Analysing volunteer engagement in humanitarian mapping: Building contributor communities at large scale

Publisher:

Association for Computing Machinery

Published version:

DOI:10.1145/2818048.2819939

(Article begins on next page)

Analysing Volunteer Engagement in Humanitarian Mapping: Building Contributor Communities at Large Scale

Martin Dittus
ICRI Cities, UCL
London, UK
martin@dekstop.de

Giovanni Quattrone
Department of Computer Science, UCL
London, UK
g.quattrone@cs.ucl.ac.uk

Licia Capra
Department of Computer Science, UCL
London, UK
l.capra@ucl.ac.uk

ABSTRACT
Organisers of large-scale crowdsourcing initiatives need to consider how to produce outcomes with their projects, but also how to build volunteer capacity. The initial project experience of contributors plays an important role in this, particularly when the contribution process requires some degree of expertise. We propose three analytical dimensions to assess first-time contributor engagement based on readily available public data: cohort analysis, task analysis, and observation of contributor performance. We apply these to a large-scale study of remote mapping activities coordinated by the Humanitarian OpenStreetMap Team, a global volunteer effort with thousands of contributors. Our study shows that different coordination practices can have a marked impact on contributor retention, and that complex task designs can be a deterrent for certain contributor groups. We close by providing recommendations about how to build and sustain volunteer capacity in these and comparable crowdsourcing systems.

Author Keywords
Crowdsourcing; Peer Production; Social Computing; Retention; Engagement; Task Design; Task Analysis

ACM Classification Keywords
H.5.3. Group and Organization Interfaces: Computer-supported cooperative work; Design

INTRODUCTION
The Humanitarian OpenStreetMap Team (HOT) aims to map all the undocumented and crisis-stricken regions of the world. The formidable scale of this ambition was illustrated during the 2014 Ebola epidemic: even after months of work by thousands of volunteers, the new maps of Central and West Africa are still not complete. An article by Médecins Sans Frontières (MSF) suggests that to reach their goal, HOT organisers need to grow their project to "the biggest instance of digital volunteerism the world has ever seen" [11]. Organisers thus not only need to consider how to produce these maps, but also how to foster a large global volunteer community in the process.

CSCW '16, February 27-March 02, 2016, San Francisco, CA, USA.

ACM 978-1-4503-3592-8/16/02...$15.00.
DOI: http://dx.doi.org/10.1145/2818048.2819939

The HOT projects presented in this study represent two aspects of this community-building challenge: disaster aid initiatives need to build volunteer capacity to provide quick emergency response, and disaster preparedness initiatives need to sustain volunteer capacity in the absence of urgent causes. While organisers have significant freedom in designing these projects, it is not clear how to they can evaluate their choices in these regards. Furthermore it is not always clear whether certain design choices involve trade-offs.

Other studies have already assessed the quality of HOT outputs, and their impact on the map [9, 34]. This study will instead focus entirely on engagement aspects: the existence of HOT presents a rare opportunity to compare different coordination practices within the same platform, involving a large number of projects and participants.

Proposed contributions
The present study is focused on a key growth challenge: to develop understanding of how best to increase volunteer capacity. Our research takes the form of a large-scale quantitative observational study. We evaluate whether individual projects can successfully activate new volunteers (*enrolment*), but importantly also retain them over time (*retention*). Together we define these as *engagement*.

A range of HOT initiatives and organisational practices offer many opportunities to evaluate specific organiser choices. We aim to assess a large number of participations in a consistent manner. To this purpose we propose three analytical dimensions: *cohort analysis* where we compare collections of similar projects, *task analysis* where we compare projects in their task complexity, and observation of *contributor performance* relating to the rate of contributions. All rely on readily available public data, and we will demonstrate that they can yield important findings.

The analytical dimensions we propose are grounded in existing theory, and have direct operational implications so that findings can be translated to organisational change. They provide minimum-effort complements to more invasive evaluation practices such as controlled experiments, A/B tests and participant observations. They are general enough to be transferrable to other online communities: their minimum requirement is a capacity to observe individual contributions over time.

A key aspect of this evaluation is our focus on first-time contributors, starting from their initial enrolment. In these first hours and days, we can expect that engagement is at least in part shaped by the specific design of this first project. We can assess whether the initial experience was so discouraging that contributors never returned, or whether it prepared them to contribute for longer periods. A more experienced contributor on the other hand may still be able to contribute to a badly designed task, which would make it harder to identify problematic design choices.

On the following pages we first present three research questions motivated by the growth challenge we presented. We then provide an introduction to HOT and its practices, describe key project initiatives, present an overview of related work, and introduce our methodology. Finally we address our research questions with a set of analyses based on contributor engagement metrics, and close with a discussion of our findings, and a brief outline of future work.

RESEARCH QUESTIONS

RQ1: Cohort analysis

Are different coordination practices associated with different contributor engagement characteristics? In large distributed online communities we are likely to find subgroups with divergent practices. In the case of HOT these are the different humanitarian causes within the same platform.

We expect that divergent coordination practices and circumstances can affect contributor engagement in different ways. In particular we believe that the perceived urgency of a cause can act as an attractor and lead to higher enrolment. On the other hand we expect that sustained promotion of needs can likely increase retention, however it is unclear how well this actually works and how long it can last.

RQ2: Task analysis

Are different task designs associated with different contributor engagement characteristics? Task analysis serves to assess the impact that contribution mechanics can have on the initial enrolment experience. It allows to distinguish between different tasks in terms of their task complexity.

We expect that a minimum degree of task complexity is needed to yield substantial contributions, however there may be diminishing returns: more complex work can be discouraging. This may be particularly true for newcomers, who may choose to abandon their participation early.

We further expect that such effects can be addressed with better contributor guidance. A gradual learning curve and other forms of guidance can help newcomers build a sense of self-efficacy. However we also expect diminishing reports in this context: too much documentation can be overwhelming as well.

RQ3: Contributor performance

Is contributor performance an early indicator of retention? A volunteer's rate of contributions, or edit pace, can be considered a key measure of their performance. Do faster contributors tend to stay longer? Can we observe performance improvements during the initial period of participation? Are these measures associated with particular retention profiles?

We expect that faster contributors tend to stay longer, they may be more confident in their abilities. We further expect to see that newcomers start slowly, but then pick up their pace. We further expect that any observed performance improvements are associated with an increase in short-term retention, as a result of increased contributor self-efficacy and enjoyment.

