

Disaster, Infrastructure and Participatory Knowledge: The Planetary Response Network

Brooke D. Simmons^{1,†}, Chris Lintott², Steven Reece³, Campbell Allen², Grant R. M. Miller², Rebekah Yore^{4,5}, David Jones⁴, Sascha T. Ishikawa⁶, Tom Jardine-McNamara⁷, Amy R. Boyer⁸, James E. O'Donnell², Lucy Fortson⁹, Danil Kuzin¹, Adam McMaster^{10,11,2}, Laura Trouille⁸, Zach Wolfenbarger⁸

¹ *Department of Physics, Lancaster University, Bailrigg, Lancaster, LA1 4YB, UK*

² *Department of Physics, University of Oxford, Keble Rd, Oxford OX1 3RH, UK*

³ *Department of Engineering Science, University of Oxford, 17 Parks Road, Oxford OX1 3PJ, UK*

⁴ *Rescue Global, c/o Niren Blake LLP, 2nd Floor, Solar House, 915 High Road, London, England, N12 8QJ*

⁵ *Institute for Risk and Disaster Reduction, University College London, London, UK*

⁶ *RAND Corporation, 1776 Main St, Santa Monica, CA 90401, USA*

⁷ *It's Ravenous Limited, 4 Aldershot Close, Stockport SK7 1AS, UK*

⁸ *Department of Citizen Science, The Adler Planetarium, Chicago, IL 60605, USA*

⁹ *School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA*

¹⁰ *School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, UK*

¹¹ *DISCnet Centre for Doctoral Training, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK*

ABSTRACT

There are many challenges involved in online participatory humanitarian response. We evaluate the Planetary Response Network, a collaboration between researchers, humanitarian organizations, and the online citizen science platform Zooniverse. The PRN uses satellite and aerial image analysis to provide stakeholders with high-level situational awareness during and after humanitarian crises. During past deployments, thousands of online volunteers have compared pre- and post-event satellite images to identify damage to infrastructure and buildings, access blockages, and signs of people in distress. In addition to collectively producing aggregated “heat maps” of features that are shared with responders and decision makers, individual volunteers may also flag novel features directly using integrated community discussion software. The online infrastructure facilitates worldwide participation even for geographically focused disasters; this widespread public participation means that high-value information can be delivered rapidly and uniformly even for large-scale crises. We discuss lessons learned from deployments, place the PRN's distributed online approach in the context of more localized efforts, and identify future needs for the PRN and similar online crisis mapping projects. The successes of the PRN demonstrate that effective online crisis mapping is possible on a generalized citizen science platform such as the Zooniverse.

Keywords

case study, crisis mapping, digital humanitarian, virtual citizen science, distributed citizen science, volunteer participants, disaster response, earth observation, satellite images, Zooniverse

[†] Corresponding author: b.simmons@lancaster.ac.uk

INTRODUCTION

During and after a natural disaster or other humanitarian crisis, there is a need for real- or near-real-time information about an affected area. The informational needs of responders¹, decision makers, and other stakeholders on the ground often fall under the broad term “situational awareness”: these needs pertain to key information about features in the environment that inform decision making at all levels. It is often the case that a relatively small group of responders has an urgent need for accurate, reliably sampled information concerning a large area of interest (AOI). In the digital age, ample relevant data frequently exists but responders often lack the additional resources required to extract this information themselves (Tapia and Moore, 2014). This tension between a flood of data and a trickle of resources is familiar to many citizen science practitioners (Bonney *et al.*, 2009; Wiggins and Crowston, 2011).

There are many types of crowdsourced responses to humanitarian crises, from locally-driven efforts (e.g., Dailey and Starbird, 2014; Brown *et al.*, 2016) to those which filter SMS messages (Munro, 2013) and social media (Hughes and Palen, 2009; Popoola *et al.*, 2013) and those which involve both local and remote participants (Rehman Shahid and Elbanna, 2015; Dittus, Quattrone and Capra, 2017). This work focuses on the application of distributed online citizen science principles and methods to the creation of humanitarian maps, sometimes called crisis maps (e.g., Ziemke, 2012; de Albuquerque, Herfort and Eckle, 2016), based on analysis of satellite imagery. Specifically, this paper offers a case study of the Planetary Response Network (PRN),² which since 2014 has provided rapid, accurate, high-value situational awareness to responders and decision-makers in a disaster context. The PRN is run as a partnership between the Zooniverse, computer science researchers, and humanitarian response and resilience organisations. In the generalized citizen science project typology of Parrish *et al.* (2018), the

PRN is a “data generated: active participation: virtual: multiple independent classifications” project type. Within the Disaster Research Center (DRC) typology (for a recent review of this typology, see Strandh and Eklund, 2018), the PRN blends aspects of both the Extending and Emergent types of disaster response organizations. It is an online, distributed project that, within a recently-described geographic citizen science framework (Skarlatidou and Haklay, 2021), uses participatory design principles to capture volunteered geographic information (VGI) in a generalized (i.e., not purely geographic) web application interface.

The field of online distributed humanitarian mapping is relatively new and still evolving (Meier, 2011, 2012; Ziemke, 2012; Sharma and Joshi, 2019; Turk, 2020). Assessments following the 2010 earthquake in Haiti (e.g., Zook *et al.*, 2010; Harvard Humanitarian Initiative, 2011) and subsequent disasters have shown both the promise of crowdsourced crisis mapping and its challenges. For example, Westrope, Banick and Levine (2014) analyzed the OpenStreetMaps response to Typhoon Haiyan and found that the rapid assessment was valuable to responders but was limited by inaccurate labels, lack of participant training, and uneven coverage of the AOI. More generally, some of the challenges of online crisis mapping are relatively specific to that application, such as the lack of shared technical language between project teams and responders, competing priorities for security, privacy and publicity, and the psychological toll that participation in a project may take on its volunteers (Ziemke, 2012; Liu, 2014). Other challenges, such as data verification and reliability (Haklay, 2013; Kosmala *et al.*, 2016; Parrish *et al.*, 2018) and boundary issues between different involved groups (Shirk *et al.*, 2012; Oswald, 2020), have found multiple solutions in the broader realm of citizen science. In many cases these solutions were known to citizen science practitioners prior to the mainstream emergence of online distributed humanitarian mapping (e.g. participant training, label aggregation and validation, and uniform data

¹ In this manuscript, the term “responder” is used to refer to those who use the results of this project to coordinate and execute humanitarian responses on the ground in affected areas. It is distinct from terms such as “volunteer”,

“classifier”, and “participant”, which refer to those who participate in the online citizen science project to provide individual feature labels.

² planetaryresponsenetwork.org

coverage; Lintott *et al.*, 2008). Blending best practices from both fields is thus of high potential value to each.

The PRN is slightly different from other crisis mapping efforts (e.g., Ushahidi, Humanitarian OpenStreetMap), in part because it runs on the Zooniverse citizen science platform instead of a platform built specifically for mapping. As such, it must approach the mapping aspect of deployments slightly differently, but benefits from over a decade of lessons learned regarding citizen science project design, data quality and community engagement. It additionally benefits from exposure to the Zooniverse community of over 2 million registered participants. The choice of platform enables the PRN to complement, rather than compete with, existing crisis mapping efforts. This case study aims to describe the project design, present its deployment statistics, and evaluate project outcomes in the context of the field of distributed online crisis mapping, including a summary of lessons learned.

PROJECT DESIGN

The PRN has deployed multiple times for specific disaster responses. To date, the PRN has exclusively made use of satellite imagery for data assessment. Satellites provide verified data over large areas, which facilitates the rapid and broad situational analysis the PRN prioritizes. Within that context, we make project design choices (described in subsections below) that align both with the technical design of the Zooniverse platform and with domain-specific needs.

Deployments of the PRN also adhere to best practices in citizen science (e.g., Lintott and Zooniverse, 2010; Gold, 2019); projects must have a genuine and beneficial outcome whose goal can be expressed in advance. In a disaster relief context, while outcomes may include e.g. providing training data for machine learning algorithms (Isupova *et al.*, 2018; Weber and Kané, 2020), the primary goal of providing useful information to improve situational awareness necessitates that the PRN include partners on the ground. Co-creating crisis maps with local stakeholders is a critical step to prevent the

abstraction of digital humanitarian projects from those they affect (Mulder *et al.*, 2016).

