Hazard Map – Real time hazard mapping using social media.

Emma Bee, Senior GIS Analyst, British Geological Survey.
Pascal Coulon, Senior GIS Consultant, SCISYS.
Joanne Robbins, Weather Impact Scientist, UK Meteorological Office.
Jacob Tomlinson, Software Developer, UK Meteorological Office.

Synopsis

During the weekend of the 21 April 2012, a number of enthusiastic software developers and geographic information enthusiasts from around the globe gave up their time to participate in a 'codeathon'. The International Space Apps challenge, an initiative of the Open Government Partnership, is a two day technology development event similar to "Random hacks of kindness" where citizens around the globe work together to solve current critical challenges for society.

One of the challenges submitted for the weekend was called Hazard Map which asked participants to develop a 'map-lication' alerting staff working within a hazard centre environment to the impacts of natural hazards by using data harvested from social media sites such as Twitter or Facebook. This paper and presentation presents the results of the International Space Apps Hazard Map challenge, highlighting the potential usefulness of OpenStreet Map and the OpenLayers Heatmap to those involved in civil contingency planning and response.

The Hazard Map Challenge

Obtaining real-time information about a hazard event as it unfolds, such as a flood or earthquake, has, until recently, largely been limited to the professional media and even then it takes time for journalists to report the situation. However, as was seen during the Japanese Earthquake in 2010 for example, social media (e.g. Facebook, Twitter, YouTube, Flickr etc) was used extensively to gain situational awareness. Videos were posted on YouTube hours before the same clips were used by the professional media. Social media was also used by individuals affected by the disaster as a medium to tell their friends and family that they were safe.

As part of the International Space Apps 'hackathon' event in April 2012, the British Geological Survey and the UK Meteorological Office submitted a challenge on behalf of recently formed Natural Hazard's Partnership[1]. The challenge asked participants to see if real-time information about hazard events could be harvested from social media (or indeed other publicly available data) and presented in such a way that could provide useful verification for scientists looking at the impact of natural hazards, or to alert them to the hazard in cases where scientific instrumentation does not or cannot generate an early warning. Participants were asked to consider whether it was possible to gather geo-tagged images and videos or 'felt reports' from tweets or Facebook posts and collate them into one portal or 'map-lication'. Ideally, the information collated would be stored for scientific verification of hazard-impact models. Although the challenge was primarily aimed at addressing the needs of the scientific community, such an application could also be useful to the disaster response side of civil contingency e.g. MapAction.

[1] The Natural Hazards Partnership (NHP), endorsed by the UK Cabinet Office, brings together expertise from across the UK's leading public sector agencies with the aim of drawing upon scientific advice in the preparation, response and review of natural hazards.

Approach

Since the challenge had to be tackled over one weekend, it had to be broken down in order to produce a manageable outcome utilising the skills within the team participating in the challenge. The team was made up of six people at the event and one person in Canada participating remotely. The skills within the team varied from GIS professionals, web and database specialists, hazard scientists, business analysts and a graphic designer. The team decided to choose one form of social media, Twitter, in which to concentrate its effort over the weekend. The reasons for this were the limiting time factor but also the fact that Twitter provides an easy to use Application Programming Interface (API) to work with and provides geo-located information in the form of tweets. Facebook usually restricts access to profile data, so whilst you may be able to obtain a geolocation, it is difficult to access the information behind the ‘post’ to alert the hazard map user to the hazard event.

The team also decided to use a heatmap approach to visualise data on a map. Heatmaps are graphical representations of the density of point data within geographic space (e.g. Figure 1). With the advance of HTML5 and javascript libraries such a jQuery, the rendering of geospatial data on the fly has vastly improved and heatmaps have become more common place for the visualisation of data on the web. A heatmap approach was deemed to be useful because it could help filter anomalous tweets (e.g. ‘I’m meeting my friends in the Earthquake café’) and enable the user to rapidly identify the areas where the greatest volume of tweets of interest were being reported by the social media.