VOLUNTEER MAPPING WITH HOT AND OSM

HOT was founded in 2010 by an informal network of experts, and the organisation has gradually refined the necessary processes and technologies that allow it to scale [30]. HOT organisers coordinated responses to typhoon Haiyan in 2013, the West African Ebola crisis in 2014, the 2015 Nepal earthquake, and many other activations [32]. The promise of these initiatives is to achieve a greater volunteer capacity for disaster response and humanitarian aid by incorporating the help of a global online volunteer force. The MSF staff member Ivan Gayton summarises the particular appeal HOT has for representatives of aid organisations: "Finally, I can give volunteers something to do that isn't just giving money" [18].

HOT organisers set up projects to address particular information needs, and promote these to potential volunteers. Projects seek to map certain geographic features in a particular region, for example to map settlements so that aid experts can understand population distributions, or to trace roads so that field teams can plan transport routes. Larger HOT activations can consist of dozens of individual projects.

The maps produced by HOT volunteers are free for all under a liberal license, and are now in use by experts at MSF, the American and British Red Cross, the World Health Organisation, and a growing number of other institutions [7]. A 2014 report by MSF discusses the impact such initiatives can have on the work of aid organisations: "Many interviewees commented that they were 'amazed' by the speed at which the area was mapped with the help of the volunteers. On his own, the GIS officer would not have been able to produce these base maps during his mission." [17]. Figure 1 visualises the global distribution of all HOT edits.

Remote mapping with the Tasking Manager

Typically the first step in the creation of a new map is a remote mapping practice involving the help of hundreds or even thousands of volunteers. In remote mapping, HOT volunteers trace maps from satellite imagery of remote places. In some cases, regional groups host mapathons to come together in a more social setting, but many contributors simply participate online.

The HOT Tasking Manager is a key technology in the contribution process, it emerged out of a need to distribute work across large numbers of remote mappers while reducing edit conflicts [23]. The Tasking Manager is also a rich source of contextual information: it publishes a list of all remote mapping projects, task descriptions, boundary polygons for project regions, and contributor lists.

Figure 2 shows a screenshot of a project and its description. Within all Tasking Manager projects, work is divided spatially

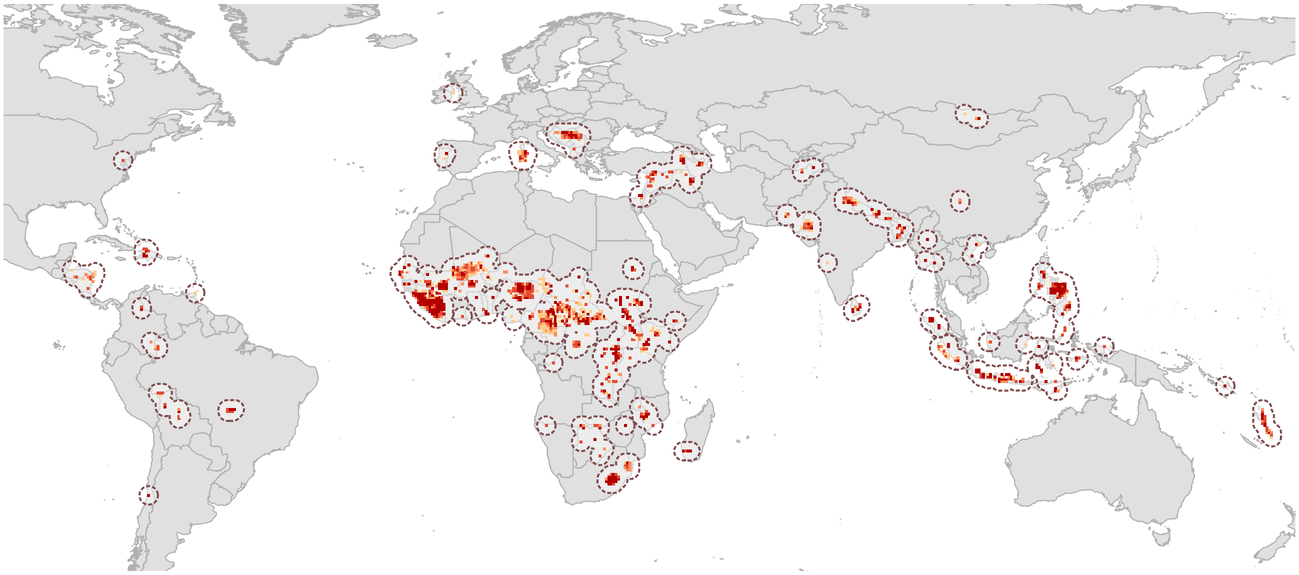

Figure 1. Geographic distribution of all HOT contributions. Activities focus on Central and West Africa, Southeast Asia, and other parts of the world.

into smaller tasks (map grid cells). Contributors choose a free task in the Tasking Manager when they begin their work, and then contribute according to a set of instructions.

All contributions are made on OSM using OSM mapping tools, and HOT contributors require an OSM account to contribute. Most of their time is spent in these editing tools rather than the Tasking Manager. The OSM changesets of their contributions are now automatically annotated with HOT-specific tags, which means HOT contributions can be identified in the OSM edit history. Further background on OSM and its relationship to HOT is provided by Soden et al [30].

In subsequent stages, local mappers with knowledge of the respective area may augment the maps by adding more detailed annotations, including place names and other details, and by making corrections based on ground observations [31].

For this study we focus on the remote mapping process, and we will not assess the contributions of local mappers: there is less less readily available data on the circumstances and outcomes of their practices, and these subsequent activities on the ground are not as easily observed at scale. As a result a large-scale quantitative evaluation of their work is harder to achieve. Their engagement characteristics are also likely different from those of remote mappers: they involve a much smaller number of contributors who are often more closely embedded with local communities or aid organisations.

RELATED WORK

A key barrier to entry for first-time HOT contributors is the fact that mapping with OSM is a complex practice: it requires specialist tools and an understanding of specialist concepts [29]. To our knowledge there is no published research on how this affects HOT contributor engagement, however there is some knowledge in related domains.