While the specifics of the PRN pipeline have evolved over time, they generally involve three phases: (1) a *planning* phase, where all parties consult, gather available data, and decide on urgent situational awareness needs; (2) a *crowd labeling* phase, where volunteer classifiers assess available imagery (typically answering questions about and marking features on images); and (3) an *analysis* phase, where crowd labels are aggregated to produce a consensus result and feature assessments are produced. These phases may be repeated several times during a single deployment as more data becomes available and/or the response needs evolve. Following each analysis phase, the PRN produces a “heat map” for each feature of interest and delivers these to responders. In addition to the post-analysis heat maps, the involvement of the crowd facilitates rapid identification of unexpected features that impact situational awareness.

Once a deployment is complete, the PRN organizational team meets to discuss successes, failures, and near-failures, so that we learn from these and improve our future pipeline. Many of the lessons learned discussed below crystallized from these post-deployment self-assessments.

Planning Phase

When a disaster is imminent or has just occurred, PRN partners consult with each other to decide whether a deployment is appropriate. For the PRN deployments to date, our primary domain-expert partner was Rescue Global, a UK response and resilience charity that operates worldwide.³ Rescue Global’s work includes search and rescue activities as well as liaising with local governments and stakeholders.

The advance relationships Rescue Global builds on the ground with local governmental and community-based organisations are crucial to the project. Given that many disasters strike communities in the global south, the need for local partnerships is especially critical for a response effort (such as the PRN) whose members are primarily from the global north. Rescue

³ rescueglobal.org

Global's ongoing partnerships have included organisations such as the Caribbean Disaster and Emergency Management Agency (CDEMA) and the Mexican Jewish non-profit Cadena, which operates local branches throughout Central and South America. These larger non-profit partnerships have also facilitated relationships with individual communities in these and other regions, which helps Rescue Global communicate local priorities to the PRN team at all stages of a deployment. By partnering directly with a single organization whose expertise includes cultivating multiple local relationships, the PRN team can maximize the chances that a deployment will appropriately address the needs of affected individuals and communities, while minimizing the costs and risks of developing separate relationships from a remote position. The necessity of involving local stakeholders is echoed by many studies of geographical citizen science projects (e.g., Hecker *et al.*, 2019; Skarlatidou and Haklay, 2021).

For each PRN deployment, once Rescue Global confirms they will deploy to the region and would benefit from improved situational awareness, the other partners begin assessing imagery data availability. Satellite data availability can be a complex landscape. Some data is fully open, such as that from NASA's Landsat or ESA's Sentinel constellations. Higher-resolution imagery often comes from commercial providers, which may have their own humanitarian data programs and may also participate in the International Disasters Charter.⁴

The date and resolution of available data varies, impacting deployment planning. Satellite imagery is typically not available for at least 24 hours after a disaster, and this can increase due to tasking delays, orbital patterns, and weather. Long delays can force deployment priorities to shift. The resolution of available data affects the labels that can be reliably collected (Battersby, Hodgson and Wang, 2012; See *et al.*, 2013; de Albuquerque, Herfort and Eckle, 2016) and the speed of collection. Considering the needs of the responders and local decision-makers in the context of evolving data availability is critical to

ensuring the relevance and utility of crisis maps (Ziemke, 2012; Turk, 2020).

The assessment time for satellite imagery also depends on the complexity of the features being assessed. Citizen science projects across all disciplines must consider tradeoffs between labelling speed and the level of detail captured. For example, collecting binary responses about an image is fast but sacrifices considerable detail compared to drawing individual polygons around each feature. For the PRN, the needs of responders and local stakeholders, not academic researchers, take priority in project design. Responders are accustomed to operating with "good enough" information (Tapia and Moore, 2014), and generally do not require highly granular maps, especially in the early days of a response. PRN deployments have thus generally asked volunteer classifiers to label features with point marks, as this prioritizes the speed of classification while sacrificing precision at a level acceptable to responders. This choice deliberately places responders' immediate needs above the future needs of our computer science partners who use PRN damage labels to train machine learning algorithms between live deployments.⁵

The processing of satellite imagery is also part of the planning phase. PRN leadership procures available data, decides which datasets to use for a given deployment, and assembles geo-referenced pre-event and post-event image mosaics. *Ad-hoc* decisions are often required to optimize tradeoffs between cloud cover, image quality, and imaging date, given sparse time-sampling of the AOI in both pre- and post-event imagery. Following assembly and resolution matching of pre- and post-event mosaics, the images are tiled into matched sections of a manageable size for data labeling. Typically, the mosaic is sliced into square sub-images of 500-600 pixels on a side, which will be assessed by volunteer classifiers via the web and mobile devices (described further below). For high-resolution imagery the data

⁴ disasterscharter.org

⁵ This tension could be alleviated by running the project, post-deployment, in a non-urgent phase to collect more

detailed labels that reach the precision required by computer scientists.

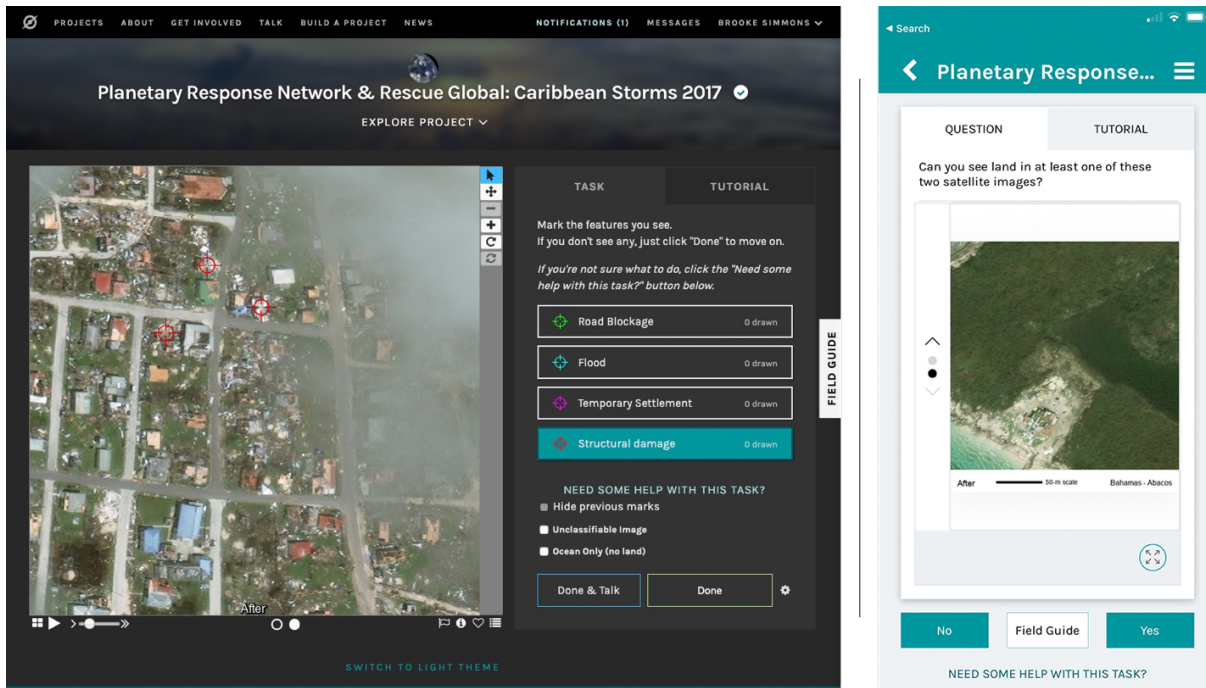


Figure 1: Example web (left) and mobile (right) interface for Zooniverse PRN projects. *Left:* Image shown: Barbuda, September 2017. *Right:* Image shown: Bahamas, September 2019. *Satellite imagery credit:* Maxar Technologies' Open Data Program.

labels are collected on image subsections as small as 150 m × 150 m, whereas for medium-resolution imagery this can be as large as 6 km × 6 km. All image subsections are large enough to provide useful context for damage assessments.

Crowd Labeling Phase

The Zooniverse is the world's largest online crowdsourcing platform for citizen research. For a detailed glossary of Zooniverse terms and infrastructure description, we refer the reader to Simpson, Page and De Roure (2014). In the PRN, Zooniverse volunteers typically classify paired pre- and post-event image subsections as a single unit of data; we refer to these image pairs as “subjects” below. A completed collection of tasks that each volunteer classifier is asked to submit within a workflow for each subject is called a “classification”.

Like most Zooniverse projects, the PRN collects multiple classifications per subject. Aggregating multiple independent classifications addresses many data quality challenges identified within citizen science generally (Haklay, 2013; Parrish *et al.*, 2018) and in Earth Observation and crisis mapping specifically (Harvard Humanitarian Initiative, 2011; Liu, 2014; Westrope, Banick and Levine, 2014; Fritz,

Fonte and See, 2017). The minimum number of classifications the PRN collects per subject and workflow is generally at least 10. Subjects are served randomly to volunteer classifiers from within a set of subjects. Our design choices contrast with those of other crisis mapping projects, which allow users to choose their own map location and to submit highly detailed labels. Our choices are designed to facilitate rapid, uniform coverage of the entire AOI and to deliver initial results to responders as quickly as possible at their required level of precision.

