

Large-scale Volunteer Engagement in Humanitarian Mapping

Martin Dittus · ICRI Cities, UCL · @dekstop
28th Feb 2016 · CSCW DC 2016, San Francisco



The Humanitarian OpenStreetMap Team (HOT)

Making maps to support field logistics by aid organisations.

Online **crowdsourcing** with thousands of volunteers.

Part of **humanitarian relief efforts**: Haiti earthquake, typhoon Haiyan in the Philippines, 2014 Ebola epidemic, Nepal earthquake, ... hundreds of initiatives.

Also an **online commons**: the maps are on OSM, freely available for everyone to use.



English Français

About Updates Projects Get Involved Donate



FEATURED PROJECT Mapping Financial Inclusion in Uganda

Access to digital financial services is fundamental to enabling poor people to become more economically stable, prosperous, and resilient. HOT, with support from the Bill and Melinda Gates Foundation, will initiate a six-month pilot program in Uganda starting in December 2015.

The Humanitarian OpenStreetMap Team [HOT] applies the principles of open source and open data sharing for humanitarian response and economic development.

RECENT UPDATES

Mapping in Mozambique to Help Reduce Child Mortality

Posted by *Blake Girardot* on Feb, 19 2016

Mozambique is a country of 25 million people on the east coast of Africa, often beset by natural disasters ranging from severe drought to devastating floods as well as civil wars and strife. Dar Ramani Huria and HOT have recently started mapping in Mozambique, to support critical public health work.

[Read more...](#)

A day mapping in Dar Es Salaam: drainage in Chang'ombe

Posted by *Paul Uithol* on Feb, 16 2016

Guest post by Kathryn Davis, Columbia University. January 14th, 10:30am. We arrive in Chang'ombe ward (an industrial area in Dar es Salaam) after weaving in and out of traffic in the project bajaj (rickshaw) through different neighborhoods. Johannes Peter, a seasoned mapper with the Dar Ramani Huria project, takes another look at his drone imagery printout of the area, and indicates that we've arrived at our starting point. Mappers like Johannes detail key drainage and water infrastructure that hasn't been digitized before, and that can be used for disaster risk reduction planning and flood resilience. Almost immediately, we see drainage on the right side of the wide dirt road, and mark it on the map...

[Read more...](#)

OSM Celebrates International Women's Day 2016

Posted by *Courtney Clark* on Feb, 11 2016

A young woman anxiously practices a talk about youth in OSM that she is slated to give at the HOT Summit the following day. Impostor's syndrome has convinced her that it would be better for her not to give the talk at all; she's relatively new to HOT and has few technical skills. She'll be easily exposed as a fraud, and the audience members surely have better uses of their time than listening to her (spoiler alert: that was me at the first HOT Summit).

[Read more...](#)

<https://hotosm.org>



Martin Dittus · ICRI Cities, UCL · @dekstop

Remote mapping and field mapping

How we work

Step 1.

Remote volunteers trace satellite imagery into OpenStreetMap



Step 2.

Community volunteers add local detail such as neighborhoods, street names, and evacuation centers