[1]http://tasks.hotosm.org/project/870

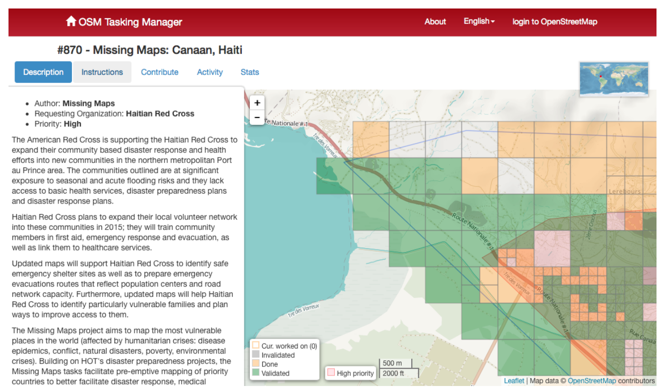

Figure 2. The HOT Tasking Manager, showing a project description on the left hand side, and a map with a task grid on the right.[1]

Crowdsourcing is the solicitation of labour from a large group of participants, typically online, accomplishing a variety of tasks such as the creation of content or solving of problems [13]. This includes forms of citizen science where a coordinating party invites scientific contributions by laypeople [35]. Crowdsourcing systems in which participants are motivated by payment and similar incentives are sometimes distinguished as *crowd work* [16].

In crowdsourcing and crowd work systems, organisers (or "requesters") prepare tasks that are then offered to contributors. Organisers need to strike a balance between organisational performance and worker satisfaction: badly designed tasks can be deterrents for participation [35]. A basic design strategy is to split the work into smaller pieces that are manageable by a single contributor in a limited amount of time.

Kittur et al. further propose to communicate more clearly with contributors, but also to better support their learning experience. “Workers may need to acquire new skills to perform unfamiliar tasks, before or in the midst of performing the actual work” [16]. In an evaluation of Mechanical Turk crowd workers, Khanna et al. found that more complex tasks that require a nuanced understanding of the domain can pose barriers to participation. Other barriers included problems related to the user interface, or misunderstandings derived from differences in cultural contexts between coordinator and contributor [15].

There is some evidence that increased activity and increased retention may not always be achievable at the same time. In online citizen science projects it was found that more prolific contributors can have shorter retention periods [25, 26]. A similar effect was found for crowdsourcing designs that aim to increase member productivity, and was attributed to either burnout or a sense of a “mission accomplished” [33].

A basic psychological model that allows us to reason about the nature of engagement barriers is the framework of self-efficacy [1]. According to this model, perceived human efficacy determines if an individual will initiate an activity, how much effort will be expended, and how long the activity will be sustained. Self-efficacy is derived from four principal information sources: performance accomplishments, vicarious experience, verbal persuasion, and physiological states. In other words, task designs that aim to increase self-efficacy must consider how best to bring about experiences of mastery, and to not overwhelm prematurely. This can at least in part also involve social processes: persuasion, observation of others, and feedback.

When designing tasks, one should also consider the motivations of contributors to participate. A general model of motivations of volunteer workers was presented by Clary et al. who describe six basic motivational categories, including values (such as altruism), social experience, and enhancement (self-improvement, a positive self-image) [8]. In a study of Wikipedia contributor motivations, this was later amended by two further categories: fun, and ideology [19]. Among early OSM contributors, Budhathoki found that an individual’s local geographic knowledge was the most significant driver to contribute [3]. In the context of charitable work it was further shown that the presence of a community of participation can affect a person’s willingness to contribute to charitable organisations [28], as does the perceived degree of social urgency behind a particular cause.

HOT can also be considered an example of commons-based peer production [2]. It emerged out of the activities of the OSM community, and is shaped by its technologies and practices: in principle anyone can contribute, and contributors choose their own tasks freely [23]. There is a substantial amount of research on peer production systems, with Wikipedia and OSM as prominent examples. However in such systems there typically is no clear distinction between organiser and contributor, while in HOT a central committee coordinates activities with aid organisations. As a result, work in HOT may be organised differently than in OSM or Wikipedia, and contribution processes tend to be more formal and goal-oriented.

Research contribution

We present the first comparative study of HOT activity across multiple large initiatives, and likely the first large-scale study of HOT community engagement. Despite increased research interest in the topic, to our knowledge there has been no published research on the ability of different HOT project designs to build volunteer capacity, and then successfully retain trained contributors over longer periods.

METHODOLOGY

Data

All our analyses are based on two data sets:

1. Project information published on the HOT Tasking Manager.[2] This data was scraped for every project.
2. The OSM edit history of all map contributions, recording the creation and modification of map objects over time. This dataset is freely available for download.[3]

Using contributor lists from the Tasking Manager as a starting point, we extracted the OSM map contributions by all known HOT participants and cross-referenced them with HOT projects based on username, date, and location. For the purpose of this study, any creation or modification of a map object is considered an edit.

Study period

We selected an 18-month enrolment window from mid-2013 to late 2014 to observe first-time contributions, further extended by a buffer period of 180 days for the observation of contributor retention. This time captures the most active period of the Tasking Manager history to date. It excludes an initial early-adopter period, but includes the first Tasking Manager use at a larger scale in late 2013 [23].

Figure 3 shows the remarkable growth of HOT remote mapping activity in this time: by early 2015, HOT organisers had created almost 1,000 remote mapping projects. Our study period is highlighted in the graph.

The specific timeframe, all dates are inclusive:

- First date of enrolment window: 16th of June 2013
- Last date of enrolment window: 15th of December 2014
- Plus 180-day buffer period: ends 16th of June 2015

Cohort selection

For this first study of HOT engagement we limited our evaluation to key remote mapping initiatives: larger collections of Tasking Manager projects with a shared singular purpose and shared organisational practices. We identified such collections based on project listings on the HOT homepage and the OSM wiki. Approximately 50% of HOT projects could not easily be allocated to a group.

In order to find collections that are suitable for both cohort analysis and task analysis we further rejected initiatives that involved less than 50 contributors, and less than 10 projects.

[2] http://tasks.hotosm.org

[3] http://planet.osm.org/planet/full-history/

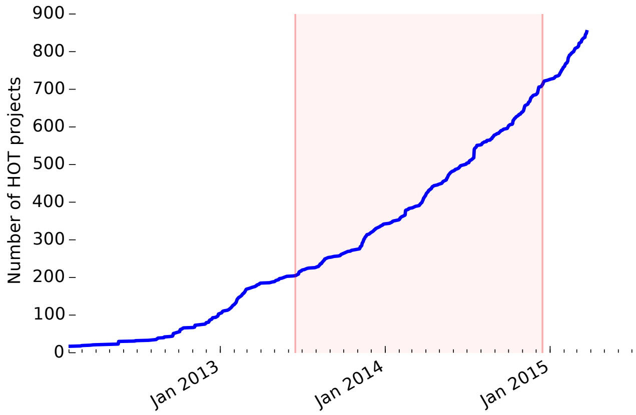

Figure 3. Growth of HOT Tasking Manager projects over time. Our study period is highlighted.