Figure 1 shows a PRN project screenshot showing the classification interface, with examples of both mobile and web interfaces. For deployments where available data may be of variable quality across an AOI, we have found best results with a combination of workflows which filter the subjects in a cascading fashion. Volunteers first assess whether images are “classifiable” (defined as having land visible). Only subjects with a majority of “Yes” responses are added to the feature marking workflow. One advantage of splitting the workflow is that the Yes/No workflow can also be deployed in the Zooniverse mobile app, whose interface facilitates rapid classification. Separating the project into multiple workflows thus optimizes for overall speed of classification without

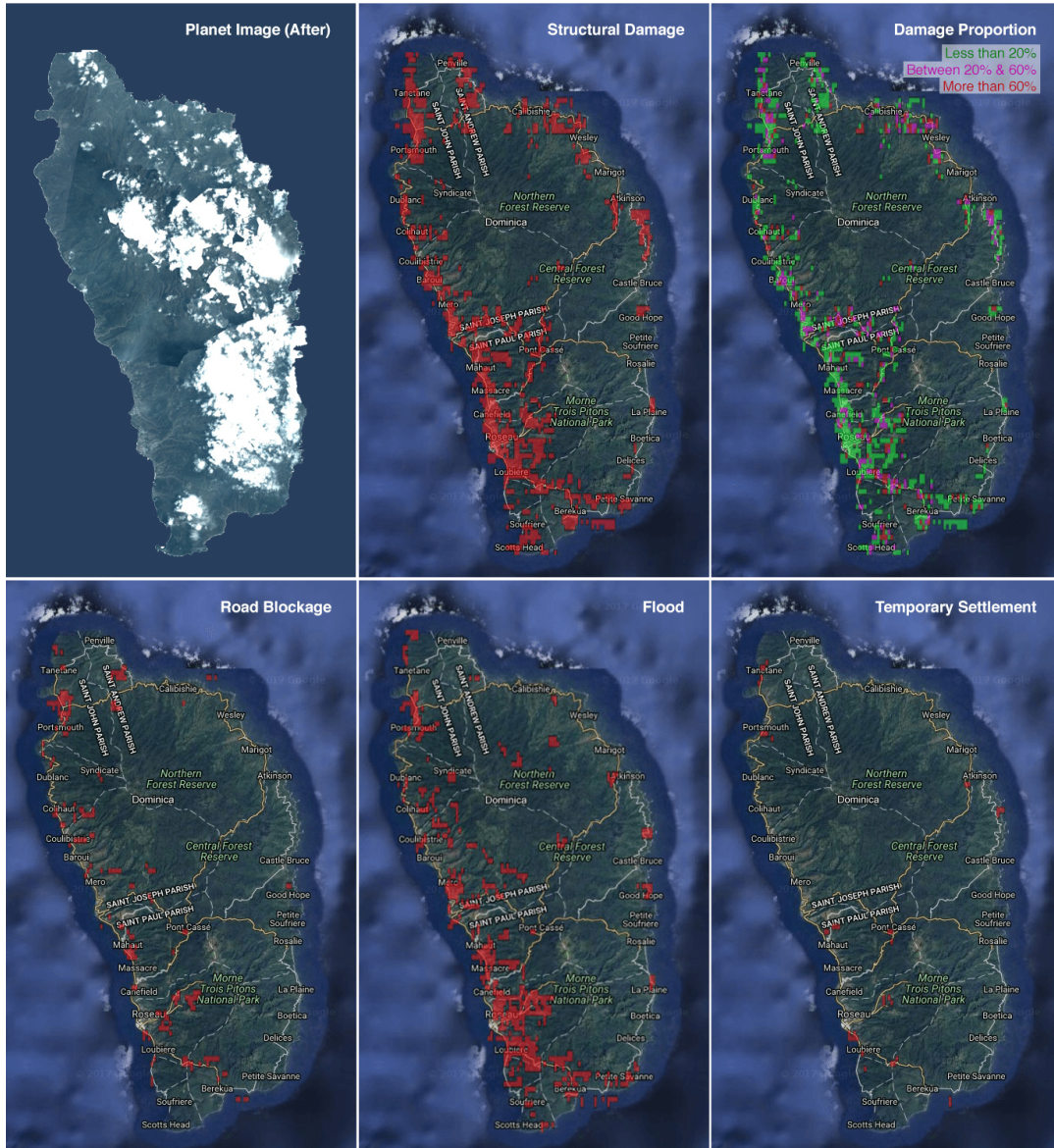


Figure 2: Heat maps for labeled features in Dominica following Hurricane Maria in 2017. Web-based maps may be zoomed in to show further detail. *Satellite imagery credit:* Planet Team (2017) License: CC-BY-SA

sacrificing completeness of coverage. Classifiers also tend to find this structure more satisfying than in early PRN deployments where feature marking workflows included high fractions of unclassifiable images. Classifiers may access additional resources for help (Katrak-Adefowora, Blickley and Zellmer, 2020) on either a specific task or on the overall project. The Field Guide feature, available for all workflows, allows classifiers to see multiple examples of the different types of labels.

When further data becomes available, the crowd labeling phase may continue with additional rounds

of imagery. The project organizers announce each image set to participants in the project’s Talk community discussion area; if additional attention is required, the Zooniverse team may also send a newsletter to either the existing project community or a wider Zooniverse audience. We have sustained high levels of engagement over several weeks due to newsletter campaigns and regular data releases. Each deployed Zooniverse project remains active until the PRN partners decide the crowd labeling phase is complete. As soon as participants classify the first image set, the analysis phase of the PRN begins.

Analysis Phase

In the analysis phase, we derive consensus from individual labels by volunteer classifiers. This aggregation step accounts for individual variations in assessment styles and minimizes the impact of the small fraction of classifications that contain errors (Lintott *et al.*, 2008; Simmons *et al.*, 2017) by resolving disagreements among the crowd and arriving at a high-confidence final label set.

In past deployments, individual labels have been aggregated using the Independent Bayesian Classifier Combination (IBCC) machine learning algorithm (Simpson *et al.*, 2013; Ramchurn *et al.*, 2016). This algorithm calculates the reliability of each classifier and combines their labels into a single map by weighting each classifier contribution according to their reliability. The IBCC algorithm is unsupervised, i.e., no ground truthing (physical and/or expert verification of feature labels) is required to produce the crisis maps. However, if expert labeling is available then the algorithm can fold these into the maps as ground truth.

The aggregation incorporates individual point-marked labels for each feature type, as well as “blank” marks where a classifier indicated there was no feature of interest in the image. The aggregated labels are then turned into “heat maps” for each feature type. A heat map is a color-coded overlay on the satellite image. Figure 2 shows an example of heat maps provided for Dominica following Hurricane Maria in 2017, based on medium-resolution imagery from Planet. The resolution of the heat map grid is chosen to reflect both the resolution of the satellite imagery and the level of map detail required by the responders. These digitized maps are bundled together and forwarded to our partner responders.

PROJECT DEPLOYMENT STATISTICS

The PRN has so far deployed live, time-sensitive projects 4 times: (1) following the two earthquakes with magnitudes 7.8 and 7.5 in Nepal in spring 2015, (2) following the 7.8-magnitude earthquake in

Ecuador in April 2016, (3) following Hurricanes Irma and Maria in the Caribbean in autumn 2017, and (4) following Hurricane Dorian in autumn 2019. We have additionally prepared other projects that did not deploy (i.e., they never entered the *crowd labeling* phase described above). Projects may fail to deploy for a number of reasons, including changes to ground access granted by local governments and revised estimates of event severity (e.g. a hurricane that changes course or dissipates). We choose to focus here on the two most recent deployments of the PRN, as these projects exemplify the general properties of PRN deployments while being similar enough to each other to facilitate comparison.⁶

Quantitative and technical details for the PRN Caribbean deployments are given in Appendix 1. The two projects jointly collected over 1 million individual classifications from thousands of online participants. Figure 3 shows classifications collected over time from logged-in and not-logged-in participants for both deployments.

In both projects, Rescue Global joined the PRN team as on-the-ground partners. In the 2019 deployment responding to Hurricane Dorian, we also partnered with 24 Commando Royal Engineers, a unit of the British Army’s Royal Engineers who provide military engineering support to 3 Commando Brigade Royal Marines and who had additional assessment needs (see Appendix 1).