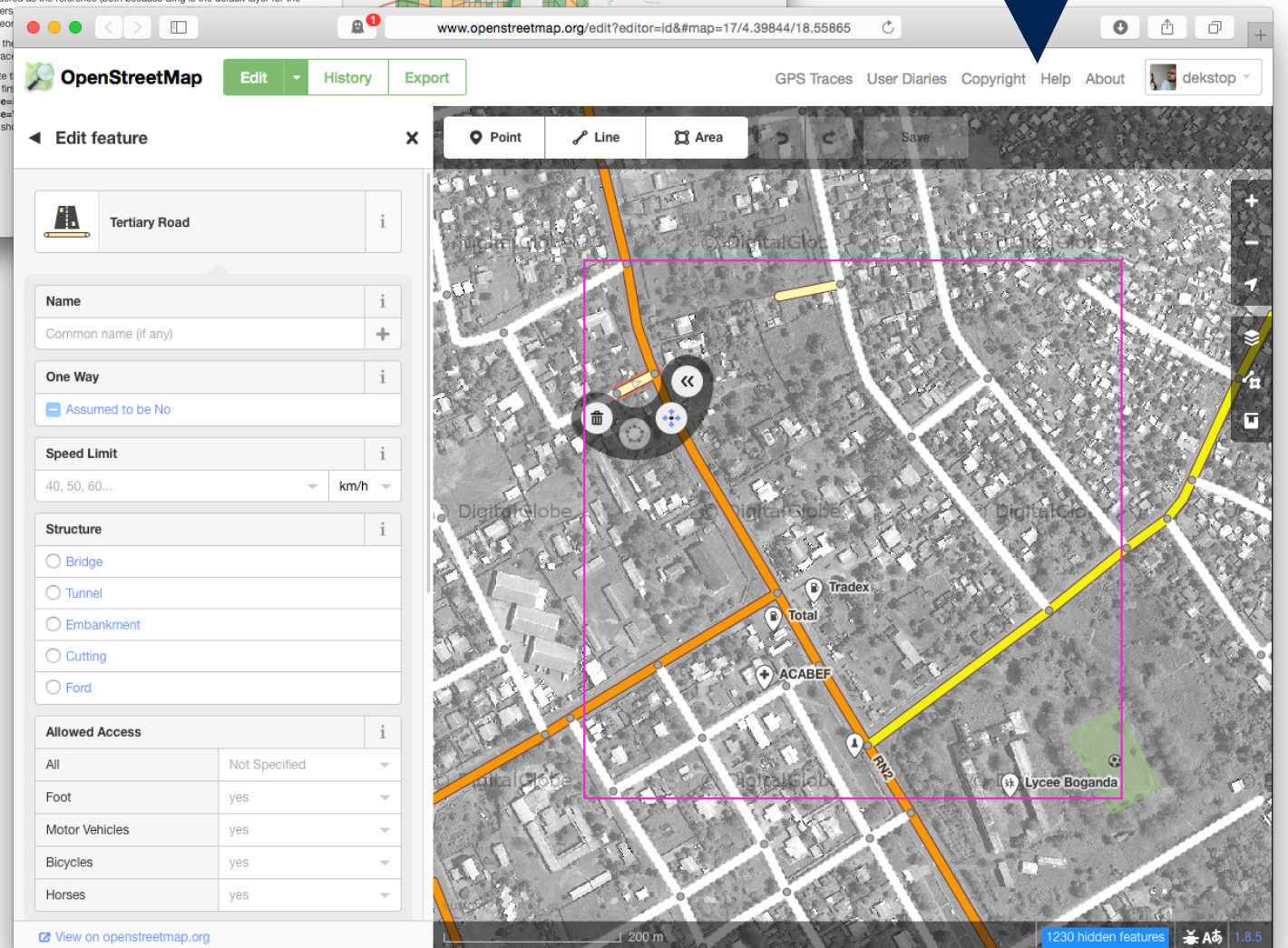
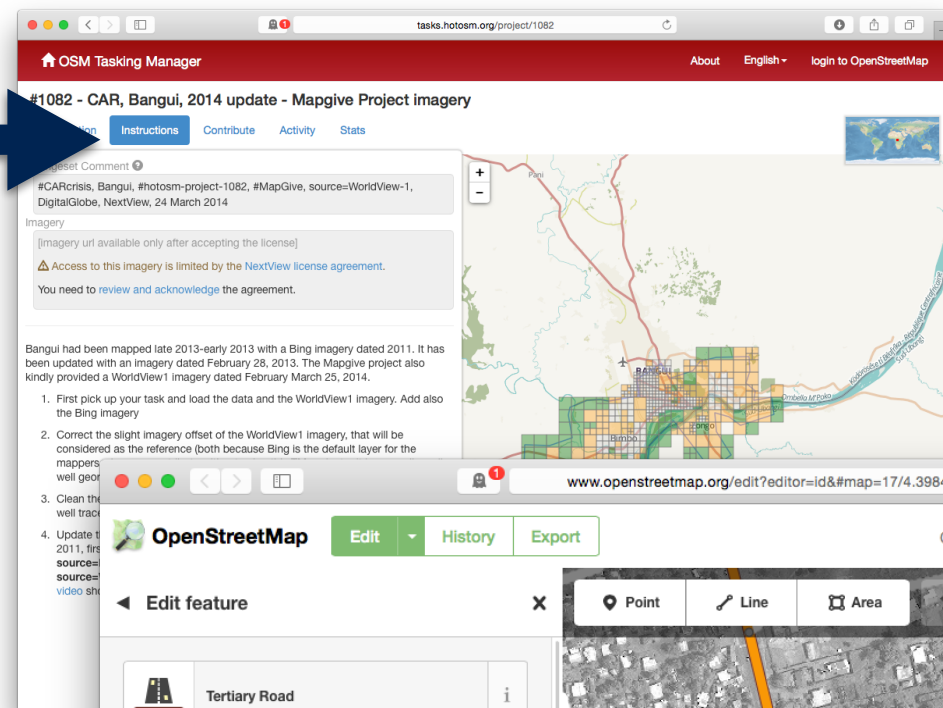
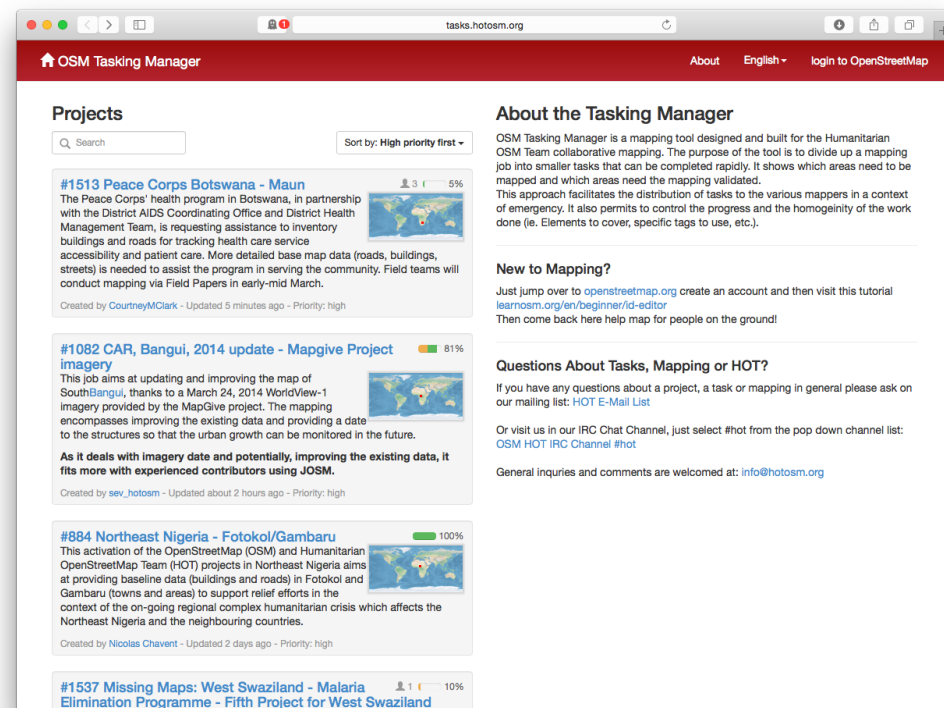
Step 3.

Humanitarian organizations use mapped information to plan risk reduction and disaster response activities that save lives