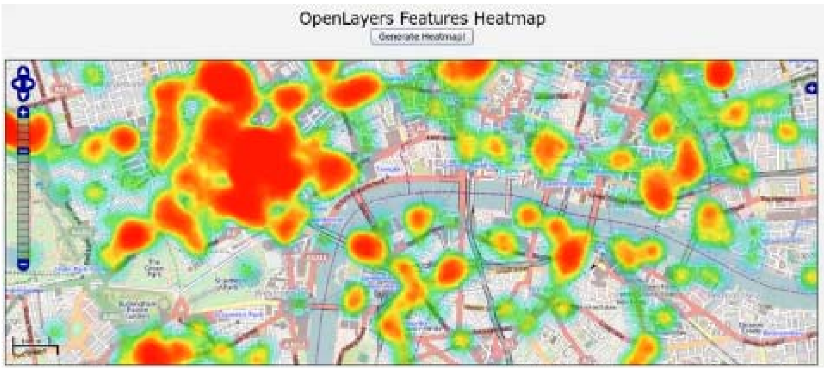

Figure 1: Example of a heatmap using OpenLayers showing pubs and restaurants in London (source: Felipe Barriga Richards).

In order to test the application, the team decided to use the London Marathon as a proxy for a hazard, since this was taking place over the same weekend and people in and around London would hopefully be tweeting about the event. To operate the hazard map, users input a word, such as ‘marathon’ or ‘earthquake’ into a user dialogue box and the ‘map-lication’ searches geotagged tweets for the string. Whilst this offers flexibility, ultimately it was envisaged that the application would also utilise a predefined cached list of words for various hazards so that natural hazard events could be identified even before the user is aware that an event has taken place.

Technical Solution and Challenges

The Hazard Map solution, represented graphically in Figure 2, is divided in three main technical components:

1. OpenLayers: Open Source API enabling the display of geospatial data and use of Open Geospatial Consortium (OGC) data streams.

2. Social Data Handler: enables the query and retrieval of data from key social networks. It translates the data into an internal format suitable for both OpenLayers and the Heatmap API. This is a server side component.
3. Data Visualisation - Heatmap API: the API which enables the display of density type map data.

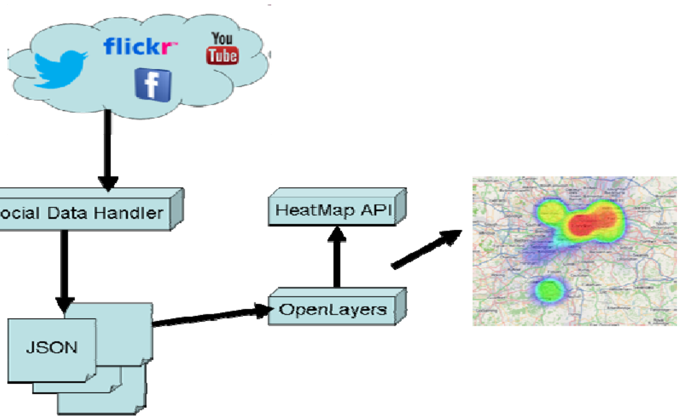

Figure 2: Schematic overview of technical solution for Hazard Map

1. Open Source API

One of the stipulations of the International Space Apps Challenge hackathon was that any code developed over the weekend should be made open for others to access and use beyond the weekend. The team therefore used an Open Source Framework and accessed Open data. The OpenLayers Map API was chosen to enable the implementation of the 'Hazard Map' web mapping application. The OpenLayers API provides a rich, yet simple to use set of interfaces enabling Rapid Application Development (RAD). The choice of OpenLayers, licensed under the Free Berkeley Software Distribution (BSD) licence, allows free deployment of the solution. The OpenLayers API supports data stream for most OGC regulated formats. This enabled the 'map-lication' to integrate data from non social media sources as required, in particular WMS (Web Map Service) and WFS (Web Feature Service).

The Hazard Map application also needed to be operational at a global scale. At the time of the challenge, licence free access to a continuous global base map did not exist. However, the OpenStreetMap community is continually increasing its global coverage of base line mapping. Furthermore, in the event of natural disaster, this mapping is often rapidly updated by volunteers across the world (as observed during the Haiti earthquake in 2010). OpenStreetMap was therefore chosen as the main base mapping solution. The OpenLayers API also enabled the team to display other freely available open data, for example weather data or earthquake related data. In a drive to remain as open as possible it was decided to integrate the OGC compliant Web Mapping Service and Web feature Service geospatial data feeds.

2. Social Data Handler – Twitter API

Social data can be derived from either social networking sites such as Twitter or Facebook (passive crowdsourcing) or through a bespoke crowdsourcing application (active crowdsourcing). The Hazard Map challenge requested participants to consider the passive crowdsourcing of information about natural hazards.