Rescue Global has good relationships with multiple governmental and non-governmental organizations (NGOs) active in the Caribbean region. As a result, the heat maps provided by the PRN had wide reach during both deployments. The maps were delivered to over 60 NGOs, the UN and the Caribbean Disaster Emergency Management Authority (CDEMA). Below we analyze deployment outcomes and critically evaluate the PRN to extract several generalized lessons that may be learned from this case study of online citizen science for humanitarian aid.

⁶ The projects we focus on here are at zooniverse.org/projects/vrooje/planetary-response-network-and-rescue-global-caribbean-storms-2017 and

zooniverse.org/projects/mrniaboc/planetary-response-network-hurricane-dorian

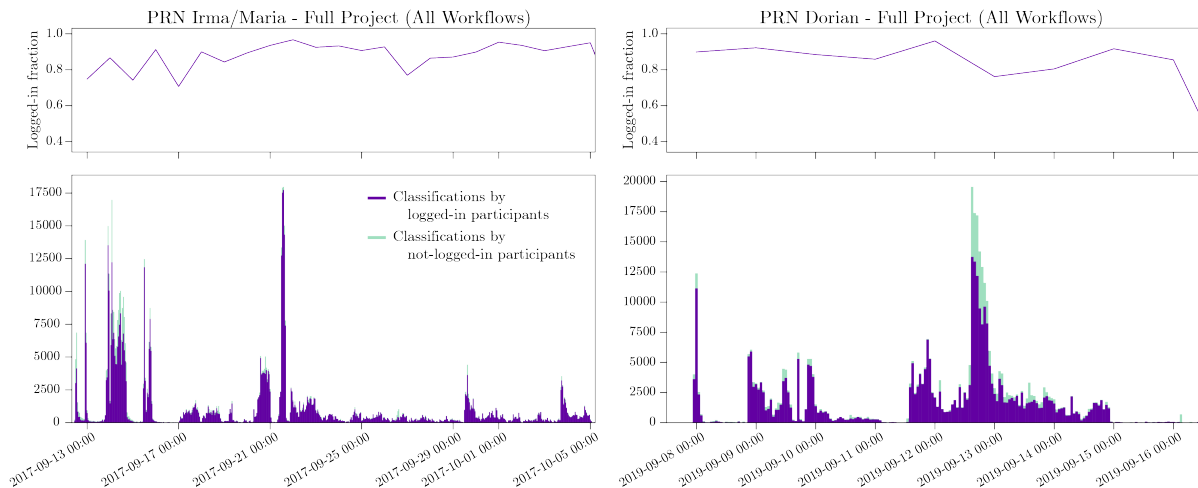


Figure 3: Classifications over time for PRN Caribbean deployments (2017, left; 2019, right). Classifications from logged-in participants are shown in purple; classifications from not-logged-in participants are added in light green, such that the combined hourly histogram shows overall classification totals. Upper panels show the daily fraction of classifications from logged-in participants.

EVALUATING PROJECT OUTCOMES

There are many relevant lenses through which to assess the outcomes of the Planetary Response Network. Some are purely related to the citizen science aspect, while others additionally consider our humanitarian objectives. Below we evaluate the PRN in the context of its success in engaging the crowd, the nature of that crowd, the speed of delivery of heat maps, the quality of the data delivered, and evidence of actual use of the maps in the field. We also comment on the process of assessing and learning from failures.

Engagement By, and With, PRN Participants

The overall number of classifiers who participate in a given Zooniverse project can vary from hundreds to hundreds of thousands. Given the duration of each PRN deployment, the fact that thousands of people have participated represents a strong level of participation compared to other short-duration Zooniverse projects. In general, the success of a Zooniverse project is related to both project design and volunteer engagement, rather than project duration (Cox *et al.*, 2015).

The project statistics (see Appendix 1) are also typical of healthy Zooniverse projects. The classification activity over time (Figure 3), while more varied than a typical Zooniverse project, is within expectations for

a project with time-sensitive data and staggered data releases (Spiers *et al.*, 2019). The fraction of classifications submitted by logged-in participants is approximately 85% throughout both projects, which is also within normal ranges for successful Zooniverse projects (Cox *et al.*, 2015).

The “Talk” discussion area, where participants can engage more deeply via open-ended discussions and tagging interesting subjects, is a valuable part of the Zooniverse ecosystem. Within the Talk area for each PRN deployment, about 10% of logged-in participants posted at least 1 comment. Figure 4 shows the average word count per post for each participant who posted on Talk. Even among those who choose to join the Talk discussion, participation is not evenly distributed: in both deployments, approximately half of participants who posted on Talk posted a single comment, with a majority (68% and 77% in the 2017 and 2019 deployments, respectively) posting 3 or fewer comments.

The nature of Talk comments varied, from single-word notes tagging an image snapshot with a hashtag (including unexpected features of interest not captured by the main classification interface) to lengthy posts with discussion, comments, and suggestions. The PRN leadership also posted regularly, using Talk to update participants with descriptions of new image datasets and sharing

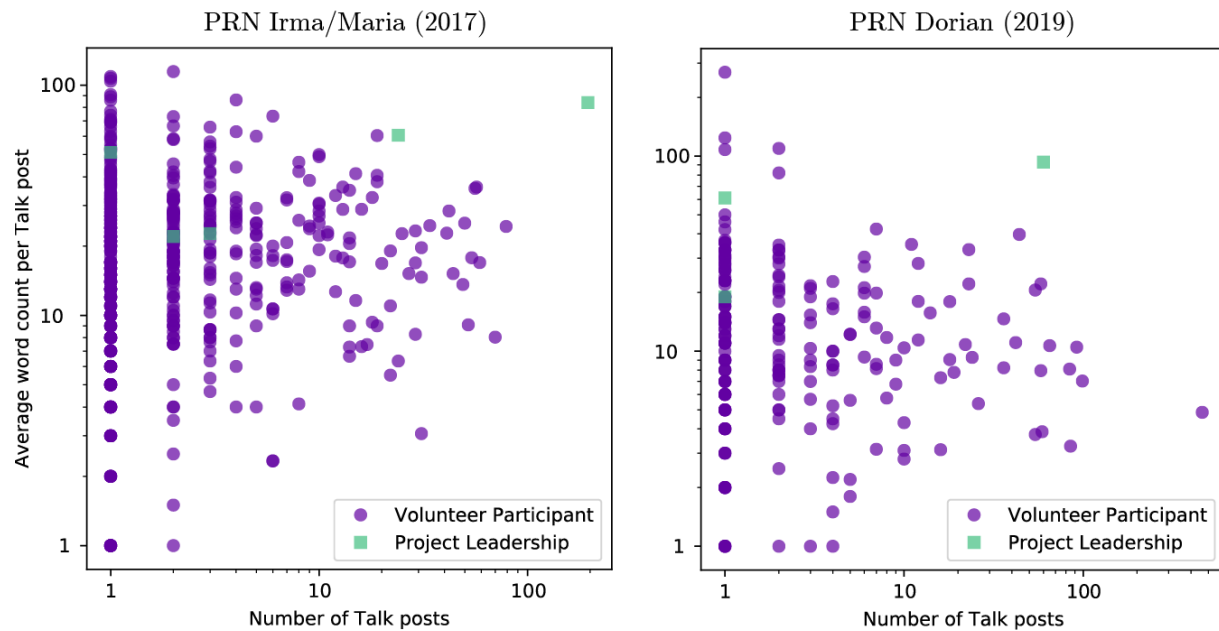


Figure 4: Average length of Talk posts for each participant versus their post count, for PRN Caribbean deployments. Volunteer participants are shown as purple circles and PRN organizational team members are shown as green squares.

preliminary heatmaps and feedback from responders. As in many Zooniverse projects, we observed “trickle-down training” occurring on Talk, in which advice and tips initially shared by the project organizers were subsequently shared by other participants in response to common inquiries from less experienced classifiers.

The Talk environment also allowed us to directly address the risk of participant burnout and secondary trauma inherent to online crisis mapping projects (Ziemke, 2012). To alleviate these risks, the PRN lead created a section of Talk explicitly for taking breaks and regularly reminded people that stepping away from the project was a healthy action that would not endanger those on the ground. Overall, this represented a small fraction of Talk interactions: it was more common that participants expressed sentiments of accomplishment and satisfaction. Still, both participant engagement generally and burnout prevention specifically are important ongoing responsibilities of teams organizing crisis mapping efforts. This is a domain-specific reflection of the need to provide a supportive environment for all

participants in a citizen science project (Resnik, Elliott and Miller, 2015; Chari, Blumenthal and Matthews, 2019).