Cohort	*Projects*	*First-time contributors*
TH	23	481
ER	65	881
MM	11	208
Total	**99**	**1,570**

Table 1. Project cohort sizes.

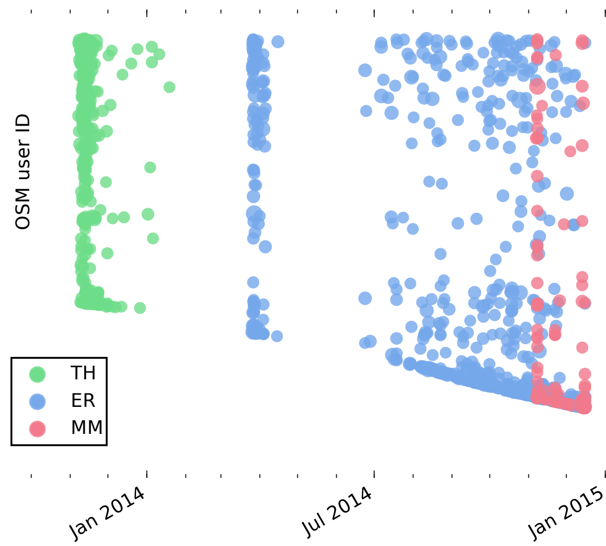

Figure 4. Timeline of contributor enrolment, broken out per cohort. Contributors are arranged vertically in order of ascending user ID.

We identified six larger initiatives as candidates. Of those we chose three as study cohorts based on their profiles: they represent a cross-section of typical HOT coordination practices, from urgent responsive mapping to sustained proactive mapping. Their aggregate size is substantial: they encompass 30% of all projects on the Tasking Manager, and account for almost 50% of all first-time contributors in the period.

Typhoon Haiyan (cohort: TH)

Also known as Typhoon Yolanda, TH was a tropical cyclone that devastated the Philippines on November 8, 2013. The HOT projects associated with it were of high urgency: the map data was used in disaster response and humanitarian aid as soon as it became available. TH was probably the first highly promoted HOT initiative to rely on the Tasking Manager for coordination. Volunteer work started only few days before the event, initially with projects to prepare a base map. The focus later switched to damage assessment of the affected areas. The OSM wiki lists 22 mapathons which were organised around TH in November 2013 in cities around the world. Typically these were one-off events [22].

Ebola response (cohort: ER)

The Ebola outbreak began in Guinea in early 2014. Aid organisations needed maps to locate and treat those infected, yet many of the affected areas were not documented on any existing maps. Initially this was treated as an urgent one-off event, and activity stopped soon. However coordination was picked up again as the epidemic spread to neighbouring countries, and the strategy changed to a more sustained effort covering larger regions. The initiative lasted until early 2015. Mapathons were organised in many parts of the world to train newcomers and coordinate volunteers, including monthly events in several cities [20]. Project descriptions indicate that many activities were coordinated by representatives of aid organisations.

Missing Maps (cohort: MM)

This effort launched in November 2014: regions vulnerable to crises are mapped early so that maps are already available when a crisis occurs [27]. In contrast to TH, this initiative is proactive rather than reactive, and continuity of contributor engagement is more important than a quick response. Where TH and ER were more ad hoc in their coordination, MM organisers set up structures to support more sustained engagement: there are regular mapathons in many parts of the world, often organised as monthly events. A signup form for volunteers allows contributors to get informed of upcoming initiatives and mapathons [21]. Similar to ER, many of the MM projects were initiated by representatives of aid organisations.

We selected contributions relating to each cohort in the study period, identifying instances where new first-time contributors joined a project. The final data set derived from this is summarised in Table 1, with a total of 1,570 first-time contributors across 99 projects.

A timeline of contributor enrolments per cohort is shown in Figure 4, this plot illustrates a number of key aspects of our data set. The cohorts were active at different periods: TH in late 2013, ER throughout much of 2014, and MM from late 2014 onwards. The contributor distribution also indicates that many first-time contributors already had existing OSM user accounts, as indicated by their user IDs. The plot also reveals a participation gap for ER between May and July 2014, this reflects the period of inactivity between its initial conclusion in early 2014, and a subsequent resurgence of activity later in the summer.

Experience d_{pre}	0 days	1-9 days	10-99 days	≥100 days
TH	30.9%	22.2%	20.3%	26.6%
ER	52.8%	24.3%	11.7%	11.2%
MM	72.8%	18.4%	5.3%	3.4%

Table 2. Distribution of prior OSM experience per cohort, in days with contributions (d_{pre}).

Contributor observation periods

We collected all contributions by first-time contributors in the first 180 days after their initial contribution. We paid particular attention to three different timeframes per contributor:

- The initial enrolment period of the first 48 hours. This period is used to observe initial contributions, and to assess how many contributors returned immediately on the second day. The short initial observation period of two days was chosen based on the median difference between the first and last moment of contribution of all first-time contributors, which is only 20 hours.
- A 90-day retention period from the moment of enrolment. This was chosen based on an analysis of average contributor lifetimes: almost 90% of contributors cease contributing after 90 days of their initial participation.
- A 180-day survival period from the moment of enrolment, to identify the last known moment of contribution. This was used for survival analysis: we considered contributors 'dead' if they had been inactive for at least 90 days by the end of this survival period. The last known date of contribution before that point marks their 'death event'.

In summary this means that we only consider the first 90 days of contribution activity for our analyses, however we observe for a full 180 days after enrolment to establish abandonment with some degree of certainty.

Engagement metrics

Based on these relative observation windows we computed engagement metrics for every first-time contributor we identified. We computed contribution sessions based on the timestamps of individual contributions, with a session timeout of one hour. Based on these we computed the number of labour hours spent on each contribution, using a process described by Geiger et al in the context of Wikipedia contributor analysis [12].

A first set of engagement metrics are *measures of activity in the enrolment period*. We chose time-based activity measures rather than simple edit counts because they allow us to more meaningfully compare contributor effort across different kinds of tasks [12]. We seek to quantify the amount of time spent contributing, rather than the volume of data that was produced, so we can compare contributor effort across different map object types.