In a drive to enable more complex searches and advanced data display, the data from the social network (e.g. twitter) ideally should be cached. However, one of the obstacles of RAD is that sometimes the terms and conditions of an API can sometimes be overlooked or misinterpreted.

Since the International Space Apps Challenge event took place, it came to light that Twitter's terms and conditions restrict the caching of their data, stating that "*You will not attempt to use or access the Twitter API to aggregate, cache or store place and other geographic location information contained in Twitter Content*" (Source: Twitter 2012)[2]. The Hazard Map team have subsequently chosen to remove the application, and any future development will need to consider using a streaming API as a possible workaround.

During the early phase of the codeathon it was identified that the lack of geo-tagged data from Twitter (approximately 1 in every 100 tweets appeared to be geotagged) maybe an obstacle and would potentially restrict the amount of data displayed. Twitter offers a number of ways to obtain the location of a tweet, some more accurate than others. Each user on Twitter can enable their account to display the location of their last tweet, however, unlike Facebook, the location feature is not turned on by default. The twitter location feature also depends on the type of device used. Although this feature is most often used by mobile users, the triangulation method on such devices is not always accurate enough (sometimes bringing back the location of a server) causing the tweet to be miss-placed.

3) Social Data Visualisation - Heatmap API

Social media data is, by nature, punctual and often represented as a simple point on a map. For Hazard Map, point data would make it challenging to ascertain the degree of importance of an event. For this reason, the team used the HeatMap API during the weekend Hackathon to visualise the data. In general, the term heatmap refers to any display that uses colour to represent quantitative data. Society is familiar with heatmaps in the form of weather maps, which use colour to encode values such as temperature or rainfall. Heatmaps also come in forms other than graphical maps. The use of heatmap was seen as a useful mechanism to visualise averaged event strength and therefore mitigate impact from abnormal records (i.e. tweet).

JavaScript Object Notation (JSON) was the chosen data format for exchanging data between the various parts of the system. JSON is lightweight and easily read by most JavaScript Web API. jQuery was used as the JavaScript framework as it easily deals with JSON data and can make Asynchronous JavaScript and XML (AJAX) requests in a simplified way compared to using straight JavaScript[3]. This meant that users could search for different hazards quickly without having to wait for the page and map to reload each time. jQuery was also used for animation and other interactive features in the project such as sliding menus.

Results

During the course of the weekend, the team produced a live demonstrator, or web 'map-lication', which displayed all geotagged tweets that included the word 'marathon' as a heatmap. As was anticipated, at a global scale, the heatmap (Figure 3) highlighted the UK as a hot spot (since the London marathon was being held that weekend, on Sunday 22 April 2012).

[2] The authors revisited the Twitter rules of the road on the 18/8/12 and the rule now states "*You will not attempt to use or access the Twitter API to aggregate, cache* ***(except as part of a Tweet)*** *or store place and other geographic location information contained in Twitter Content". This implies that the caching of geotagged tweets may now be acceptable to Twitter.*

[3] *AJAX allows a page to request more data without reloading the whole page using JavaScript*

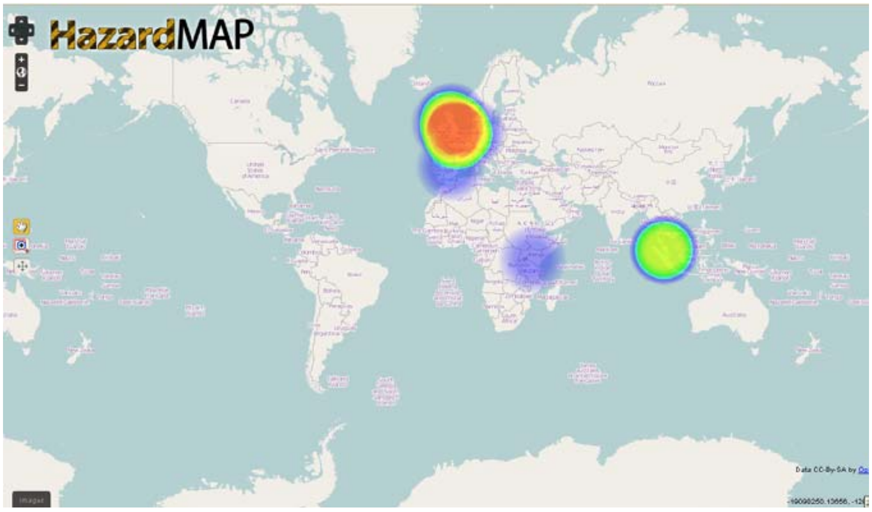

Figure 3: World view of heatmap displaying the results for geotagged tweets showing the word 'marathon' at 1000 GMT on 22 April 2012.