The PRN is a virtual and distributed crisis mapping project. Analytics for the landing pages on both Caribbean projects indicate a wide global reach of visitors (over 130 countries represented overall). For both deployments, over 85% of web browser sessions originated in North America and Europe, which is generally consistent with overall Zooniverse traffic during deployment periods. More local participation from Caribbean countries represented less than 1% of browser sessions in either project; however, this fraction is higher than Caribbean traffic Zooniverse-wide (<0.1% to non-PRN projects). This difference is statistically significant⁷ and reflects an increased local interest in the PRN even while overall participation is much more widely distributed.

Therefore, while the PRN does generate some local activity, it primarily provides an opportunity for a global community to meaningfully contribute to a humanitarian aid effort, even (and possibly especially) when its members are too far from the

⁷ Using Bayesian binomial confidence intervals and comparing PRN traffic to overall Zooniverse traffic during the same dates and during a 1-week period outside hurricane season (March 2019), we estimate the probability

that the fraction of browser sessions from the Caribbean for PRN deployments is consistent with the fraction outside PRN projects is $p < 5 \times 10^{-6}$.

affected area to offer help in person. This complements humanitarian crowdsourcing projects that are more “ground up” in their origins: whereas those projects often provide highly localized and detailed individual information, the PRN can provide rapid and uniform coverage of a large affected area at a broad level of detail suitable for responders seeking to inform their initial and ongoing allocation of resources. This complementarity reflects the similarities and differences of these two approaches. Specifically, both “ground up” and “top down” crowdsourced crisis mapping efforts often strive to improve knowledge of a specific disaster by blending VGI with traditional sources of geospatial information (Zook *et al.*, 2010), without placing the burden on responders to become experts in either. Locally-driven efforts often harness high levels of relevant local factual and cultural knowledge (Goodchild and Glennon, 2010) that complements the humanitarian skills of response organizations (Strandh and Eklund, 2018). In contrast, the more distributed online projects allow anyone to participate regardless of whether they have the resources or skills to join a locally organized effort. A distributed project such as the PRN, which is hosted on an established citizen science platform, also has access to a high fraction of participants with significant prior experience participating in citizen science, which facilitates accurate label collection and aggregation. Furthermore, the PRN team includes members with substantial experience running citizen science projects, which allows us to translate between our citizen science community and our responder partners. This significantly alleviates boundary issues when planning and deploying a response. We would stress, however, that it is extremely important for a distributed project such as ours to continually center local needs and priorities, including sharing results with local communities (e.g., Mulder *et al.*, 2016) as soon as it is safe to do so.

Data Quality and Delivery Speed

The need for high-quality image labels was a key motivator for hosting the PRN on the Zooniverse. The platform is designed to enable high-quality data collection via proven methods such as collecting multiple independent classifications per subject (Kosmala *et al.*, 2016; Parrish *et al.*, 2018). Ensuring

data quality is also a factor in ethical considerations in citizen science (Resnik, Elliott and Miller, 2015). Zooniverse projects have produced data labels whose quality matches and even exceeds that of a single expert (e.g., Lintott *et al.*, 2008; Swanson *et al.*, 2015). Additionally, the aggregation method we use is able to reduce the effect of noisy inputs from individual classifiers and account for individual skill levels in reaching consensus. This is especially important as ground truthing is generally not available in advance, which makes precise calibration challenging. We thus rely on feedback from the field to regularly assess the quality of our heat maps.

There are several potential bottlenecks to delivering heat maps rapidly enough to be of use to responders. These include:

- **Domain Expertise:** Humanitarian crisis mapping is inherently multi-disciplinary (Ziemke, 2012), and several types of expertise are required to successfully deploy all stages of the PRN. These include knowledge of Geographical Information System (GIS) data sources and formats, disaster response and resilience, data science and statistical methods, and citizen science project design.
- **Data Availability:** Satellite and/or aerial imaging data may be unavailable for several reasons. Some of these are outside the project team’s control: tasking delays, weather issues, and corrupted data may all mean that needed data is either unavailable or severely delayed.
- **Data Access:** Image data may exist, but not be accessible to the project team. Obstacles to data access can take the form of paywalls, bureaucratic delays, or technological problems (e.g., bandwidth issues).

The PRN has been able to deploy projects with very rapid turnaround, including initial heat map delivery just hours after beginning the crowd labeling phase. This deployment speed is possible in large part due to the work that takes place prior to, and in between, active project phases. Advance preparation is thus a key solution to all of the obstacles described above.

The PRN’s advance preparation alleviates the issues described above in several ways. We have carefully

assembled the PRN partnership specifically to address the domain expertise needs of a distributed online crisis mapping project. These needs were identified as the PRN initially formed, but have been refined following assessments of successes and failures of deployments. Some of the PRN partnership assembly has included negotiating and building relationships with partners; other aspects have involved training existing team members in new skills, developing new features on the Zooniverse platform that are useful to the PRN, and documenting end-to-end procedures and guidelines for all phases of PRN deployment. These procedures have enabled us to redirect crowd attention to alternative workflows when post-event data is scarce. Pre- and post-event data has sometimes been available to the PRN days before the same data is made openly available due to previously established partnerships with commercial satellite companies.

We therefore strongly agree with the findings of other studies (e.g., Harvard Humanitarian Initiative, 2011; Liu, 2014) that preparation is critical to a successful crisis mapping deployment. We would also note a need to inform preparation with the need to be flexible for each deployment. This is consistent with the idea that advance preparation must prioritize “articulation” work (Hughes and Tapia, 2015), which develops means of inter- and intra-organization information exchange so that this flexibility is possible during time-critical periods without sacrificing efficacy.

Open data is a major benefit to crisis mapping efforts. However, improvements are possible in this area. While some sources of satellite data are technically open, they are not always open in a way that actually encourages their use. Since the PRN was created, we have encountered various issues with “open” data that have measurably slowed active deployment efforts. These have included restricted bandwidth for downloads of uncompressed GeoTIFF image tiles, previously open image search tools becoming paywalled with little or no notice, and unsearchable image lists presented in raw form with no separate geographic metadata available.

The best implementations of open data have allowed us to save hours or days in the planning phase of the PRN. For example, humanitarian users of Planet⁸ data have access to the full commercial search and download area of the Planet website and API, which significantly streamlines data acquisition. Additionally, Amazon Web Services’ Open Data program hosts a copy of processed, mosaiced ESA Sentinel-2 image data with no restrictions on transfer bandwidth.⁹ This additionally facilitates GIS image processing in the cloud, which can save further time during live deployments. If more sources of satellite imagery took similar approaches in the future, this would encourage more rapid and more successful crisis mapping projects, including but not limited to the PRN.

Improvements to Crisis Response

Ultimately, the success of a crisis mapping project depends on whether it achieves its stated goals of positively impacting the response effort during and after a particular deployment. This framing encapsulates several of the ECSA’s ten principles of citizen science within a humanitarian context.

Given that every disaster is different, it is difficult to rigorously quantify the effect of adding a distributed crisis mapping effort to a disaster response. While a project team may be highly motivated (e.g. by academic metrics or funding pressure) to answer questions such as “how much faster will the recovery be now that heat maps are available?”, it is not trivial to extract this information even by comparing to previous disasters where the project did not deploy. Attempts to collect uniform quantified feedback *in situ* during an ongoing response represent a significant local resource demand. For a distributed project such as the PRN, it would be particularly inappropriate for the organizers to either make these demands from their position of safety, or to risk sending personnel into an ongoing response for this purpose. Therefore, feedback to online distributed crowd mapping efforts on the utility of crisis maps is typically qualitative and often arrives after the most active periods of a response effort.

⁸ planet.com

⁹ registry.opendata.aws/sentinel-2

Evidence of improvements during the example deployments considered here vary. They include broad messaging that the heat maps provided were actively used on the ground to inform the ongoing effort, as well as specific examples. The project team collected specific qualitative feedback on the use of the results of the analysis phase. These include the use of road-blockage heat maps to optimize personnel allocation and more quickly restore critical national infrastructure, the incorporation of features flagged directly on Talk into flight plans for aerial assessment and evaluation of airstrips, and the use of building-damage heat maps to target priority areas for rapid ground-truth assessments and subsequent allocation of aid. We also received feedback that responders generally found maps of proportional damage (the fraction of structures in a given area which are damaged) extremely useful, especially as they worked their way to more isolated communities and health centers.

The above examples represent a minimum assessment of the utility of PRN heat maps. Rescue Global also distributed the maps widely to other organizations on the ground, and the remote PRN team did not receive feedback from those organizations as to whether and how the maps were used. Preparation for future deployments will include establishing contact with a wider group of organizations in advance, in part to facilitate end-of-response collection of feedback from these groups.