More specifically, we captured the number of labour hours l_{48h} in the first 48 hours. These are also calculated separately for the first and second day in this initial period: l_{d1}, l_{d2}. We further capture the rate of contributions in the firth 48 hours c_{48h}, measured in edits per hour, and contribution rates for the first and second day c_{d1} and c_{d2}. These allow us to determine a change in pace between the first and second day to test for the presence of performance improvements. This change in pace is described by the ratio c_{d2}/c_{d1}.

A second set of metrics are *measures of retention*. These are calculated per project: what share of first-time HOT contributors could later be retained for further activities on any HOT project? To quantify short-term retention per project we determine R_{d2}, the percentage of first-time contributors that are still active on the second day of their participation. Additionally we calculate long-term retention metrics R_{m2} and R_{m3}, the retention rates in the second and third month after enrolment. These 30-day periods were chosen to reflect monthly mapathon cycles observed by some HOT initiatives.

Quantifying prior domain experience

We further quantified each contributor's *degree of prior OSM experience* as d_{pre}, the number of days on which they had contributed to OSM before they joined their first HOT project. In certain analyses we used this measure as a control variable: contributors with prior OSM experience may find it easier to contribute to HOT.

In an initial assessment of the impact of prior experience we correlated d_{pre} with our engagement measures. We found that less experienced users contribute for less hours during their enrolment (Spearman correlation coefficient $\rho_S = 0.19$), that they contribute at a lower pace ($\rho_S = 0.23$), and that they are retained less often than others in the second month ($\rho_S = 0.12$) and third month ($\rho_S = 0.13$, all with $p < 0.001$).

Comparison across cohorts shows that the three groups have different constituencies: the most experienced group was TH (median: 5 days of prior OSM contributions, mean: 113 days), the most inexperienced group was MM (median: 0 days, mean: 23 days), and ER was in between (median: 0 days, mean: 52 days). Table 2 visualises the distributions as a histogram. Pairwise comparison of these distributions with Kolmogorov-Smirnov statistic (KS) confirmed that these are statistically different populations ($p < 0.001$), with the greatest difference measured between TH and MM (KS statistic: $\alpha = 0.41$).

Task analysis

We rely on the fundamental assumption that some tasks are more challenging than others, and that this affects contributor engagement. We expect that we can identify this effect by observing participation across different tasks. We specifically aim to assess how easy it is for a newcomer to become a productive contributor.

In "Task Complexity: Definition of the Construct", Wood distinguishes between the complexity of the required work itself, and the amount of information cues and guidance necessary to

Aspect	Variable	Description
Motivation	*has_context*	Does the project description state an explicit purpose?
Visual complexity	*urban_density*	Is the mapped region rural (simple), mixed, or urban (complex)?
Task complexity	*num_concepts*	How many different types of map objects are to be mapped?
Task complexity	*building_trace*	Are buildings to be mapped as points (simple) or polygons (complex)?
Guidance	*num_cues*	Number of information cues provided in the documentation?
Guidance	*num_tag_ex*	Number of tag examples listed?

Table 3. Task design feature vector produced by our task analysis.

produce it [36]. These can be quantified as the number of acts and number of information cues involved in the work. Both are measures of task complexity: simple tasks require processing fewer cues than complex tasks [24]. They reflect the consideration that task designs must consider how best to bring about experiences of mastery without being overwhelming.

Table 3 lists the six task design features we considered for this study, including motivational factors, visual complexity of satellite imagery, task complexity, and forms of guidance. These were labelled by the first author in an iterative process, involving a detailed study of a large number of projects. We categorised all requested map features by their geometries and semantic role: natural features, roads and highways, settlement boundaries, buildings, urban infrastructure, and other features. We similarly classified and counted the distinct number of information cues per project: stated priorities, descriptions of map object types, explicit sequences of work steps, and external links to coordination pages, reference documents, or instruction manuals.

These features were then used to compute task design feature vectors for each project as described in Table 3. We standardised these feature vectors using z-scores, so that all variables have a mean of 0 and a standard deviation of 1. We found no multicollinearity across the standardised variables: the condition index of all variables is 2.33, and we found no near-zero eigenvalues in their cross-correlation matrix.

FINDINGS

RQ1: Cohort analysis

Table 4 shows median engagement statistics per cohort for the enrolment period of the first 48 hours. TH and ER have similar enrolment profiles, while MM contributors appeared to work longer but also contributed more slowly than the other cohorts. We confirmed the pairwise difference of these distributions with a KS statistic: with one exception these differences in distribution were statistically significant ($p < 0.01$). The only instance where a difference could not be asserted were the distribution of labour hours in the TH and ER cohorts, however their rate of contributions differed.

Table 5 shows the corresponding median retention rates per cohort, broken out for short- and long-term retention periods. These measures indicate that TH and ER have the highest short-term retention, with almost 30% of contributors returning on day 2. However the long-term retention rates of TH are the lowest of all cohorts. In contrast MM starts with the lowest short-term retention of only 10% returning on second day, followed by the most stable long-term retention of all cohorts: almost 9% contributors are still retained by month three. We computed survival functions for each cohort based on a Kaplan-Meier estimate with a 95% confidence interval. A pairwise logrank test confirmed statistically significant differences in survival rates between ER and TH ($p < 0.001$), no other pairing was found significant.

Cohort	l_{48h}	c_{48h}
TH	1.14	640.4
ER	1.16	620.0
MM	1.29	529.2

Table 4. Median contribution activity by cohort: labour hours (l_{48h}) and contribution rate (c_{48h}) in the first 48 hours.

Cohort	R_{d2}	R_{m2}	R_{m3}
TH	26.8%	4.2%	4.6%
ER	27.2%	13.6%	8.9%
MM	10.1%	9.6%	8.7%

Table 5. Median retention for day 2 (R_{d2}), and months 2 and 3 (R_{m2} and R_{m3}). These indicate the percentage of HOT contributors who are still active in the respective period.

When limiting survival analysis to OSM newcomers only, where the first HOT contribution was also the first OSM edit, we observed a statistically significant difference in survival rates between TH and MM, as well as TH and ER (both $p < 0.001$). The corresponding survival plot for OSM newcomers in Figure 5 illustrates this: TH has the lowest overall retention. ER and MM differ in the short term, with MM having lower initial retention, yet then achieve similar long-term retention characteristics. According to the plot MM may even have the highest long-term retention, yet this difference was not confirmed by the logrank test.