<http://www.missingmaps.org>

Remote mapping with HOT: tracing imagery



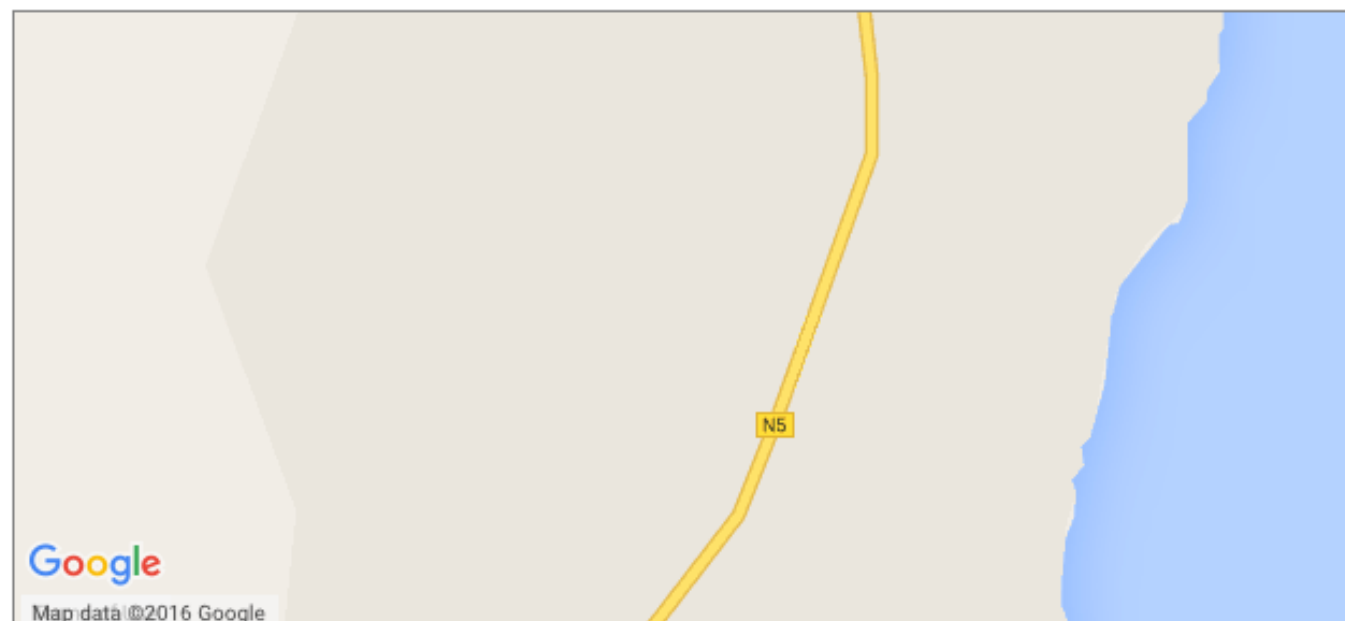
<http://tasks.hotosm.org>

E.g. putting Baraka, DRC on the (digital) map.