Interestingly, the global map also highlighted hotspots in Kenya and Malaysia. The team believed the hotspots in Kenya could be attributed to the fact that Kenyan athletes won both the men's and women's London Marathon races. The hotspot over Malaysia was deemed to be a result of the 'Energizer Night 42km Race' that took place the evening before. The fact that an endurance event had occurred in Malaysia over the weekend, even though the development team did not know about the event in advance, demonstrated the usefulness of using twitter in this way to display information.

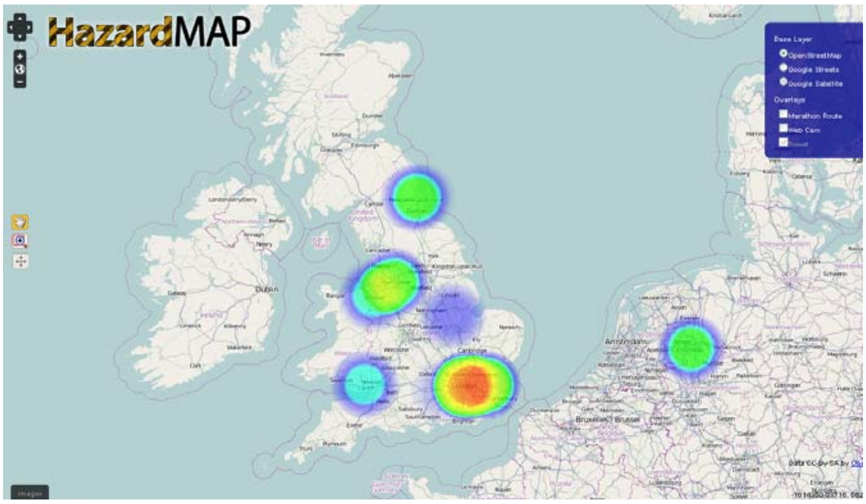

Figure 4: UK scale view of heatmap displaying the results for geotagged tweets showing the word 'marathon' at 1000 GMT on 22 April 2012.

At a national scale, the heatmap continued to highlight London as the dominate location for geotagged tweets for the word 'marathon' (Figure 4). However, when zoomed into London City, and the marathon route event, the heatmap appeared to become less useful, showing just three heat spots (Figure 5), a consequence of the small proportion of geotagged tweets that were observed.

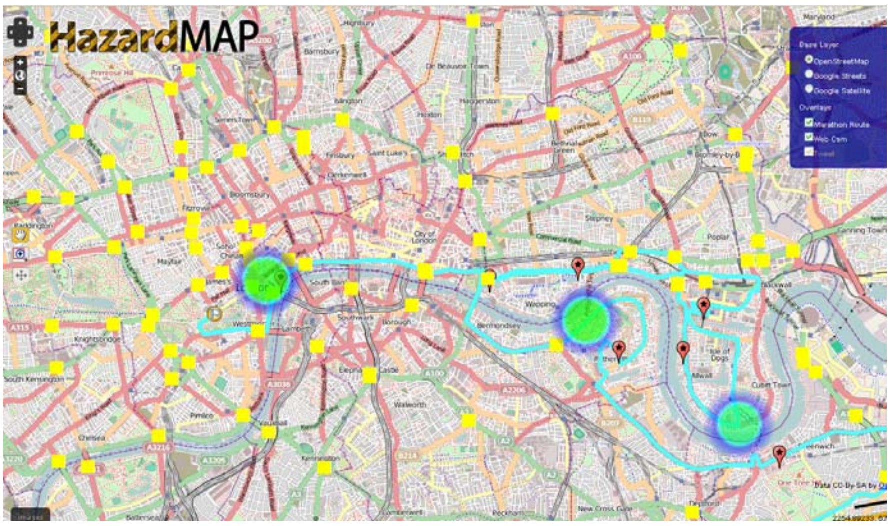

Figure 5: View of central London showing heatmap results. Yellow squares show locations of CCTV cameras. Turquoise line shows marathon route.