RQ2: Task analysis

We prepared a regression model to observe the impact of the six task design features on early abandonment, measured in labour hours l_{48h} as the dependent. We included two control variables: prior user experience d_{pre}, and the size of the project (its number of tasks) to account for goal-setting effects with larger projects.

Regression analysis was performed on all data, and for each cohort separately. The only significant model was found for the MM cohort (using ordinary least squares: adjusted R^2=0.37, F-statistic of 18.72, with $p < 0.001$ at 7 degrees of freedom).

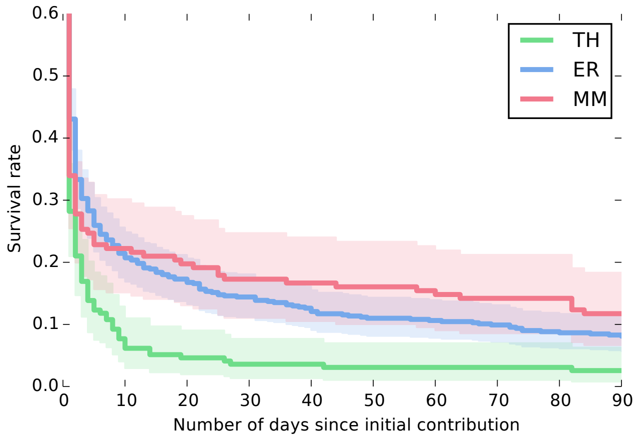

Figure 5. Survival functions for each cohort, indicating the rate at which participants with no prior OSM experience ceased contributing. Shaded regions show 95% confidence intervals.

According to this model, individuals contributed for longer periods if they had prior OSM experience ($\beta_{d_{pre}}$=0.5), and were given fewer map object types to map ($\beta_{num\_concepts}$=-0.15, with a 95% confidence interval between -0.232 and -0.064).

These coefficients are based on standardised z-scores. The negative coefficient $\beta_{num\_concepts}$ indicates that the strongest response is below the mean value of $num\_concepts$, which before standardisation is at 2.7. This result suggests that people remained active for longer on tasks that involved the mapping of less than three distinct map features.

Other models yielded no improvements in fit. This includes models that only included OSM newcomers, or only involved the two significant features d_{pre} and $num\_concepts$. The contribution rate c_{48h} was not explained by any regression model.

RQ3: Contributor performance

We found that faster contributors tend to remain active for longer: a correlation analysis found that the initial contribution rate c_{d1} is associated with longer participation in the enrolment period (Spearman correlation coefficient $\rho_S = 0.15$), and with increased retention in month 2 ($\rho_S = 0.15$, both with $p < 0.01$), however this effect is slightly reduced in month 3 ($\rho_S = 0.11$, $p < 0.05$).

During a further analysis we found that contributors who abandoned tasks early and those who stayed the longest tended to contribute more quickly than those in between. We segmented contributors into three engagement classes based on their initial labour hours l_{48h}, this variable follows a normal distribution. Segmentation at the 25th and 75th percentile yields three groups of short-term (under 30 minutes), average, and long-term contributors (2 hours or more). Figure 6 shows the distribution of contribution rates for each engagement class. Short-term contributors tended to contribute at a faster pace: their median contribution rate (650 edits per hour) was almost 30% higher than that of the average group (515 edits per hour). This effect was found in all cohorts. A pairwise comparison

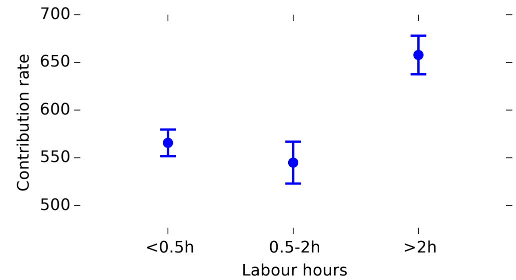

Figure 6. Mean contribution rate (c_{48h}) segmented by contributor engagement in the first 48h, with 95% confidence intervals.

across the three bins with a KS statistic over the respective distributions of l_{48h} was significant for all pairs ($p < 0.001$).

On the other hand it cannot be said that there always are performance improvements within the first sessions: the median change in pace is around 0% between the first and second day of enrolment across cohorts. We found no correlation between performance improvements and long-term retention.

DISCUSSION

In the following section we will summarise our results by highlighting the key effects we observed, and close with a set of implications and design recommendations.

The three cohorts have different volunteer constituencies. A comparison of their distributions of prior experience revealed markedly different distributions. This may be a result of a difference in engagement strategies: TH required a quick response by an existing community, while ER and MM could build volunteer capacity in a more sustained manner over longer periods. Furthermore they were active at different periods: by the time ER and MM were active, HOT had already gained some prominence outside of the OSM community.

The three cohorts have different contribution profiles during enrolment. The initial contribution profiles of TH and ER are quite similar, although TH contributors contribute at a slightly higher pace. In contrast, MM volunteers have the lowest rate of contributions, yet were working longer hours during initial enrolment than the other cohorts. The lower pace may be an effect of relative inexperience, while increased initial activity may indicate a greater motivation to contribute.

Coordination practices have a marked impact on contributor retention. Long-term retention was highest for the ER and MM cohorts which were specifically set up as more sustained initiatives, and which relied on a range of volunteer engagement practices such as mapathons and social media use. It was further found that for the MM cohort, short-term retention was lowest yet long-term retention was highest: contributors do not tend to come back on the second day, however they are more likely to remain active a month or two later. This may indicate a greater reliance on mapathons, where contributors do not necessarily return a day later, but at the time of the next mapathon event. In contrast to this, TH had a large number of contributors yet the lowest retention rates, likely because of an absence of any such sustained engagement practices. Its contributors may have been attracted by a perceived urgency around the disaster event, however they were not successfully retained for any subsequent HOT activity.

Complex task designs can be a deterrent for certain contributor groups: projects involving three or more map object types saw shorter activity periods for the MM cohort. In other words, more complex task requirements may be demotivating to first-time contributors, regardless of their prior OSM experience. We found no evidence of an impact of documentation and guidance on engagement.

Most first-time HOT contributors tend to operate at a fairly steady pace, however contributors with prior OSM experience tend to work faster and stay a little longer. This effect is consistent across cohorts.

Early abandonment is associated with higher contribution rates. Volunteers who stopped contributing within the first 30 minutes tended to contribute at a slightly higher pace than those who stayed an hour longer. This may suggests an instance of a burnout effect, where some first-time contributors begin their engagement at a relatively high pace but then lose motivation quickly. The effect was found in all cohorts.

