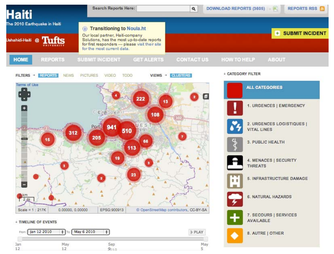
Population Movement & Accuracy of Citizen-Generated Information—Ushahidi Haiti

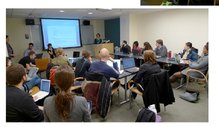
BACKGROUND

In response to the 2010 earthquake in Haiti, members of the technology community rallied together in an effort to alleviate traditional information shortages during a crisis. One map, the Ushahidi Haiti mash-up, was able to gather and geolocate over 2,500 reports of need within the first two weeks. This map compiled information from Twitter, Facebook, and traditional media in combination with messages translated from a free short-code, 4636. Although this was an unprecedented use of technology, the response to the earthquake still fell short. Traditional responders were not equipped to manage these new information flows and this information went underutilized. Responders mentioned that one of the key problems with these tools, especially Ushahidi Haiti, was that it was raw, unanalyzed information. Although the map demonstrated clusters of points, it still did not accurately convey trends over time. Furthermore, this was a new information collection tool to most of them - they were not sure if citizen-generated (or crowd sourced) information could be trusted.

This analysis seeks to answer the question: *to what degree of accuracy did the Ushahidi Haiti data indicate population concentrations and their movement over time?*



Ushahidi Haiti Map
Source: http://haiti.ushahidi.com



Ushahidi Haiti Team (above)
Ushahidi Haiti Volunteer Training (below)
© Carol Waters 2010

METHODS

This analysis answers that question using a two parts. However, prior to this, the Ushahidi Haiti data required some "cleaning up." I analyze only reports concerned with larger populations of 50 or above. This is because the goal is to see if this data demonstrates how internally displaced persons (IDPs) moved. This required deleting all reports of small populations, unrelated events, or duplicate reports. Then, using information within the message, I added a population field. See the example below:

"15,000 people with only four days' supply of food on ile a vache. Can't get a response from the orgs working in that area..."

From this message, it's clear that the population of the camp was 15,000 people. However, numerous messages indicate a large camp or group of people, but there is no clear population data. These reports are included in the dataset with the number "-999" in the population field. See the example below.

"Since most houses are destroyed, many families have gathered at soccer field."

The first part of the analysis examines only the Ushahidi Haiti reports over the first two weeks in order to determine if there are any trends. This analysis uses two heat maps, one for each week pictured to the right. The search radius is 5 km with an output cell size of 50 m. Because it is clear there is a large amount of change in and around Port-au-Prince (PaP), the next two maps of display Week 1 and Week 2 in only the greater PaP area with a search radium of 1 km and an output cell size of 25 m. The last map displays the difference.

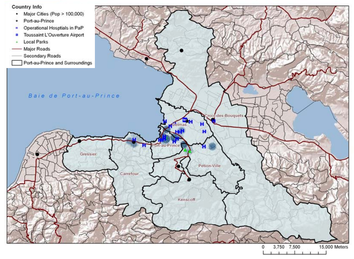
The second part of the analysis compares the Ushahidi Haiti reports to official datasets concerning IDPs from international organizations. The first dataset, created by the Office of Foreign Disaster Assistance (OFDA), indicates the camps in place as of January 24, 2010. The second dataset, created by the Camp Coordination and Camp Management Cluster (CCCM), indicates camps in place as of February 16, 2010. From there I compare two Ushahidi Haiti data sets, one with all reports until January 24, and the second with all reports until February 16. However, please note that both the OFDA and CCCM datasets are only of greater Port-au-Prince (PaP) and therefore both Ushahidi Haiti datasets are also confined to the same region. To analyze the accuracy of the Ushahidi Haiti reports, I used two processes. First, I simply select the Ushahidi Haiti reports that were 5 km within an official report. Second, I calculate the distance from an Ushahidi Haiti report to a nearest official camp using the Near tool.