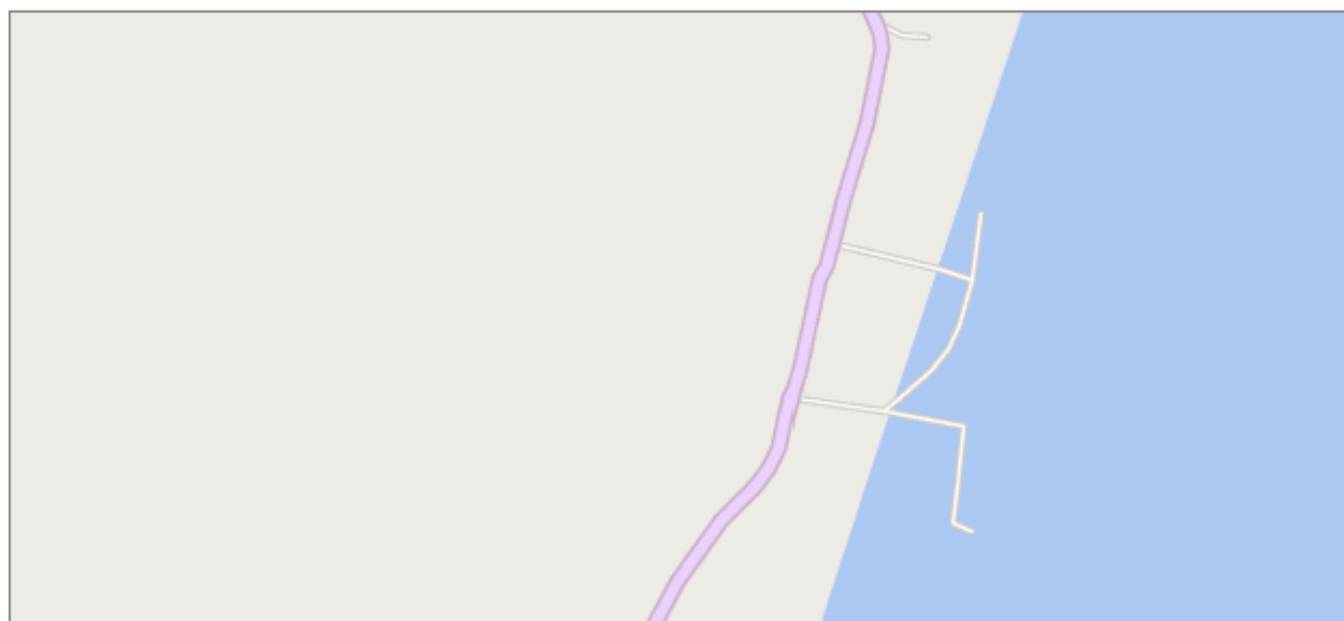
Choose map type: Google Satellite



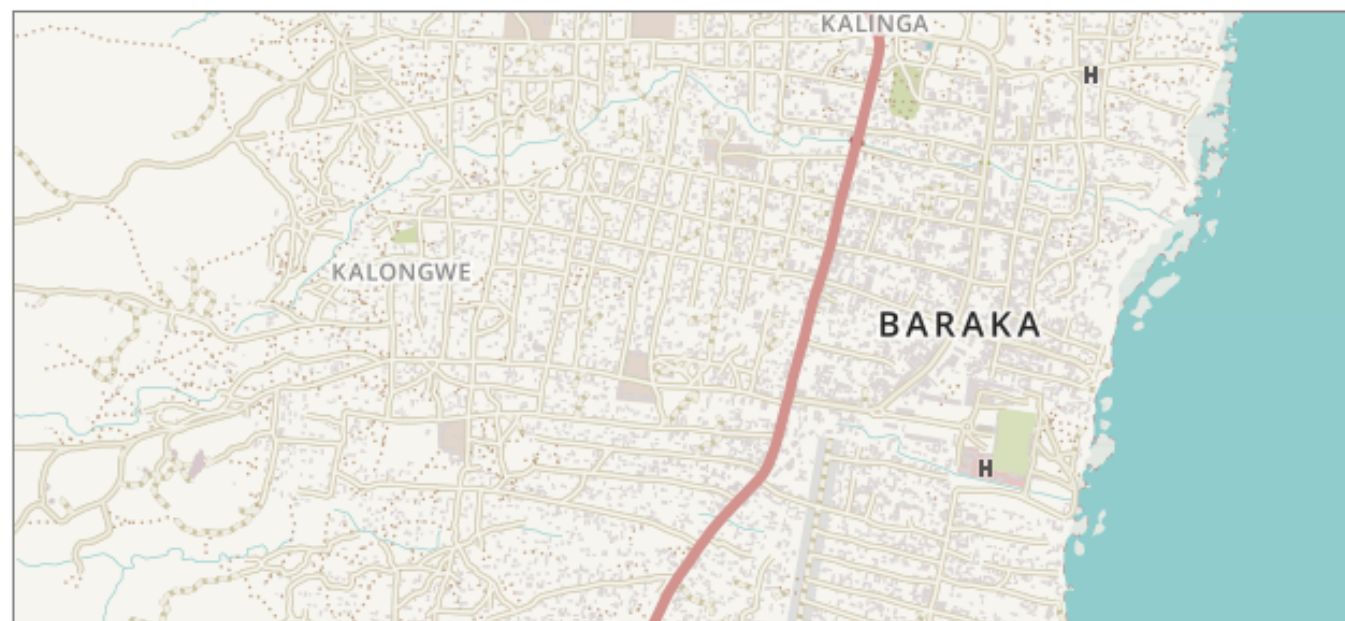
Choose map type: Google Map



Choose map type: Bing Map



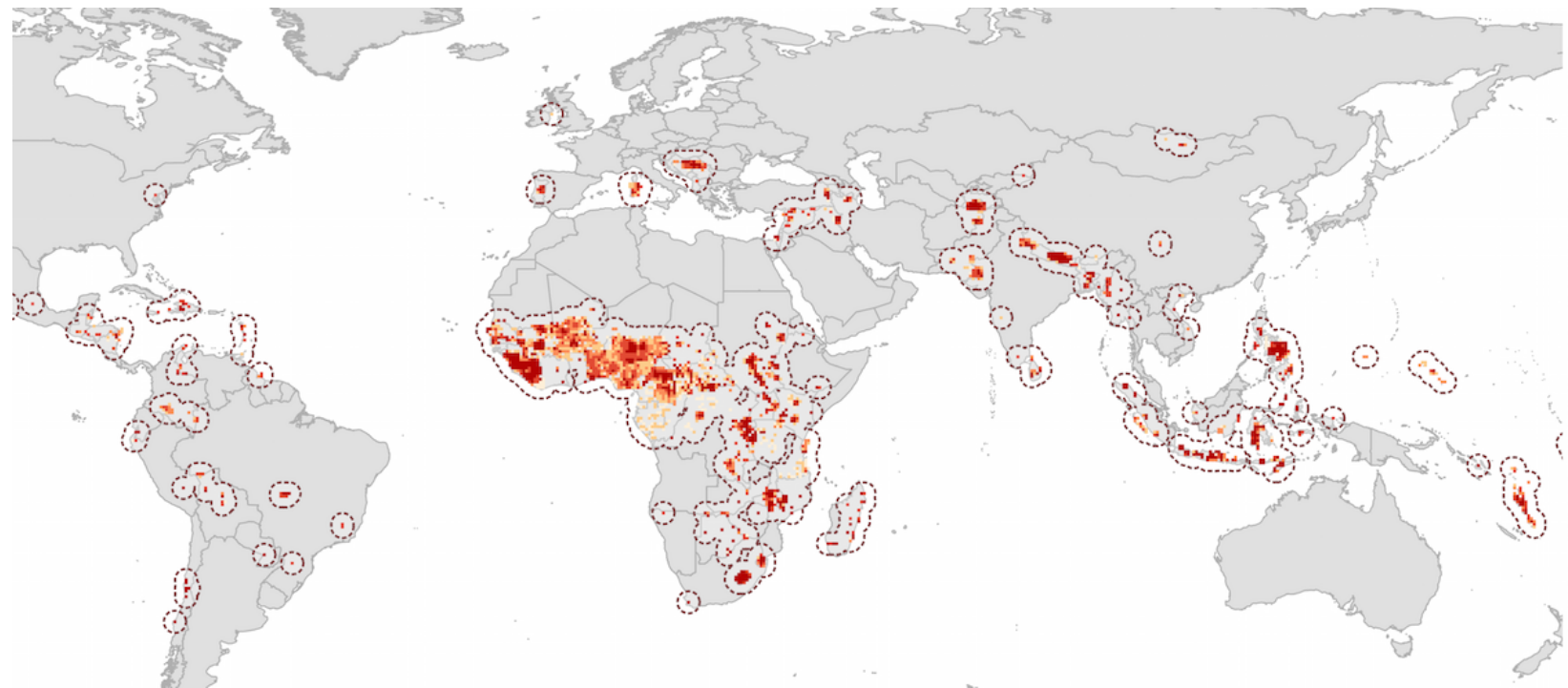
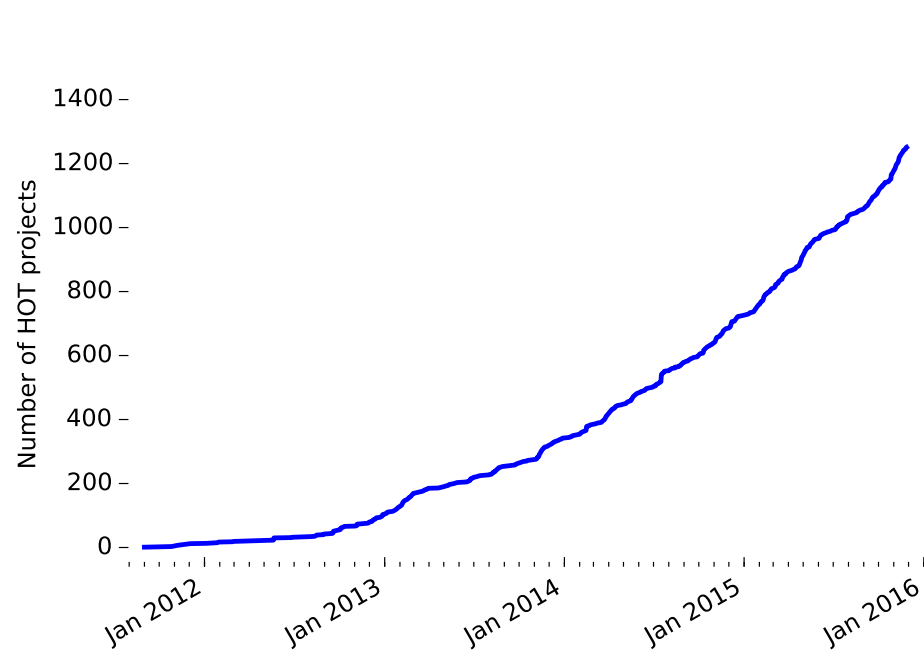
Choose map type: OSM Mapnik (humanitarian)



<http://tools.geofabrik.de/mc/>

HOT contributor activity to date

- 1,000 projects,
- 20,000 contributors,
- 120M changes made to the map,
- involving an estimated 150,000 hours of volunteer work.