Performance improvements are not associated with increased retention: contributors whose performance increased within the enrolment period were not necessarily retained for longer, suggesting that the presence of performance improvements is not an early indicator of increased long-term engagement.

Implications

The aim of our study is to understand engagement factors in crowdsourcing communities that help organisers reason about how to build and retain volunteer capacity. We believe we have identified some key aspects that can inform the design choices of HOT and related large-scale crowdsourcing systems.

Our findings suggest that the capacity-building strategies of the ER and MM cohorts work well, and we encourage organisers of other initiatives to adopt their practices: a combination of highly promoted projects over a sustained period, a steady stream of new efforts, regular mapathons and other training environments, and the use of email notifications in the MM cohort as a means of notifying interested contributors of new causes. In these two cohorts, a larger share of contributors kept coming back. In contrast to this, the greater urgency of the TH initiative (a response to a discrete crisis event) may have attracted many volunteers, however it did not contribute to an increase in retention.

These retention effects may further be affected by the marked difference in cohort constituents, attributable to self-selection effects and the presence of existing social ties. According to project descriptions, both ER and MM initiatives were largely coordinated by representatives of aid organisations, while TH projects were coordinated by OSM community members. This may have affected how the initiatives were promoted, and to whom. It is possible that ER and MM participants already had existing connections to aid organisations as part of other outreach efforts, and as a result had a higher motivation to remain engaged.

With one exception, the task designs encountered in our study were found to be remarkably consistent in their engagement characteristics. One cohort saw a reduction in activity during the enrolment period for tasks that requested contributions of a higher complexity, which suggests that organisers should limit the number of map feature types requested per project.

This finding of relatively consistent contributor performance across designs may also indicate a limitation of our observational study: we do not compare radically different task designs, and instead merely observe existing tasks which rely on the same tools and interfaces for the actual contribution process. Additionally, many of these designs have already been informed by years of prior experience [23]. Consequently we encourage HOT organisers to also experiment with new task designs to identify alternative strategies for further improvement. An example of this could be a micro-tasking interface that offers smaller and simpler tasks, which could allow newcomers to become productive more easily and quickly.

An alternative interpretation is that current HOT contribution mechanics do not have a major impact on engagement, and that other aspects may be more important. In particular our review of prior work suggests that intrinsic participant motivations, participant enjoyment, association with particular humanitarian causes, and social aspects of the HOT contribution experience may be more important to contributor engagement than the specifics of HOT contribution mechanics.

A further conclusion from our findings is that the best means of increasing output capacity is to grow the volunteer base, particularly considering the vast amount of uncharted territory that HOT aims to map. We could observe improvements in the performance of individual contributors, however an increase in contributors would raise output capacity more quickly.

It remains open what constitutes good training conditions for absolute newcomers. We believe that given a choice, newcomers are best placed in projects where they have a higher likelihood of being retained. In our case this would be the ER and particularly MM cohorts: projects that are specifically set up as long-term initiatives. Additionally there are indications that particularly the MM cohort was successful at retaining and training absolute newcomers with no prior OSM experience.

CONCLUSION

We presented an observational study to assess the relationship between project designs and contributor engagement, with a specific focus on the experience of first-time contributors. We compared project designs along three analytical dimensions: cohort analysis, task analysis, and observation of contributor performance. Under consideration of these aspects we evaluated different projects in their ability to foster contributor engagement in the short and long term. The analytical dimensions yielded plausible findings: we found that different coordination practices did have a marked engagement impact, and that differences in task design can have an impact for some groups. Additionally, we found that prior domain experience in first-time contributors is likely to increase their engagement.

We believe that these findings have some external validity: we encounter similar contributor engagement across different cohorts, but also some differences, and we believe that most of these differences have been explained by the observable factors discussed above. For this reason, we suggest that our findings are transferrable to other crowdsourcing systems.

The nature of this work and the available evidence however place limits on our ability to build more nuanced understanding, and there are many aspects which we cannot gauge from the available data. For example the contributor context is generally not known: in which situation does mapping take place? Yet the large scale of this study allow us to identify some general engagement trends that tend to be shared across different kinds of projects.

Future work

Our findings suggest that the participation setting may play an important role in contributor engagement: mapathons and other social learning environments can play an important role in contributor onboarding. Although there is some existing knowledge on these effects, we believe that their impact warrants further study. Some early observations were made by Hristova et al. in a study of OSM mapping parties. The authors found that participants sustained engagement even after the event, however this effect was limited to contributors with at least some prior experience, and newcomers to OSM could not be activated by these settings [14]. Other studies have found evidence that the socialisation experience of first-time contributors in online communities can increase their contributions and long-term retention. In a recent observational study of Wikipedia contributors, Ciampaglia et al. find that a successful early socialisation experience is associated with and can sometimes predict increased contributor engagement, however it was also found that the causal structure between socialisation, motivation, and participation is not entirely clear [6]. Further studies identified similar effects [10, 4, 5].

Furthermore our analysis of task complexity could be augmented with research to understand the actual contribution process, such as participant observation, usability studies, and ethnographic studies. Such methods could asses which forms of guidance are most useful to newcomers, to what extent volunteers spend time reading task instructions before they start contributing, and related aspects of the contribution process.

We have not studied the impact of project designs on output *quality* because it is a separate concern from that of contributor engagement, but also since it would be significantly harder to measure. These maps are typically the first ever of their kind, which means there is no ground truth data, and a manual quality assessment would require substantial effort at this scale. Additionally the mapped geographies tend to be different across initiatives, which makes it hard to develop intrinsic quality measures that have validity across different regions of the world. HOT does have its own validation process, however it is highly informal and not sufficiently consistent for a rigorous evaluation. Yet there is an opportunity for further studies to assess how coordination practices affect the quality of contributions, and whether there are interactions between quality concerns and contributor engagement.

This work could further be augmented with a broader understanding of the contributor experience, by means of participant observation or surveys. In particular there is currently no published knowledge on the specific motivations of HOT contributors, but also on the role of participant enjoyment, contributor interactions, and user interfaces, each worthy of a study of their own.

ACKNOWLEDGMENTS

This research was funded by the EPSRC and the Intel Collaborative Research Institute: Cities. We are grateful to Andrew Braye, Pete Masters, and countless others in the HOT community for their support. We would further like to thank the reviewers for their feedback and suggestions.