National Soccer Stadium IDP Camp
Source: http://www.eqclearinghouse.org/20100112-haiti/general-information/social-impact-themes-shelter-and-livelihoods/attachment/miles_haiti_day3_4-383

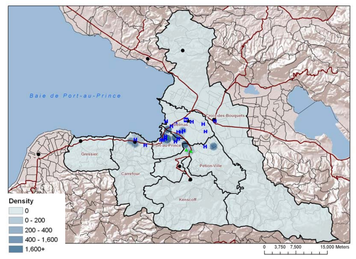
D. Rosalind Sewell
The Fletcher School of Law & Diplomacy
GIS for International Applications
Director of Crisis Mapping, Ushahidi Haiti Project

Projection: UTM Zone 18 N WGS 1984
Sources: OFDA, CCCM, MINUSTAH, ESRI, Google, HHS

Density of IDPs around Port-au-Prince - January 12 - 18, 2010



Density of IDPs around Port-au-Prince - January 18 - 25, 2010



MAP ANALYSIS

From the original 3,272 reports, there were only 347 total usable reports from January 12 until February 16. The table below breaks down the Ushahidi Haiti reports concerning IDPs for Part 1, the heat maps. SMS reports are those received through the short-code 4636, and verified reports are those that were confirmed either by multiple sources or getting in touch with the person reporting the information. Please note that any number in parenthesis indicates the number of reports in that category with data in the population field.

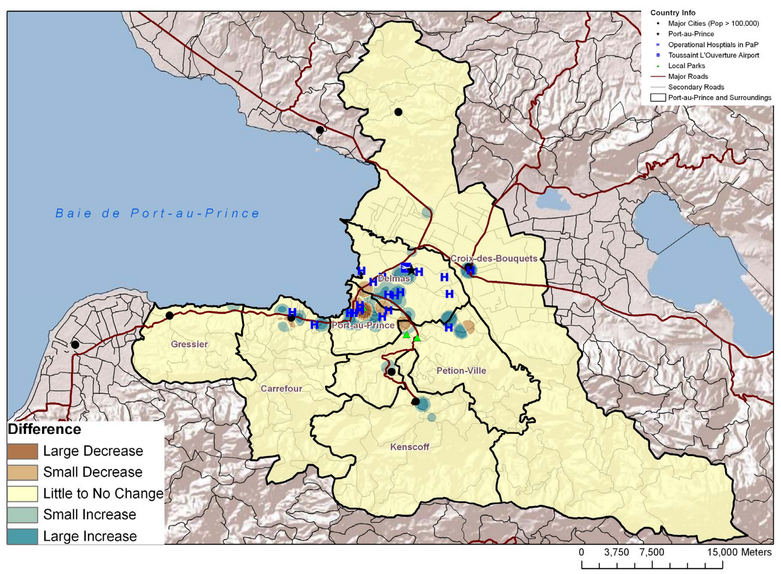
Part 1 - First Two Weeks

Time Period	# Verified Reports	# From SMS	Total
1/12 - 1/18 (Week 1)	9 (6)	10 (6)	61 (28)
- Week 1 (PaP)	5 (3)	10 (0)	51 (22)
1/19 - 1/25 (Week 2)	12 (6)	115 (39)	198 (84)
- Week 2 (PaP)	12 (6)	89 (33)	154 (71)

By looking at the maps of Haiti as a whole, it is clear that the populations remained concentrated in PaP. However, after Week 1 there is some movement along the main highway west toward the larger cities of Leogane and Carrefour. In Week 2, it is also possible to see movement toward the North, but the short-code wasn't as well advertised in these regions, which could account for the smaller populations. These maps are picture below to the right.

The maps of greater PaP (above) demonstrate how the population began to disperse after the earthquake. For instance, in Week 1 the population is centered to the West towards the Port. However, in Week 2, the population gets closer to the airport. It seems as if they are also moving towards the parks, but the difference map below demonstrates that the largest increases were towards the airport, the major outlying cities of PaP, or hospitals in the West. These results are consistent with UN and U.S. Military Situational Reports of population movements after the earthquake.