The HOT challenge...

Vast regions of the world remain unmapped.

Ebola epidemic: it can take months to map larger regions

Missing Maps: a proactive HOT mapping initiative.

Not all HOT causes draw their own crowds.

“CNN moments”: certain disasters advertise themselves.

How to foster sustained volunteer capacity in the absence of urgent causes?

E.g. for proactive mapping, map maintenance, etc.
HOT organisers have many opportunities for interventions.

What should organisers focus on?

A complex socio-technical system. A wide range of initiatives, practices, purposes, cliques, ... many intersecting concerns.

Where to begin looking?

A rare opportunity:

Much empirical data is freely available.

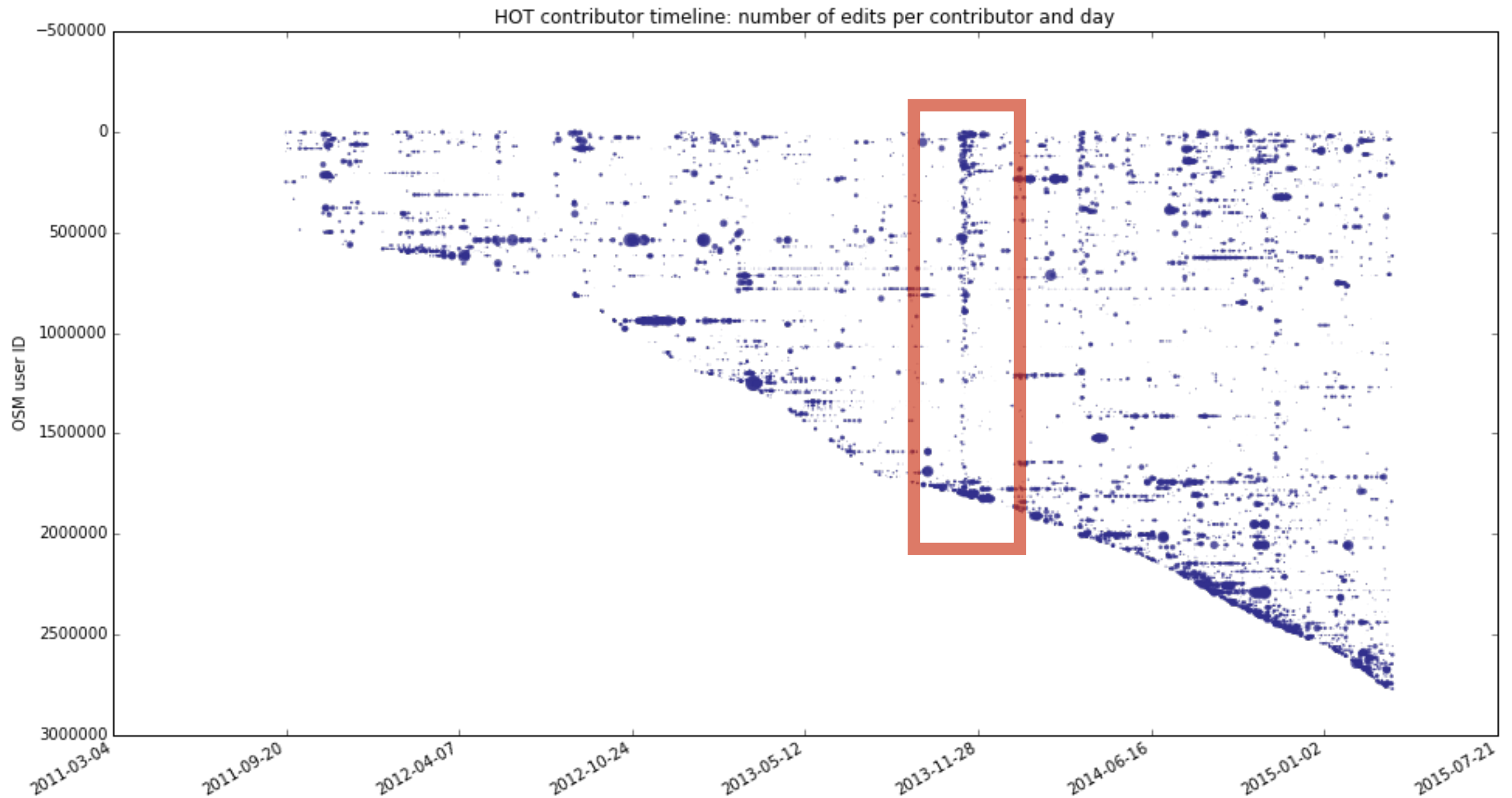
The full contribution history is public.

Part of the OpenStreetMap edit history.
~100GB compressed XML/PBF.

This means:

We can observe past outcomes of particular initiatives.

I spent much time exploring the data...



... and embedded in the community.

As contributor, participant observer.

A public research diary to share early findings.

Many discussions online and in person.

Building domain knowledge, collective understanding.



dekstop's diary



A researcher's scrapbook: understanding contributor engagement in humanitarian mapping

Posted by [dekstop](#) on 7 June 2015 in [English \(English\)](#)

Hallo! My name is Martin Dittus, and I'm a PhD student at the [ICRI Cities](#) at University College London. I research community engagement in the [Humanitarian OpenStreetMap Team \(HOT\)](#), a volunteer initiative with thousands of contributors. At its core this is quantitative work, and my main outputs are statistics and data visualisations. I also spend a lot of time with the HOT community, am a contributor myself, and have spent much of the last decade with a range of similar community organisations.

I like that my job allows me to combine my experience in large-scale data analysis with my personal interest in community organisations. I spend a lot of time exploring data sets, producing things like this:

Although I was surprised by this, this is not actually entirely unexpected. JOSM use tends to be associated with higher engagement: the most active mappers are often JOSM users.

However this does not necessarily mean that JOSM is the key trigger. It might simply reflect that the JOSM mappers at our events are a great bunch of people, fun to hang out with, and many of them know each other quite well; whereas the people at our ID tables are typically newcomers who are not yet as well-connected to the community. So maybe the difference is in the people, not the editor.

In closing I would say that we need many more observations across different kinds of settings to make these statistics meaningful. At the moment this is little more than anecdotal evidence. There's definitely space for further experiments!