Learning from Failures

Across all PRN deployments, lessons learned from failures underpin the majority of our subsequent successes. As described in the Project Design section above, several of the general best practices described herein arose from specific challenges and failures during deployments, some of which occurred during earlier deployments than those in focus in this paper. For example, prolonged cloud cover immediately following the Nepal earthquakes in 2015 forced the PRN to shift deployment priorities from damage assessment in post-event satellite images to prediction of locations likely to need urgent aid, based on comparing recent pre-event images with (then incomplete) existing building maps. This *ad hoc* shift subsequently improved our ability to statistically

incorporate other sources of geographic information (such as earthquake severity maps) into our analysis pipeline. The value of preparation was further reinforced by another lesson from the PRN Ecuador deployment in 2016, when post-project reflection on deployment delays led us to create project templates including logos, disclaimers and other boilerplate language that could be pre-approved by funders, enabling the team to focus on more pressing issues during a live deployment.

Additionally, communication with other crisis mapping teams and leaders has enabled us to learn from (and thus not repeat) external challenges. For example, early informal discussions with people involved in the 2010 Haiti earthquake and 2013 Typhoon Haiyan responses highlighted the need to ensure our partnerships include local connections and underscored the interdependence between teams who focus on technology-driven solutions (such as the PRN) and more traditional, hierarchical aid organizations (Zook *et al.*, 2010). Discussion with external crisis mapping experts has also been considered alongside the PRN team's expertise in citizen science methods to inform our project design. For example, the design choices described in the Crowd Labeling Phase section above reflect an intent to minimize the biases that can appear in VGI data following a crisis (Goodchild and Glennon, 2010; Zook *et al.*, 2010). All these considerations are especially important in applications where organizations based in the global north deploy to the global south, which is common in disaster and humanitarian aid.

Overall, it is critically important to both communicate with other experts and include an internal reflection phase following each deployment, where the team aggregates both successes and failures into lessons for the next deployment. This should be part of the normal process for any crisis mapping project.

CONCLUSION

The Planetary Response Network is a distributed online crisis mapping project that has deployed multiple times since its creation in 2014. The project approaches crisis mapping through a strong citizen science lens, with particular focus on global community engagement, data quality, and producing

outputs with clear utility to responders, decision makers, and other stakeholders. This case study, which focuses on the most recent deployments of the PRN, has produced several lessons learned following evaluation of the project's structure, deployments and outcomes:

- Distributed online crisis mapping projects, a particular type of humanitarian citizen science, play a positive role in the digital humanitarian sphere. It is critical that distributed projects such as the PRN continue to center the requirements of local stakeholders at all stages of project deployments.
- There is a strong worldwide interest in response efforts following a disaster; distributed online crisis mapping provides an excellent way for people to help even when they are distant and/or cannot afford to financially support aid efforts.
- Crisis mapping projects such as the PRN become robust when end-to-end response procedures are established early and the collaboration has prepared as much as possible in advance of deployments. With the addition of citizen science as a key response component, project design and planning become even more important, to ensure that the project makes ethical use of participants' time and contributions. Pre-established procedures must remain flexible to the needs of each specific deployment.
- While some labelling requirements vary depending on the type of deployment, other features (e.g. infrastructure damage, access blockages, signs of *ad hoc* shelters) tend to be high priorities for situational awareness needs across multiple types of disasters.
- Responders can generally tolerate more uncertainty in crisis maps than a purely academic study. This affects the design and deployment of a crisis mapping project, another reason it is vital to regularly liaise with responders and local stakeholders.
- Point markings, even of extended features, are sufficient to flag features of interest during live deployments. However, these pose challenges when using these labels to train advanced damage detection algorithms. Communication between academic and responder team

members is critical to establish priorities in advance and explore ways to alleviate tensions between urgency and precision that satisfy all parties.

- Discussion software such as Zooniverse Talk provides an important way for crisis mapping project participants to identify serendipitous features of interest, train each other in advanced feature detection, and remain engaged throughout a deployment. Supervision of the discussion by team leaders is important for guiding self-training as well as identifying and intervening with participants showing signs of secondary distress. The latter is a particularly important ethical consideration.
- Feedback between remote project organizers and responders on the ground is often qualitative and anecdotal rather than quantitative or statistical, necessitated by practical reasons of resource allocation priority and risk management. Qualitative feedback can still be extremely useful, provided team members accustomed to quantitative methods adjust their self-assessment techniques.
- It is immensely valuable when satellite image providers make their data open and easily accessible to crisis mapping projects. Providers can enable humanitarian projects to significantly improve deployment and response times by ensuring their open data includes processed data products and that such data is easy to search and acquire.

These lessons may be generalizable to other crisis mapping efforts, particularly those which conform to the virtual and distributed typology of citizen science projects. The successes of the PRN provide evidence that online distributed crisis mapping projects can be effective even when run on a generalized citizen science platform (such as the Zooniverse) not specifically designed for geographical citizen science. By applying best practices of citizen science and involving responders and local stakeholders at all stages of project execution, online distributed crisis mapping can add a valuable layer of information to complement purely community-based response efforts.

Looking forward, work within the PRN partnership continues. In particular, the team leadership is pursuing promising avenues for streamlining the planning phase using more automated image processing techniques. We are also developing a machine learning pipeline, trained on labels provided by project participants, to provide early estimates of heat maps even of new geographical regions (Kuzin *et al.*, 2021). This will allow responders to access high-value information for early resource allocation. It will also enable the project to direct participant attention to higher-level tasks, easing the tension between urgency and the need for detailed information.

There is also significant potential to develop the PRN further to include more frequent deployments as well as longer-term deployments. Deploying projects in partnership with local stakeholders to address risk reduction and resilience needs, for example, would enable the PRN to provide value at all stages of the disaster life cycle. These deployments would also benefit from the reduced time pressure compared to a response deployment, and they would allow the PRN community to remain active on an ongoing basis.

SUPPLEMENTAL MATERIALS

Appendix 1: Quantitative Details of PRN Caribbean Deployments

Following partner discussions, the PRN activated the planning phase for its 2017 Caribbean deployment on September 8, 2017. After the team processed the available imagery and decided on a project design, the crowd labeling phase was launched on September 12, 2017. Between September 12 and October 5, 5,331 logged-in participants and 6,422 not-logged-in IP addresses provided approximately 650,000 classifications of the islands of Antigua, Barbuda, Dominica, Guadeloupe, Puerto Rico, Turks and Caicos, and the US and British Virgin Islands. This effort, which labeled damage from both Hurricanes Irma and Maria, represents approximately 2.8 years' worth of full-time person-effort. Effort is measured from summing the duration of individual classifications (recorded by the Zooniverse software) and assuming 1 year of FTE is 40 hours' work per week for 52 weeks. A small number of classification durations are extremely high (e.g. 10 minutes or more where the median classification is 10 seconds); this is typical among Zooniverse projects. For the 0.4% of classifications meeting this criterion for this project we assume this represents the classifier pausing their work without closing their browser window, and replace those duration values with the average duration across all non-outlier classifications. For purposes of label aggregation, each not-logged-in IP address is treated as a unique classifier/participant.

The Gini coefficient (Gini, 1936) for classifications per volunteer was 0.82. The Gini coefficient measures inequality in a distribution and ranges from 0 (all classifiers contribute equal amounts) to a limit of 1 (one classifier contributes ~all classifications). In a citizen science context, Gini values between 0.7–0.9 generally indicate that a project can both recruit new participants and retain a loyal community of classifiers (Cox *et al.*, 2015).

The first image set was labeled by volunteer classifiers in approximately 2 hours. The first heat maps were delivered to Rescue Global on the same day, including labels for building damage, road blockage, flooding, and temporary shelters. These features, particularly building damage and road blockage, tend to be of high value to responders following a wide variety of types of disaster. Overall, classifiers submitted over 180,000 individual marks labelling the features of interest. Exact mark counts for each feature type are given in Table A1, with the time series of submitted classifications shown in Figure 3. Additionally, 537 participants from this deployment posted at least 1 comment on the project's "Talk" discussion boards.

The second PRN deployment in the Caribbean began its planning phase on September 5, 2019. This project focused on the northern Bahamas, which were severely affected by Hurricane Dorian. The initial launch of the crowd labeling project was on September 7, 2019; the first data set was labeled by the crowd in under 5 hours. Between September 7 and 16, the project collected approximately 365,000 classifications from 2,404 logged-in participants and 2,105

not-logged-in IP addresses (Gini coefficient: 0.81). Overall, classifiers submitted over 135,000 individual labels of features of interest (see Table A1). The total classification effort represents approximately 1 year of full-time person-effort (replacing 0.2% of high-duration outliers with the average duration, as above). Of the 2,404 logged-in participants who contributed to the project, 219 posted at least 1 comment on Talk.