Despite the lack of tweets, the three spots appear to highlight three of the four key locations that the London Marathon suggest spectators watch the event (see London Marathon spectator guide) i.e. Bermondsey (Mile 11), Canary Wharf (Mile 18), and covered by one heat spot, The Embankment (Mile 25) and The Mall (Mile 26.2) at the end of the race. The heatmap, therefore, is potentially useful in confirming where the number of spectators watching the event is greatest.

The Hazard Map challenge suggested that it would be useful to use data harvested from the social media alongside data from trusted or official sources to obtain better situational awareness. In an attempt to demonstrate this, the Hazard Map challenge team decided to import a Web Map Service (WMS) feed of publically available CCTV camera locations within Central London to see if images could be obtained from the ground for where heat spots were being shown from twitter. Unfortunately, time only permitted the team to locate the CCTV cameras, not display their still images as hyperlinks (see yellow squares in Figure 5). However, this was enough to prove the concept in the time available.

A condition of the International Space Apps Challenge is that all code developed during the weekend is made open and available for others to use and develop. People can, therefore, access the results from the weekend through the International Space Apps Challenge website [4].

Northern Italy Earthquake 20 May 2012

A month after the NASA space apps challenge event, on 20 May 2012, a powerful magnitude 6.0 earthquake occurred in Northern Italy, in the Emilia Romagna region, just north of Bologna (Figure 6), leaving thousands of people homeless, and causing at least seven fatalities. Unable to do anything more to help those affected, the Hazard Map team decided to test whether the hazard 'map-lication' could have provided a valuable additional source of information about the impacts of the hazard to the authorities.

[4] http://spaceappschallenge.org/challenge/hazardmap-real-time-hazard-mapping-scraping-social/solution/33

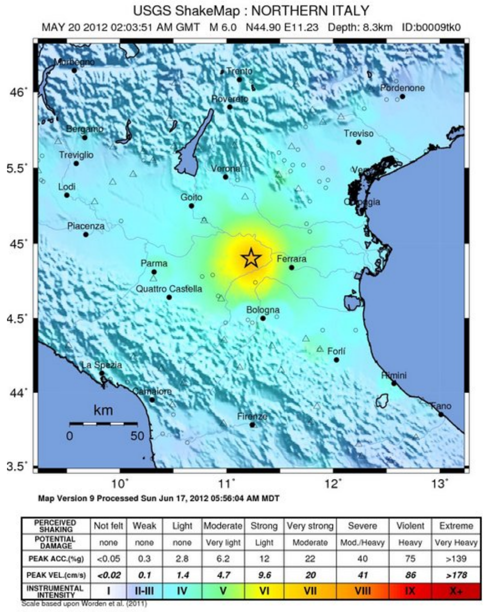

Figure 6: USGS Automated 'shakemap' showing the epicentre of the Northern Italy magnitude earthquake on 20 May 2012. (Source:USGS Shake Map)

Initially, the team searched twitter for 'earthquake' but the results were meaningless. However, once the team searched using the Italian word for earthquake i.e. 'terremoto' a useful image of the impacts of the earthquake was obtained (Figure 7). The actual tweets were also viewable alongside the heatmap, but have been left out of the figures to protect the users.

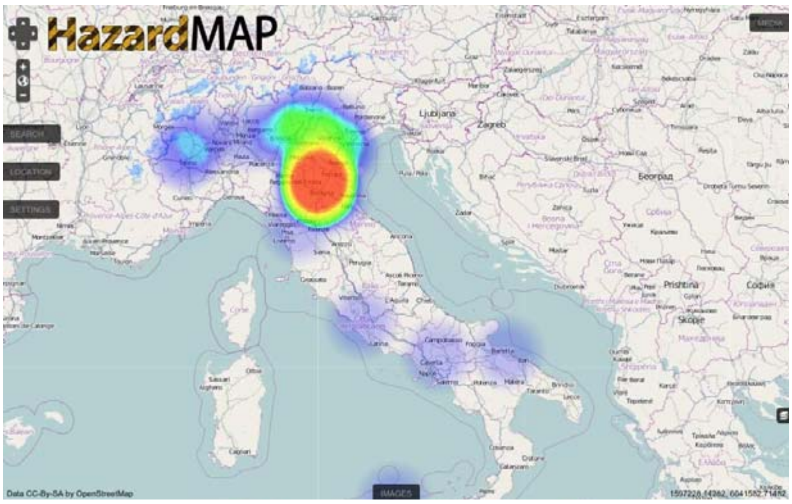

Figure 7: National scale heatmap of the word 'terremoto' search in hazard map on 21 May 2012.