Difference Between IDP Densities From Week 1 to Week 2



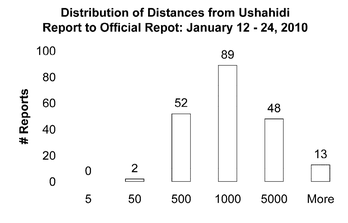
ACCUARACY ANALYSIS

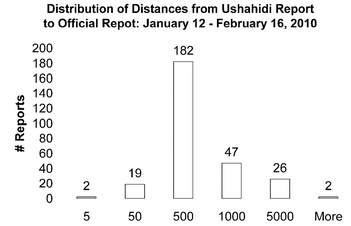
Part 2 - IDP Reports in PaP

Time Period	# Verified Reports	# From SMS	Total
1/12 - 1/24	19 [5]	100 [27]	205 [54]
1/12 - 2/16	24 [20]	135 [93]	278 [203]

The table above indicates the number of reports used in the second part of the analysis. The number in brackets indicates the reports that are within 5 km of an official IDP report. The jump from 54 to 203 reports within 5 km shows that this method of data collection improved over time. By February 16, 2010 over 73% of the reports were accurate within a distance of 5 km, opposed to 26% by January 24. Although 5 km can make a difference for a food distribution point, it is clear that there is a reliable degree of accuracy in this type of information for some purposes, such as understanding general trends.

After removing the outliers, the frequency graphs below demonstrate the distribution of the distance from an Ushahidi report to an official IDP location for both time periods.





For the time period until **January 24**, the range of distance was **11.24 m to 11,208.56 m** with an average of 1,570.27 m. However, by **February 16**, the range was **4.02 m to 8,635.27 m** with an average of 516 m. This analysis also demonstrates that accuracy improved over time. One of the many reasons for this is that the majority of the Ushahidi Haiti team was not familiar with Port-au-Prince, but the more we worked with the maps and got to know the city the more accurate our reports became. Another primary reason is that the media campaigns advertising 4636 emphasized that participants should be as specific as possible when giving location details. This also ensured greater accuracy within the final report. Once this information improved it does display a degree of accuracy that could be useful for some organizations—by February 16, **over 200 reports are accurate within 500 m.**

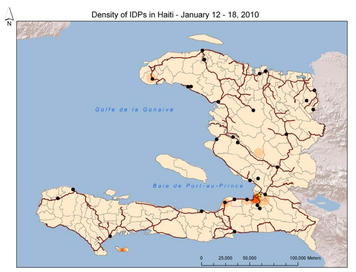
LIMITATIONS & CONCLUSIONS

Future research could evaluate the Ushahidi Haiti data alongside data from the Chilean earthquake, the floods in Pakistan, or the Japanese earthquake. Since the Haiti earthquake there have been numerous deployments of crisis mapping tools without any clear analysis on the accuracy of the information. There has clearly been a proof of concept with these tools. They are changing the way traditional response organizations managed a crisis, but without a clear understanding of their strengths and weakness this information has a danger to again become underutilized.

As the results show, the accuracy of this information is dependent on those who report the information to the system as well as those who process the data. Accuracy changed over time because the members of the Ushahidi Haiti project understood the country better with time. However, in the immediate aftermath of the earthquake, the data had a low degree of accuracy (some reports were over 11,000 m away from an actual IDP camp) and the density analysis simply demonstrated a concentration of people around PaP, which was already obvious given the concentration of people living in the capital. This indicates the necessity of training and a clear information campaign.

However, at the same time, with certain conditions in place, these results indicate that citizen-generated information via a web-mapping platform can be a viable alternative to traditional information collection practices. There needs to be clear training the team geolocating information with an emphasis on locals or those who have lived in the area before.

Density of IDPs in Haiti - January 12 - 18, 2010



Density of IDPs in Haiti - January 19 - 25, 2010