The features labeled during the 2019 Caribbean deployment were building damage, road blockage, flooding, temporary shelters, underwater hazards and helicopter landing sites (HLSs). The latter two features were added after consultation with 24 Commando Royal Engineers, who were deployed in partnership with Rescue Global to the Bahamas. For the HLS feature assessment, classifiers were asked to examine each subject and assess via a Yes/No question interface whether it contained a 50 m × 50 m plot of clear land (using scale bars added to subjects; see Figure 1, right panel). This example of a specialized label added in consultation with active responders resulted in the identification of 381 high-confidence potential sites, details of which were forwarded for further assessment and potential use in reaching hard-to-access areas of the islands. High-confidence potential sites are defined as those where at least 80% of classifiers (of at least 10) have identified an image subsection as a potential HLS.

Table A1: Counts for marks on images applied by classifiers in the PRN, for the two deployments in focus in this case study. The Underwater Hazard feature was not assessed during the 2017 deployment. One additional feature, Helicopter Landing Sites, was assessed in the 2019 deployment using a single-question task (images were not marked).

<i>Label (Image Marking)</i>	<i>PRN Irma & Maria (2017)</i>	<i>PRN Dorian (2019)</i>
<i>Building Damage</i>	95,713	82,870
<i>Flooding</i>	53,981	21,403
<i>Road Blockage</i>	19,317	20,168
<i>Temporary Shelters</i>	11,947	2,392
<i>Underwater Hazard</i>	N/A	8,738

ACKNOWLEDGMENTS

PRN participants who have classified while logged into the Zooniverse are individually acknowledged on each project’s Team page.

The PRN organizational team thanks Patrick Meier for early support and partnership. We thank Planet and Maxar for opening satellite data to humanitarian efforts, NASA and ESA for open satellite data, and AWS’ Open Data program for open image hosting.

We use software packages `MATPLOTLIB` (Hunter, 2007), `PANDAS` (McKinney, 2010), `NUMPY` (van der Walt, Colbert and Varoquaux, 2011), `GDAL` (GDAL/OGR contributors, 2020), `QGIS` (QGIS Development Team, 2009), `IMAGEMAGICK` (The ImageMagick Development Team, 2021), and `TOPCAT` (Taylor, 2005).

FUNDING INFORMATION

The PRN acknowledges current and previous support (UKRI: BB/T018941/1; STFC: ST/S00307X/1; EPSRC: EP/I011587/1; ESA: Crowd4Sat). BDS acknowledges Lancaster University IAA funding and fellowship support (NASA: Einstein PF5-160143; UKRI: FLF MR/T044136/1). SR is funded by the Lloyd’s Register Foundation as part of the Alan Turing Institute’s Data Centric Engineering programme. This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.

DATA AVAILABILITY

Satellite imagery availability is subject to permission from data providers, who hold the copyright. Heat maps have previously been made public in PDF format for ease of cross-platform distribution. Other data, such as raw project statistics and project classifications, are available with the published version of this manuscript, or by contacting the corresponding author.

ETHICS AND CONSENT

All data used herein was collected under the Zooniverse privacy and data analysis policy at zooniverse.org/privacy. The PRN has received ethical approval from Lancaster University.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHORS' CONTRIBUTIONS STATEMENT

BDS leads the PRN, manages deployments, and led the manuscript writeup. CL, SR, CA, GRMM, RY, DJ and DK are primary partners in the PRN and have played various roles in forging partnerships, project deployments, and follow-up analysis. STI, TJM, ARB, AM, JEO'D, and ZW have developed features on the Zooniverse platform for use in the PRN and have provided significant support during deployments. LF and LT are experts in citizen science and have contributed to the design and implementation of the PRN. All authors contributed to the manuscript via discussions, ideas development, and writing text.

REFERENCES

- de Albuquerque, J., Herfort, B. and Eckle, M. (2016) 'The Tasks of the Crowd: A Typology of Tasks in Geographic Information Crowdsourcing and a Case Study in Humanitarian Mapping', *Remote Sensing*. MDPI AG, 8(10), p. 859. doi: 10.3390/rs8100859.
- Battersby, S. E., Hodgson, M. E. and Wang, J. (2012) 'Spatial resolution imagery requirements for identifying structure damage in a hurricane disaster: A cognitive approach', *Photogrammetric Engineering and Remote Sensing*. American Society for Photogrammetry and Remote Sensing, 78(6), pp. 625–635. doi: 10.14358/PERS.78.6.625.
- Bonney, R. *et al.* (2009) *Public participation in scientific research: Defining the field and assessing its potential for informal science education*. Washington, DC. Available at: <https://eric.ed.gov/?id=ED519688>.
- Brown, A. *et al.* (2016) 'Safecast: Successful citizen-science for radiation measurement and communication after Fukushima', in *Journal of Radiological Protection*. Institute of Physics Publishing, pp. S82–S101. doi: 10.1088/0952-4746/36/2/S82.
- Chari, R., Blumenthal, M. and Matthews, L. (2019) *Community Citizen Science: From Promise to Action*. RAND Corporation. doi: 10.7249/RR2763.
- Cox, J. *et al.* (2015) 'Defining and Measuring Success in Online Citizen Science: A Case Study of Zooniverse Projects', *Computing in Science and Engineering*. IEEE Computer Society, 17(4), pp. 28–41. doi: 10.1109/MCSE.2015.65.
- Dailey, D. and Starbird, K. (2014) 'Journalists as Crowdsourcers: Responding to Crisis by Reporting with a Crowd', *Computer Supported Cooperative Work: CSCW: An International Journal*. Kluwer Academic Publishers, 23(4–6), pp. 445–481. doi: 10.1007/s10606-014-9208-z.
- Dittus, M., Quattrone, G. and Capra, L. (2017) 'Mass participation during emergency response: Event-centric Crowdsourcing in Humanitarian mapping', in *Proceedings of the ACM Conference on Computer Supported*

- Cooperative Work, CSCW*. New York, NY, USA: Association for Computing Machinery, pp. 1290–1303. doi: 10.1145/2998181.2998216.
- Fritz, S., Fonte, C. and See, L. (2017) 'The Role of Citizen Science in Earth Observation', *Remote Sensing*. MDPI AG, 9(4), p. 357. doi: 10.3390/rs9040357.
- GDAL/OGR contributors (2020) '{GDAL/OGR} Geospatial Data Abstraction software Library'. Available at: <https://gdal.org>.
- Gold, M. (2019) 'ECSA 10 Principles of Citizen Science'. OSF. doi: 10.17605/OSF.IO/XPR2N.
- Goodchild, M. F. and Glennon, J. A. (2010) 'Crowdsourcing geographic information for disaster response: A research frontier', *International Journal of Digital Earth*, 3(3), pp. 231–241. doi: 10.1080/17538941003759255.
- Haklay, M. (2013) 'Citizen science and volunteered geographic information: Overview and typology of participation', in *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Springer Netherlands, pp. 105–122. doi: 10.1007/978-94-007-4587-2_7.
- Harvard Humanitarian Initiative (2011) *Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies*. Washington, D.C. and Berkshire, UK. Available at: <https://hhi.harvard.edu/publications/disaster-relief-20-future-information-sharing-humanitarian-emergencies>.
- Hecker, S. *et al.* (2019) 'Innovation in open science, society and policy – setting the agenda for citizen science', *Citizen Science*. UCL Press, pp. 1–24. doi: 10.2307/J.CTV550CF2.8.
- Hughes, A. L. and Palen, L. (2009) 'Twitter adoption and use in mass convergence and emergency events', *International Journal of Emergency Management*. Inderscience Publishers, 6(3–4), pp. 248–260. doi: 10.1504/IJEM.2009.031564.
- Hughes, A. L. and Tapia, A. H. (2015) 'Social Media in Crisis: When Professional Responders Meet Digital Volunteers', *Journal of Homeland Security and Emergency Management*. Walter de Gruyter GmbH, 12(3), pp. 679–706. doi: 10.1515/jhsem-2014-0080.
- Hunter, J. D. (2007) 'Matplotlib: A 2D Graphics Environment', *Computing in Science Engineering*, 9(3), pp. 90–95. doi: 10.1109/MCSE.2007.55.
- Isupova, O. *et al.* (2018) 'BCCNet: Bayesian classifier combination neural network', in *NeurIPS Workshop on Machine Learning for the Developing World*. Available at: <https://arxiv.org/abs/1811.12258>.
- Katrak-Adefowora, R., Blickley, J. L. and Zellmer, A. J. (2020) 'Just-in-Time Training Improves Accuracy of Citizen Scientist Wildlife Identifications from Camera Trap Photos', *Citizen Science: Theory and Practice*, 5(1), p. 8. doi: 10.5334/cstp.219.
- Kosmala, M. *et al.* (2016) 'Assessing data quality in citizen science', *Frontiers in Ecology and the Environment*, 14(10), pp. 551–560. doi: 10.1002/fee.1436.
- Kuzin, D. *et al.* (2021) 'Disaster mapping from satellites: damage detection with crowdsourced point labels', in *3rd Workshop on Artificial Intelligence for Humanitarian Assistance and Disaster Response (NeurIPS 2021)*. Available at: <https://arxiv.org/abs/2111.03693>.
- Lintott, C. ~J. *et al.* (2008) 'Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey', 389, pp. 1179–1189. doi: 10.1111/j.1365-2966.2008.13689.x.
- Lintott, C. and Zooniverse (2010) *What Makes A Good Zooniverse Project?*, *The Zooniverse Blog*. Available at: <https://blog.zooniverse.org/2010/06/30/what-makes-a-good-zooniverse-project/> (Accessed: 23 January 2021).
- Liu, S. B. (2014) 'Crisis Crowdsourcing Framework: Designing Strategic Configurations of Crowdsourcing for the Emergency Management Domain', *Computer Supported Cooperative Work: CSCW: An International Journal*. Kluwer