Comment from [Joost Schoupe](#) on 7 December 2015 at 19:24

Hi Martin,

This is a slightly more elaborate answer than what I expected :)

Here's some thoughts: - the longer term retention is worth keeping an eye on as numbers increase, keeping in mind selection and social effects of course - are you sure the difference in labour hours at the events is real and not an artifact of the way JOSM and ID save information to the database? For example, if you take the difference between first and last object saved as labour time, that might affect JOSM negatively. Did you (or could you try) look at the last save during an event minus the start time of the event? Probably difficult on a larger scale, but might be worth a check at a single event. - you filtered away the people who used both editors. It might be interesting to see if they started off using JOSM and gave up to fall back to ID or if they took the other way around. If it looks like the former, one can imagine more people giving up before having anything worth saving.

Thanks again. We'll think about how we can experiment more at our events in Belgium.



Comment from [dekstop](#) on 7 December 2015 at 19:29

Ah you're right -- it is possible that JOSM captures changeset timestamps differently. In past explorations I've seen JOSM preserve timestamps for individual edits within a changeset, but I don't know enough about the editor to understand what exactly is going on.

Only 6 people used both editors at their first event, so I felt it's better to simply ignore these :)



Comment from [Warin61](#) on 7 December 2015 at 20:22

What is the coverage factor in the confidence interval? Around 2 is usual for metrologists.

I don't think difference of the initial contribution should be a deciding factor. People learn differently. Most would 'test' at a later time, so sampling as, say 2 months later for contributions, and the number of nodes added/modified/deleted? With possible follow up at 1 and 2 years?

A forensic approach: “follow the evidence”.

Exploratory instead of hypothesis-driven. Starting with simple models, incorporating more over time.

Our starting point: observe newcomer contributions across different settings.

See what works well, identify barriers to engagement.

- Where do first-time mappers leave early?
- Where do first-time mappers remain active for longer
- Infer how this may affect long-term engagement.

Goldberg (2015), “In defense of forensic social science.”

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.

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 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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CSCW '16, February 27–March 02, 2016, San Francisco, CA, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3592-8/16/02...\$15.00.
DOI: <http://dx.doi.org/10.1145/2818048.2819939>

Are task designs a barrier for newcomers?

Are particular forms of guidance more helpful?

The screenshot shows the OSM Tasking Manager web interface. The browser address bar displays 'tasks.hotosm.org/project/727'. The page has a red header with 'OSM Tasking Manager' and navigation links for 'About', 'English', and 'desktop'. Below the header, the project title '#727 - Missing Maps: Baraka, Democratic Republic of Congo' is shown, along with tabs for 'Description', 'Instructions', 'Contribute', 'Activity', and 'Stats'. The 'Instructions' tab is active, displaying a list of tasks with a legend indicating their status: 'Cur. worked on (0)' (orange), 'Invalidated' (grey), 'Done' (yellow), and 'Validated' (green). The map on the right shows the Baraka area in the Democratic Republic of Congo, with a scale bar indicating 3 km and 2 mi. The map data is attributed to OpenStreetMap contributors.

OSM Tasking Manager About English desktop

#727 - Missing Maps: Baraka, Democratic Republic of Congo

Description Instructions Contribute Activity Stats

Instructions

Please trace areas of housing (landuse=residential), buildings (building=yes) and map roads and smaller paths for as far as you can follow them (see below for useful tags and resources). Also trace waterways such as streams, rivers, canals etc. Use Bing Imagery.

Save with credit to HotOSM. Comment the changeset with: `DRC, Baraka, #hotosm-task-727 #MissingMaps source=Bing`

Commonly used tags and useful resources:

Residential areas:

- landuse=residential for a border around a cluster of residential buildings
- building=yes for rectangular & round buildings - trace the building outline
- building=construction for walls without a roof

Road networks

- highway=primary, secondary or tertiary for main roads connecting towns
- highway=residential inside residential boundaries
- highway=track only goes to farmland
- highway=path - not even a 4x4 can get down this

Waterways

- waterway=river for centreline on waterway
- waterway=riverbank for river banks!
- waterway=stream
- waterway=wadi - path of a seasonal river or stream - depends on the time of year as to whether water present
- natural=water & water=lake for lakes

Other useful tags

- leisure=common for large open areas (not fields, just smooth open ground next to towns) as as these are potential helicopter landing sites.
- amenity=school - school areas are easily identified as 1 or 2 long buildings at the edge of the village, sometimes with two small toilet buildings behind them. The schools are almost always accompanied by a large open area for the children to play in, tag the whole area of the school complex with amenity=school and trace the school buildings and bathrooms.

See the [OSM Highway Tag Africa](#) page for a localised description of some of these tags & the [OSM map features](#) page for more info

Cur. worked on (0)
Invalidated
Done
Validated

3 km
2 mi

Leaflet | Map data © OpenStreetMap contributors

We assessed task complexity for 100 projects.

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

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

First-time mappers spent more time on tasks that involved the mapping of *less than three distinct map features*.

No other factors had a consistent impact.

Dittus, Quattrone, Capra (2016): “Analysing Volunteer Engagement in Humanitarian Mapping: Building Contributor Communities at Large Scale”, CSCW 2016.

Do coordination practices matter?

Perceived urgency may serve as attractor, but how long do people stick around?

We compared three cohorts:

- **Typhoon Haiyan:** one-off, urgent, highly promoted.
- **Ebola epidemic:** sustained, first uses of mapathons.
- **Missing Maps:** proactive rather than reactive. Focused on community-building. Mapathons, email alerts, ...

