AIR TWITTER: USING SOCIAL MEDIA AND SCIENTIFIC DATA TO SENSE AIR QUALITY EVENTS

E.M. Robinson and W.E. Fialkowski

Washington University in St. Louis

1. INTRODUCTION

The ability to easily expose content through the web using social media sites like YouTube, Flickr, Blogger and Delicious have given the Earth a "skin" of photos, videos and citizen reporting that enhance our understanding of our surroundings. Businesses are taking advantage of this constant stream of information by "listening" to the social-media chatter on the web. Social listening allows businesses to better identify their customers and provide tailored service to that group. News agencies are also using social listening techniques and have implemented sites like iReport, since it is more and more likely that citizen reporters will 'break' news stories and identify major events. Scientist can benefit from social listening as well. Community remote sensing can incorporate the new and evolving social media 'sensors' along with remotely sensed surface and satellite data to provide another dimension of contextual understanding about what is occurring in the natural environment.

2. AIR TWITTER

Air Quality (AQ) events such as fires and dust storms are highly visible and impact daily life, thus the pictures, videos, blogs and tweets are shared through web within minutes of the event occurring. Air Twitter is a social media listening tool that operates using a service-oriented architecture approach (SOA) (Fig. 1). User generated content is published on sites like twitter, blogs, Delicious and Flickr and described using terms like air quality, fire and smoke. Each of these sites allows a query to be made for the AQ terms and to specify the output of the query as an RSS feed. The feeds from multiple social media sites are aggregated using existing aggregation services and filtered to remove content that isn't relevant to outdoor air quality, like 'quality of Nike Air'. Thus allowing Air Twitter to find relevant outdoor air quality, user-generated content. Air twitter then binds to the content by tagging the filtered stream with #AirQuality.

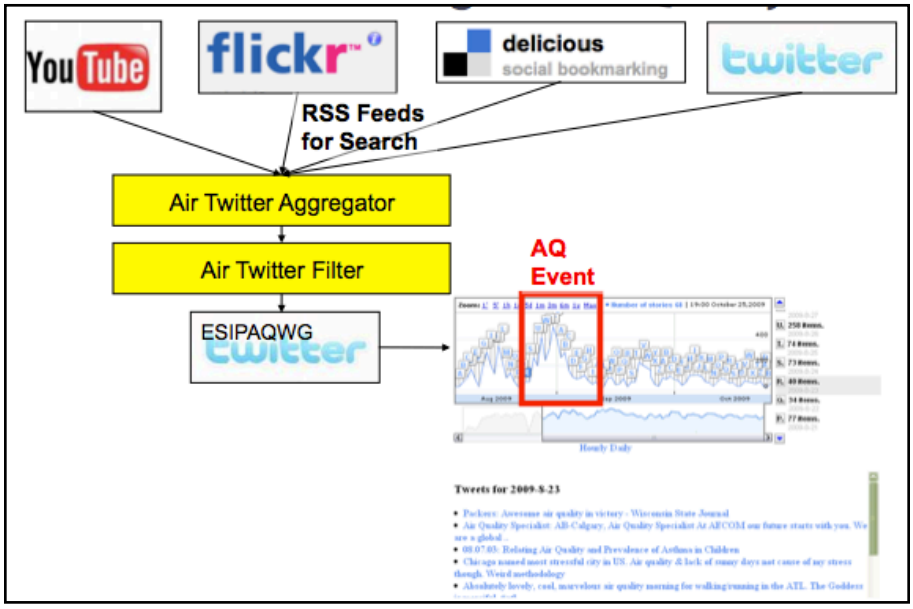

Figure 1. Air Twitter Information Flow

Through a separate twitter account for the ESIP Air Quality WG (@ESIPAQWG), the AQ relevant material is re-tweeted. A unique and unexpected outcome of this is that it has allowed a community of over 250+ people to follow this stream. Followers include Gov. Schwarzenegger and Boris Johnson, the mayor of London, as well as many local communities AQ agencies that publish their real-time surface monitoring data through Twitter.

3. AIR QUALITY EVENT IDENTIFICATION

The aggregated Air Twitter stream is also saved in a database, which allows time series of the number of tweets hourly and daily. Monitoring the time series (Fig.1) AQ events are identified from the background chatter about air quality. As the number of tweets increases, we click on the hour or day and see if there is a trending topic. The red box in Figure 1 highlights the increase in tweets seen in August 2009. Analyzing those tweets, we saw that most tweets were about the Southern California Fire. This event identification occurs hours to a full day ahead of event identification with only scientific data.

As the events are identified, collaborative EventSpaces (Robinson, 2008) are created using the ESIP wiki to collect the social and scientific information about the event. During the Southern California Fires (August 2009), an EventSpace was created within the first day of significant burning and updated throughout the entire period with relevant content (SoCal Fire, 2009). As science data was available, this EventSpace brought together relevant satellite data such as MODIS for true color images of the smoke, MODIS Fire pixels to identify fire locations and OMI Absorbing Aerosol Index to provide

additional evidence for the spatial extent of the smoke. Surface observations from EPA's Airnow were incorporated to identify the elevation of the aerosol and models such as the Naval Research Laboratory's NAAPS smoke models provided further validation for the given space and time to support AQ event analysis. All of the air quality data is accessed through the federated data system, DataFed (Husar, 2007) and displayed in the EventSpace using screencast timeseries to show the California Smoke Event evolution. Links to the data in multiple formats were also exposed. Within the EventSpace, the science data is combined with additional contextual data from the social media streams providing photos or webcams that are focused on the event, blogs about current conditions and news articles. For the California Fires, there was a webcam on top of the UCLA observatory providing half-hour images. The EventSpace captured many of these photos as a way to describe the event. Once the EventSpace has been created, the ESIPAQWG twitter account tweets the link to the relevant EventSpace wiki page every few hours in an effort to involve the public AQ community. During the California fires, the link was picked up by news agencies and there over 1000 views and was also shared through related groups on Flickr and CNN's iReport.