Figure 7 clearly demonstrates a hive of activity in Northern Italy as people tweet about the earthquake, in some cases, simply just to say 'earthquake'. When visualised at a regional scale (Figure 8) and a local scale (Figure 9), the heatmap continues to be useful at pinpointing the hazard.

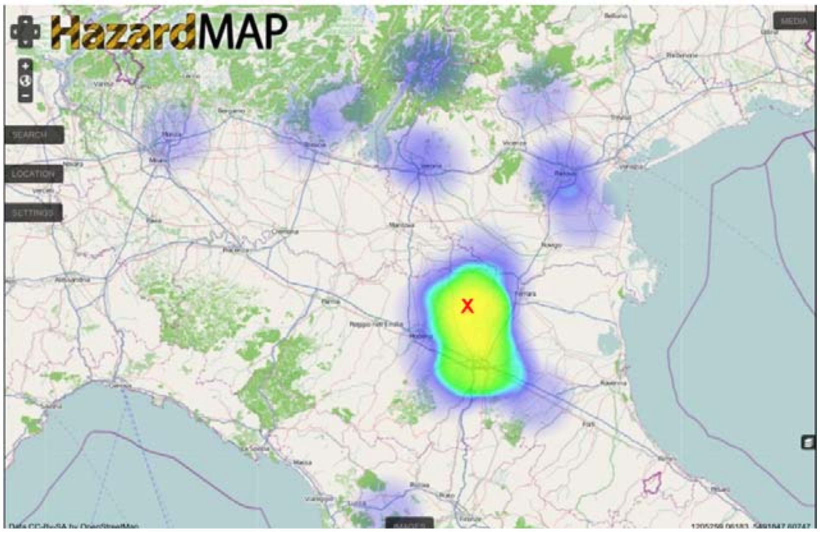

Figure 8: Regional scale heatmap of the word 'terremoto' search in hazardmap on 21 May 2012. The red cross marks the approximate location of the epicentre of the earthquake that occurred on the 20 May 2012.

The images shown in Figure 8 and Figure 9 clearly show that activity on twitter about the earthquake was greatest near to the epicentre and within population centres, with the town of Centro producing the most 'heat'. Although it is difficult to compare the heatmap against the actively crowd-sourced community intensity map produced by the United States Geological Survey (USGS) (Figure 10) because of scale and style of map, the heatmap does highlight the locations the actively crowd sourced map highlights. A more detailed analysis would be useful to see whether the passively sourced hazard map produced results more quickly than the actively sourced USGS community intensity map.

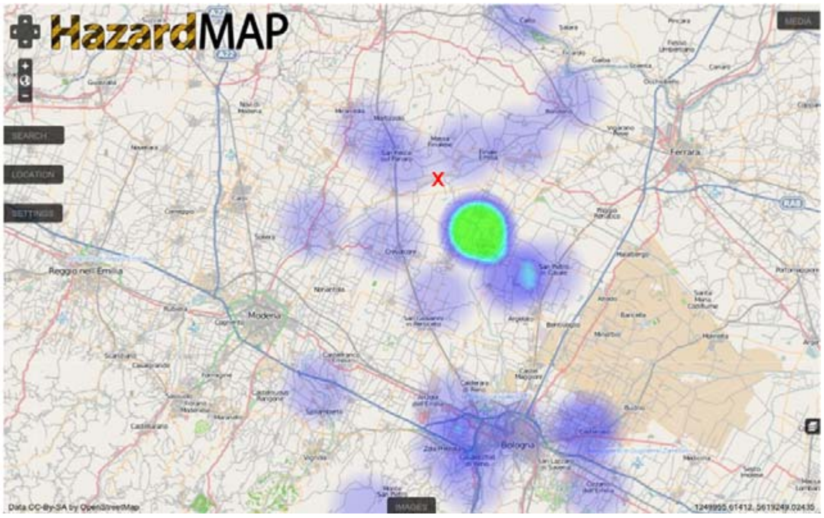

Figure 9: Local scale heatmap of the word 'terremoto' search in hazardmap on 21 May 2012. The red cross marks the approximate location of the epicentre of the earthquake that occurred on the 20 May 2012.