REFERENCES

1. Albert Bandura. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* 84, 2 (1977), 191.
2. Yochai Benkler and Helen Nissenbaum. 2006. Commons-based peer production and virtue. *Journal of Political Philosophy* 14, 4 (2006), 394–419.
3. Nama R Budhathoki. 2010. *Participants' motivations to contribute geographic information in an online community*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.
4. Moira Burke, Cameron Marlow, and Thomas Lento. 2009. Feed me: motivating newcomer contribution in social network sites. In *Proc. SIGCHI '09*. ACM, 945–954.
5. Boreum Choi, Kira Alexander, Robert E Kraut, and John M Levine. 2010. Socialization tactics in Wikipedia and their effects. In *Proc. CSCW '10*. ACM, 107–116.
6. Giovanni Luca Ciampaglia, Dario Taraborelli, and San Francisco. 2015. MoodBar: increasing new user retention in Wikipedia through lightweight socialization. In *Proc. CSCW '15*. 734–742.
7. Liat Clark. 2014. The race to contain West Africa's Ebola outbreak - Wired UK. `http://www.wired.co.uk/news/archive/2014-04/11/ebola-open-street-map`. (2014).
8. E Gil Clary, Mark Snyder, Robert D Ridge, John Copeland, Arthur A Stukas, Julie Haugen, and Peter Miene. 1998. Understanding and assessing the motivations of volunteers: a functional approach. *Journal of personality and social psychology* 74, 6 (1998), 1516.
9. Melanie Eckle and João Porto de Albuquerque. 2015. Quality assessment of remote mapping in OpenStreetMap for disaster management purposes. *ISCRAM 2015* (2015).
10. Rosta Farzan, Robert Kraut, Aditya Pal, and Joseph Konstan. 2012. Socializing volunteers in an online community: a field experiment. In *Proc. CSCW '12*. ACM, 325–334.
11. Médecins Sans Frontières. 2014. MSF: the Missing Maps project. (2014). `http://www.msf.org.uk/missing-maps-project`.

12. R Stuart Geiger and Aaron Halfaker. 2013. Using edit sessions to measure participation in Wikipedia. In *Proc. CSCW '13*.

13. Jeff Howe. 2006. The rise of crowdsourcing. *Wired magazine* 14, 6 (2006).

14. Desislava Hristova, Giovanni Quattrone, Afra Mashhadi, and Licia Capra. 2013. The life of the party: impact of social mapping in OpenStreetMap. In *Proc. ICWSM '13*. 234–243.

15. Shashank Khanna, Aishwarya Ratan, James Davis, and William Thies. 2010. Evaluating and improving the usability of Mechanical Turk for low-income workers in India. In *ACM DEV*.

16. A Kittur, J Nickerson, and M Bernstein. 2013. The future of crowd work. *Proc. CSCW '13* (2013).

17. Timo Lüge. 2014. GIS support for the MSF ebola response in Guinea 2014. (2014). Médecins Sans Frontières, Logistics Department. `http://preparecenter.org/resources/gis-support-msf-ebola-reponse-guinea-2014`.

18. Chris Michael. 2014. Missing Maps: nothing less than a human genome project for cities - The Guardian. `http://www.theguardian.com/cities/2014/oct/06/missing-maps-human-genome-project-unmapped-cities`. (2014).

19. Oded Nov. 2007. What motivates Wikipedians? *Commun. ACM* 50, 11 (2007), 60–64.

20. OpenStreetMap. 2015a. 2014 West Africa Ebola Response - OpenStreetMap Wiki. `https://wiki.openstreetmap.org/wiki/2014_West_Africa_Ebola_Response`. (2015).

21. OpenStreetMap. 2015b. Missing Maps Project - OpenStreetMap Wiki. `https://wiki.openstreetmap.org/wiki/Missing_Maps_Project`. (2015).

22. OpenStreetMap. 2015c. Typhoon Haiyan - OpenStreetMap Wiki. `https://wiki.openstreetmap.org/wiki/Typhoon_Haiyan`. (2015).

23. Leysia Palen, Robert Soden, T Jennings Anderson, and Mario Barrenechea. 2015. Success & scale in a data-producing organization: the socio-technical evolution of OpenStreetMap in response to humanitarian events. In *Proc. SIGCHI '15*. 4113–4122.

24. John W Payne. 1982. Contingent decision behavior. *Psychological Bulletin* 92, 2 (1982), 382.

25. Lesandro Ponciano and Francisco Brasileiro. 2014. Finding volunteers' engagement profiles in human computation for citizen science projects. *Human Computation* 1, 2 (2014).

26. Lesandro Ponciano, Francisco Brasileiro, Robert Simpson, and Arfon Smith. 2014. Volunteers' engagement in human computation astronomy projects. *Computing in Science & Engineering* (2014).

27. Missing Maps Project. 2015. Missing Maps - putting the world's vulnerable people on the map. `http://www.missingmaps.org`. (2015).

28. Paul G Schervish and John J Havens. 1997. Social participation and charitable giving: a multivariate analysis. *Voluntas: International Journal of Voluntary and Nonprofit Organizations* 8, 3 (1997), 235–260.

29. Manuela Schmidt, Silvia Klettner, and Renate Steinmann. 2013. Barriers for contributing to VGI projects. In *Proc. ICC '13*.

30. Robert Soden and Leysia Palen. 2014. From crowdsourced mapping to community mapping: the post-earthquake work of OpenStreetMap Haiti. (2014).

31. Sam Sturgis. 2014. Open-source mapping the world's most vulnerable regions will save lives. (2014). `http://www.citylab.com/design/2014/12/open-source-mapping-the-worlds-most-vulnerable-regions-will-save-lives/383337/`

32. Humanitarian OpenStreetMap Team. 2015. Featured HOT projects. `http://hotosm.org/projects`. (2015).

33. Loxley Sijia Wang, Jilin Chen, Yuqing Ren, and John Riedl. 2012. Searching for the Goldilocks zone. In *Proc. CSCW '12*. 989.

34. Clay Westrope, Robert Banick, and Mitch Levine. 2014. Groundtruthing OpenStreetMap building damage assessment. *Procedia engineering* 78 (2014), 29–39.

35. Andrea Wiggins and Kevin Crowston. 2011. From conservation to crowdsourcing: A typology of citizen science. *Proc. HICSS '11* (2011).

36. Robert E Wood. 1986. Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes* 37, 1 (1986), 60–82.