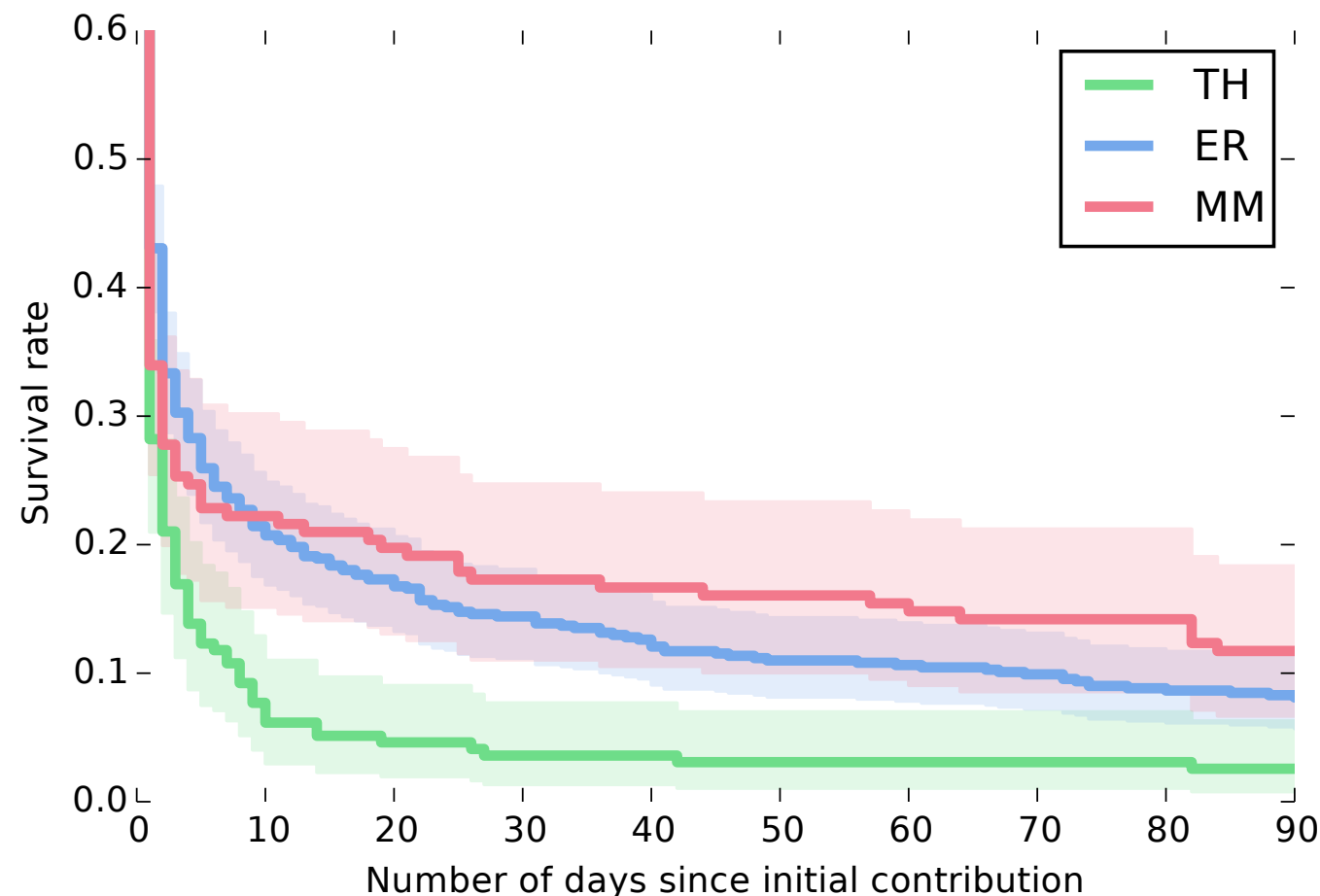
This is what a mapathon looks like.



Social contribution environments for newcomers:

- In-person training and guidance
- Safe spaces to make early mistakes

Retention of newcomers w/o prior OSM experience: Missing Maps is doing very well.



Disaster response project
Haiyan: low retention.

Ebola, MM: high retention.

**Possibly related to
coordination practice?**
Mapathons, activities over
longer periods, email
notifications, ...

Dittus, Quattrone, Capra (2016): “Analysing Volunteer Engagement in Humanitarian Mapping: Building Contributor Communities at Large Scale”, CSCW 2016.

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

Under submission

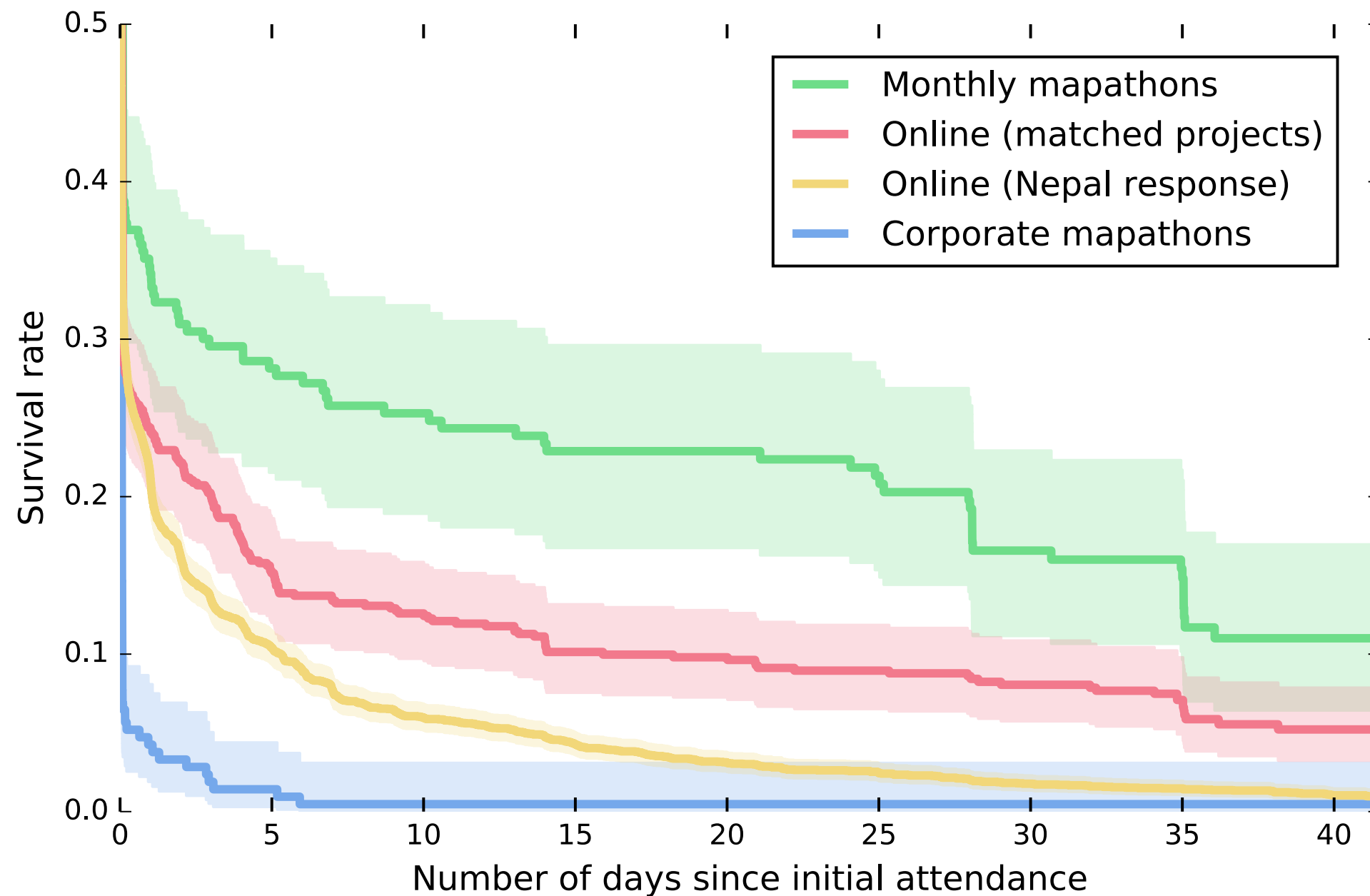
Martin Dittus · ICRI Cities, UCL · @dekstop