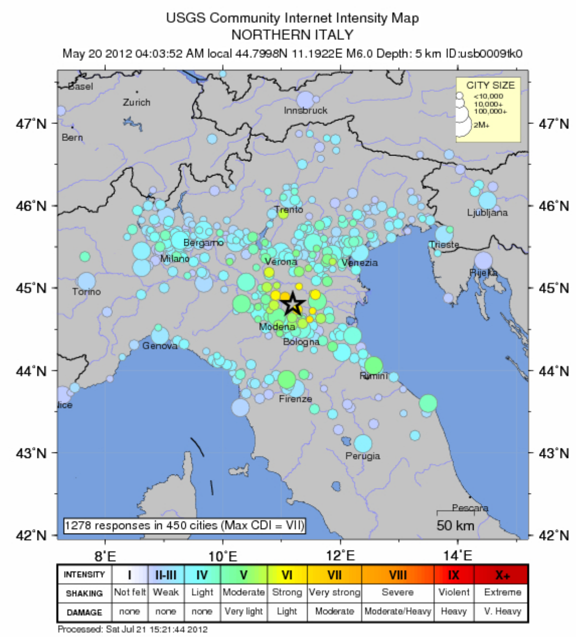

Figure 10: USGS community intensity map showing the epicentre of the Northern Italy magnitude earthquake on May 20 2012 alongside crowdsourced 'felt' reports. (Source: USGS Community Intensity Map)

Future Development

The Hazard Map challenge owners are exploring opportunities to develop the Hazard Map concept further. Some suggestions for future development are bulleted below.

- Twitter, unlike Facebook, lacks geo-tagged data. It would therefore be useful to develop the application so that it could interface with other social media sites with provide access to more geographically rich content.

- Hazard Map could be combined with other data, such as crowd sourced data, to provide a central source of 'on the ground' information during a crisis. A mobile solution of the application could be developed to support this use.

- Further development would be necessary to enable the Hazard Map to display harvested information about hazards automatically, perhaps from a catalogue, rather than to rely on a single word search string entered by the user. Any catalogue of words would need to be mutli-lingual or passed through a translator to ensure that the Hazard Map picks up all hazard information, irrespective of language.

- A more detailed analysis would also be useful to see whether the passively sourced Hazard Map generates results more quickly than the actively sourced USGS community intensity map.

- Development is required to rectify the data caching issue identified. A streaming API could provide a useful work around.

Conclusions

The Hazard Map's unique approach of using a heatmap to display point data has demonstrated that it could be a to be a successful method for displaying hazard information harvested from the social media. The case studies described here highlight the potential of this methodology for obtaining additional situational awareness during hazard events. However, one of the principle objectives of the Hazard Map challenge was to ascertain whether information could be harvested from social media to help verify scientific models attempting to model the likely impacts of natural hazard events. Without the ability to cache and store some of this data, the ability to use this wealth of information to help verify scientific models is unlikely. Despite this, the Hazard Map, with further development, could still provide a useful situational awareness tool for experts working within a hazard centre environment or for Non-Governmental Organisations such as Map Action or Shelter Box.

Acknowledgements

The authors would like to acknowledge the contributions made by the other Hazard Map team members Alex Ramsden, Pete Pojuner, Martin Sharpe, Robert Dunn and Aaron Huslage, each of whom contributed their expertise to the solution during the International Space Apps Challenge weekend, and in some cases, beyond. In addition we would like to thank the International Space Apps Challenge event team and UK Meteorological Office, and in particular Michael Saunby, for organising and hosting the codeathon event.

Emma Bee publishes this article with the permission of the Executive Director, British Geological Survey (BGS), Natural Environment Research Council (NERC).

References

USGS, Shake Map, Retrieved (8/8/2012), http://earthquake.usgs.gov/earthquakes/shakemap/global/shake/b0009tk0/)

USGS ,Community Intensity Map, Retrieved (8/8/2012), http://earthquake.usgs.gov/earthquakes/dyfi/events/us/b0009tk0/us/index.html

Twitter, Developer Rules of the Road, Retrieved (29/06/2012) https://dev.twitter.com/terms/api-terms

Felipe Barriga Richards Blog, OpenStreetMap Heatmap Using OpenLayers – demo for pubs and restaurants in London, Retrieved (8/8/2012) http://blog.felipebarriga.cl/programacion/openstreetmap-heatmap-using-openlayers/