4. RESULTS AND FUTURE WORK

The EventSpaces are monitored using Google Analytics. During the August California Fires the traffic increased five-fold to the ESIP wiki (Fig.2a). Furthermore, the increase in traffic was entirely due to views of the SoCal Fire EventSpace (Fig. 2b). A top driver to the site was through tweeting the link to the EventSpace and having that link re-tweeted by others like the LA Times. An interesting an unexpected observation, was that most of the increased traffic was coming from Southern California (Fig. 2c). So the right people were finding the right information at the right time. The overall benefit of using the online community as an AQ event indicator, allows specific effort to be made for initial documentation of air quality events and the result is a catalog of events with some sparse analysis that can be followed-up.

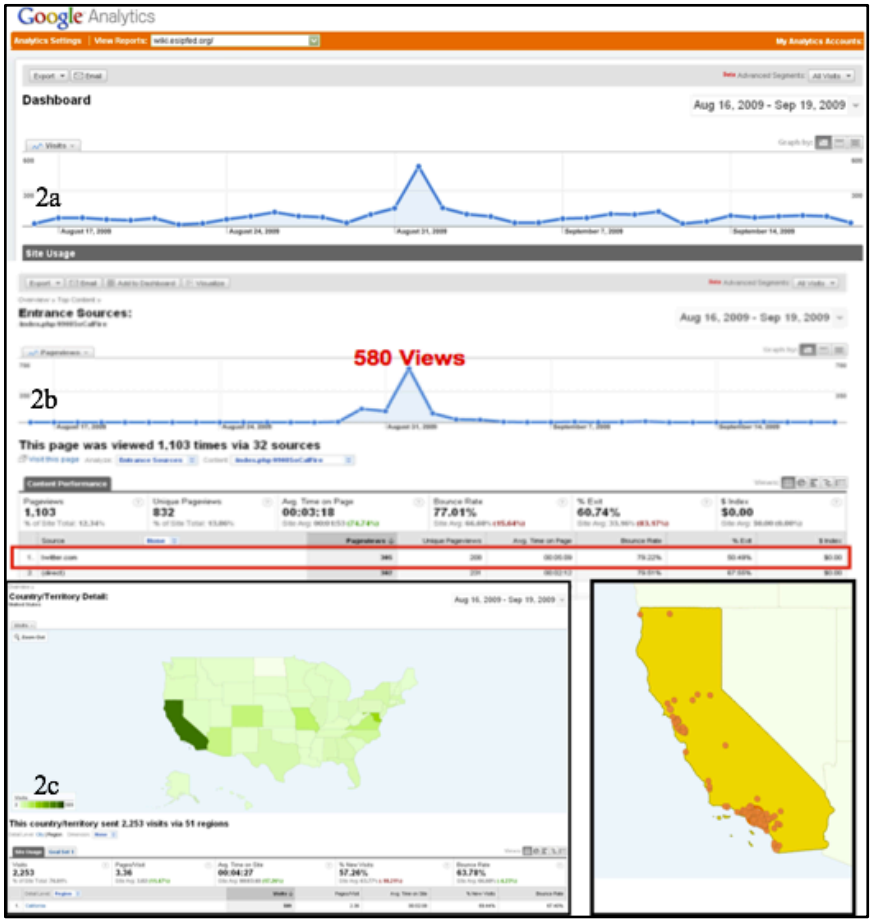

Figure 2a. Entire site traffic for ESIP wiki Aug. 16-Sept.16; 2b. Traffic to SoCal EventSpace driven from Twitter; 2c. Geographic location of traffic.

Future work includes reusing the technologies implemented for the air quality application for other environmental applications, such as drought. Additionally, the current method of identifying events is clumsy and could be improved by some method of georeferencing tweets in order to identify visually geographic hotspots. Finally, the concept of EventSpaces is a work in progress. Improvements are needed to better define relevant content, preserve the pages once the event is over and increase community contributions both from the general public as well as from scientist.

5. ACKNOWLEDGEMENTS

This work is supported by the Earth Science Information Partners Funding Friday Grant received July, 2009.

6. REFERENCES

[1] E.M. Robinson and R.B. Husar, "DataSpaces: Using Community Workspaces to Enable Rich Air Quality Metadata," AGU Fall Meeting Abstracts, 2008, p. 06.

[2] http://wiki.esipfed.org/index.php/0908SoCalFire

[3] R.B. Husar, K. H\öij\ärvi, and S.R. Falke, "DataFed: An Architecture for Federating Atmospheric Data for GEOSS," *IEEE Systems Journal*, vol. 2, 2008, pp. 366–373.