Academic Publishers, 23(4–6), pp. 389–443. doi: 10.1007/s10606-014-9204-3.

McKinney, W. (2010) 'Data Structures for Statistical Computing in Python', in Millman, J. and van der Walt, S. (eds) *Proceedings of the 9th Python in Science Conference*, pp. 51–56.

Meier, P. (2011) 'New information technologies and their impact on the humanitarian sector', *International Review of the Red Cross*, 93(884), pp. 1239–1263. doi: 10.1017/S1816383112000318.

Meier, P. (2012) 'Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time', *Journal of Map and Geography Libraries*, pp. 89–100. doi: 10.1080/15420353.2012.663739.

Mulder, F. *et al.* (2016) 'Questioning Big Data: Crowdsourcing crisis data towards an inclusive humanitarian response', *Big Data & Society*. SAGE Publications Ltd, 3(2), p. 205395171666205. doi: 10.1177/2053951716662054.

Munro, R. (2013) 'Crowdsourcing and the crisis-affected community: Lessons learned and looking forward from Mission 4636', *Information Retrieval*. Springer, 16(2), pp. 210–266. doi: 10.1007/s10791-012-9203-2.

Oswald, E. (2020) 'Getting to Know Other Ways of Knowing: Boundary Experiences in Citizen Science', *Citizen Science: Theory and Practice*. Ubiquity Press, Ltd., 5(1), p. 25. doi: 10.5334/cstp.310.

Parrish, J. K. *et al.* (2018) 'Exposing the Science in Citizen Science: Fitness to Purpose and Intentional Design', *Integrative and Comparative Biology*. Oxford Academic, 58(1), pp. 150–160. doi: 10.1093/icb/icy032.

Popoola, A. *et al.* (2013) 'Information verification during natural disasters', in *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion*. New York, New York, USA: ACM Press, pp. 1029–1032. doi: 10.1145/2487788.2488111.

QGIS Development Team (2009) 'QGIS Geographic Information System'. Available at: <http://qgis.org>.

Ramchurn, S. D. *et al.* (2016) 'A disaster response system based on human-agent collectives', *Journal of Artificial Intelligence Research*, 57, pp. 661–708.

Rehman Shahid, A. and Elbanna, A. (2015) *The Impact of Crowdsourcing on Organisational Practices: The Case of Crowdmapping, ECIS 2015 Completed Research Papers*. doi: 10.18151/7217474.

Resnik, D. B., Elliott, K. C. and Miller, A. K. (2015) 'A framework for addressing ethical issues in citizen science', *Environmental Science and Policy*. Elsevier Ltd, 54, pp. 475–481. doi: 10.1016/j.envsci.2015.05.008.

See, L. *et al.* (2013) 'Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts', *PLoS ONE*. Edited by T. Preis. Public Library of Science, 8(7), p. e69958. doi: 10.1371/journal.pone.0069958.

Sharma, P. and Joshi, A. (2019) 'Challenges of using big data for humanitarian relief: lessons from the literature', *Journal of Humanitarian Logistics and Supply Chain Management*, 10(4), pp. 423–446. doi: 10.1108/JHLSCM-05-2018-0031.

Shirk, J. L. *et al.* (2012) 'Public participation in scientific research: A framework for deliberate design', *Ecology and Society*. Resilience Alliance, 17(2). doi: 10.5751/ES-04705-170229.

Simmons, B. D. *et al.* (2017) 'Galaxy Zoo: quantitative visual morphological classifications for 48 000 galaxies from CANDELS', *Monthly Notices of the Royal Astronomical Society*, 464(4), pp. 4420–4447. doi: 10.1093/mnras/stw2587.

Simpson, E. *et al.* (2013) 'Dynamic bayesian combination of multiple imperfect classifiers'. Edited by T. Guy, M. Karny, and D. Wolpert. (Intelligent Systems Reference Library), 474, pp. 1–35. doi: 10.1007/978-3-642-36406-8_1.

Simpson, R., Page, K. R. and De Roure, D. (2014) 'Zooniverse: Observing the World's Largest Citizen Science Platform', in *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee (WWW Companion '14), pp. 1049–1054. doi: 10.1145/2567948.2579215.

Skarlatidou, A. and Haklay, M. (eds) (2021) *Geographic Citizen Science Design, Geographic Citizen Science Design*. London: UCL Press. doi: 10.14324/111.9781787356122.

Spiers, H. *et al.* (2019) 'Everyone counts? Design considerations in online citizen science', *Journal of Science Communication*. Sissa Medialab, 18(1), p. A04. doi: 10.22323/2.18010204.

Strandh, V. and Eklund, N. (2018) 'Emergent groups in disaster research: Varieties of scientific observation over time and across studies of nine natural disasters', *Journal of Contingencies and Crisis Management*, 26(3), pp. 329–337. doi: 10.1111/1468-5973.12199.

Swanson, A. *et al.* (2015) 'Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna', *Scientific Data*, 2(1), p. 150026. doi: 10.1038/sdata.2015.26.

Tapia, A. H. and Moore, K. (2014) 'Good Enough is Good Enough: Overcoming Disaster Response Organizations' Slow Social Media Data Adoption', *Computer Supported Cooperative Work: CSCW: An International Journal*. Kluwer Academic Publishers, 23(4–6), pp. 483–512. doi: 10.1007/s10606-014-9206-1.

Taylor, M. ~B. (2005) 'TOPCAT STIL: Starlink Table/VOTable Processing Software', in Shopbell, P., Britton, M., and Ebert, R. (eds) *Astronomical Data Analysis Software and Systems XIV*. (Astronomical Society of the Pacific Conference Series), p. 29.

The ImageMagick Development Team (2021) 'ImageMagick'. Available at: <https://imagemagick.org>.

Turk, C. (2020) 'Any Portal in a Storm? Collaborative and crowdsourced maps in response to Typhoon Yolanda/Haiyan, Philippines', *Journal of Contingencies and Crisis Management*, 28(4), pp. 416–431. doi: 10.1111/1468-5973.12330.

van der Walt, S., Colbert, S. C. and Varoquaux, G. (2011) 'The NumPy Array: A Structure for Efficient Numerical Computation', *Computing in Science Engineering*, 13(2), pp. 22–30. doi: 10.1109/MCSE.2011.37.

Weber, E. and Kané, H. (2020) 'Building Disaster Damage Assessment in Satellite Imagery with Multi-Temporal Fusion', *arXiv preprint arXiv:2004.05525*.

Westrope, C., Banick, R. and Levine, M. (2014) 'Groundtruthing OpenStreetMap building damage assessment', in *Procedia Engineering*. Elsevier Ltd, pp. 29–39. doi: 10.1016/j.proeng.2014.07.035.

Wiggins, A. and Crowston, K. (2011) 'From conservation to crowdsourcing: A typology of citizen science', in *Proceedings of the Annual Hawaii International Conference on System Sciences*. doi: 10.1109/HICSS.2011.207.

Ziemke, J. (2012) 'Crisis Mapping: The Construction of a New Interdisciplinary Field?', *Journal of Map & Geography Libraries*. Jen Ziemke, 8(2), pp. 101–117. doi: 10.1080/15420353.2012.662471.

Zook, M. *et al.* (2010) 'Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake', *World Medical & Health Policy*. Wiley-Blackwell, 2(2), pp. 6–32. doi: 10.2202/1948-4682.1069.