Newcomer attendance at mapathons

Date	Cohort	Attendees	Newcomers	% newcomers
2014-11-24	Monthly	64	37	57.8%
2014-12-15	Monthly	58	24	41.4%
2015-01-27	Monthly	52	16	30.8%
2015-02-12	Corporate	50	44	88.0%
2015-02-24	Monthly	49	25	51.0%
2015-03-31	Monthly	62	29	46.8%
2015-04-28	Monthly	51	19	37.3%
2015-05-15	Corporate	191	174	91.1%
2015-06-02	Monthly	27	6	22.2%
2015-07-07	Monthly	51	15	29.4%
2015-08-04	Monthly	87	49	56.3%
2015-09-01	Monthly	41	15	36.6%
2015-10-06	Corporate	30	28	93.3%
2015-10-06	Monthly	69	24	34.8%

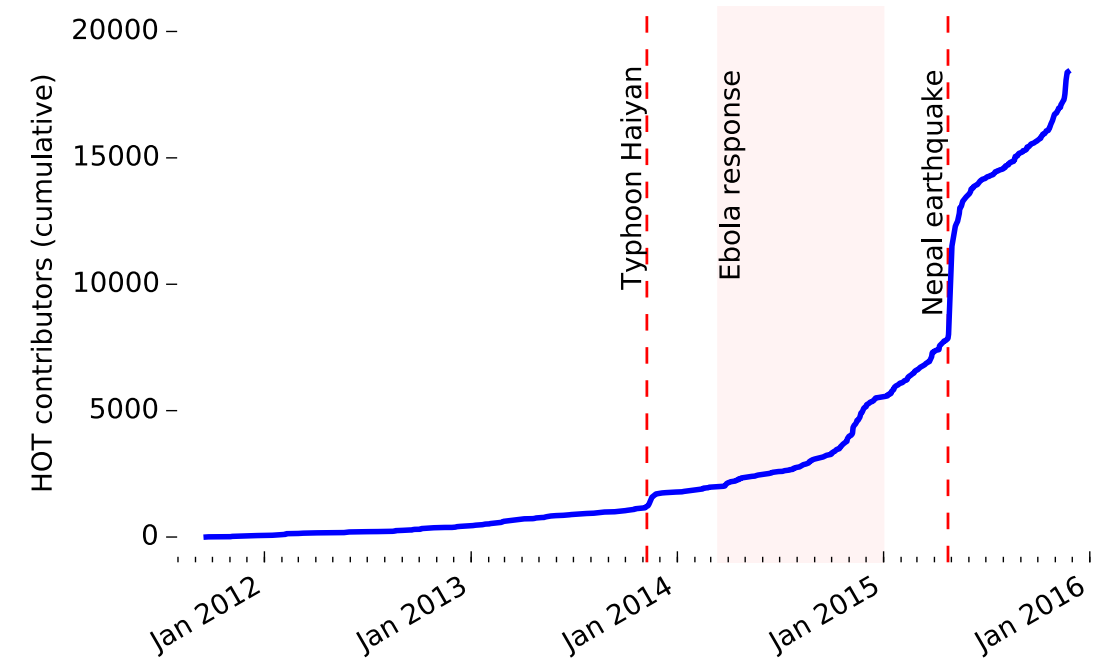
Table 1: Estimated attendance at the 14 mapathons under study, including the number and share of first-time attendees.

Mapathons as social attractors?



Implications (1/2)

Overall, HOT seems to be doing very well. It appears **relatively easy to get people interested** at first.



However it's **harder to get sustained engagement.**

That's only partly about the technology, the process.
Also **social factors**, factors relating to the **individual**.
The human experience.

Can we imagine **online support experiences that imitate the mapathon setting**? Expert guidance, peer support, the presence of a community of practice, ...

Implications (2/2)

Likely also self-selection effects. Promotion, social ties, ... the cohorts had different contributor profiles.

If the aim is sustained engagement, then **indiscriminate growth may not be advisable.**

Instead:

- **Identify highly-engaged contributors early** and give them support.
- **Identify participation contexts** that are more likely to yield highly motivated contributors.

Many new questions raised by early studies.

A more comprehensive model of **volunteer capacity**?

Volume of contributions, activity periods, ...

Long-term effects among more experienced mappers?

Are mappers reactivated by high-profile events?

Are there generational effects to consider — are early joiners different from more recent joiners?

How do any of these aspects relate to **data quality**?

... etc.

Impact of the work to date

First paper published at CSCW 2016.

Research diary now widely referenced in the community.

<https://www.openstreetmap.org/user/dekstop/diary>

Working closely with organisers at HOT, MSF, BRC, ARC, ...

A recurring conversation: "Can you help us understand X?"

HOT groups have already picked up many of my insights.

E.g. Missing Maps organisers have made changes to task design, email notifications for mapathon attendees, and other aspects of their coordination practice.

Methodologically...

Forensic social science as a means of laying the groundwork for future research.

Don't need to be hypothesis-driven. Instead can combine *empirical work with exploration, embedded practice*.

Now have a *clearer picture* of contributing factors, and starting points for more focused work.

The embedded/collective approach was mutually beneficial.

We have deeply improved our domain understanding.

The community is already adopting our recommendations.

Request for feedback:

What aspects to focus on?

I know why I'm doing the work.

I know what it means to the community.

I have a fair sense of what else I need to to to graduate.

What of this should I present to the CSCW community?

HOT as an object of study?

The study: task designs, coordination practices?

The research methods; forensic approach, embedded practice, collective understanding?

The impact of the work to date? (Only partially related to the paper)

... etc.

Thank you.

Martin Dittus · @dekstop

<https://openstreetmap.org/user/dekstop/diary>