



Social Media Monitoring: Big Data Analytics
New Arrival Predictions: DRC Situation, Angola (update 3)
 The Innovation Service

ABOUT SOCIAL MEDIA MONITORING DATA ANALYTICS

UNHCR recognizes that big data can help improve its understanding of the protection environment. For this reason, the UNHCR Innovation Service started a sentiment analysis exercise together with UN Global Pulse to support the Europe Refugee Crisis - with an aim to provide decision makers with additional context to the situation as it unfolded. Part of the exercise was to determine how big data – in this case unstructured social media data – could be used to improve UNHCR's understanding of a complicated, and unique situation, by providing structured insights. More information is available in previous updates (1 and 2).

APPLICATION TO THE ANGOLA CONTEXT

A complex emergency is unfolding in the Kasai region in the Democratic Republic of the Congo (DRC). Individuals continue to seek safety in Northern Angola from inter-communal clashes, generalized violence, disorder and a shortage of basic items. Protection concerns and human rights violations have been reported. Most civilians in affected areas are at risk of serious human rights violations, including physical mutilation, killing, sexual violence, arbitrary arrest and detention.

The violence in the Kasai region in the Democratic Republic of Congo (DRC) which erupted in August 2016, has resulted in the displacement of large numbers of Congolese. It is estimated that 1.27 million people have been internally displaced. In addition, as of today, over 30,000 people are newly registered in Angola, having escaped violence in Kasai. Initially, the daily arrival figure into Angola fluctuated significantly with spikes in arrivals difficult to predict; the Angola operation began exploring solutions to this challenge, including manual social media monitoring (Twitter) for reports of violent incidents in the Kasai region. On this occasion, the machine-led monitor was set-up to help improve efficiency and strengthen the systematization of this monitoring in order to support the operation to better predict and prepare for new arrivals. More information is available on the machine-led data query in previous updates (1 and 2).

Due to the evolving dynamics of the conflict in Kasai, access to 'real-time' information on violent incidents - considered the key driver of displacement in this context - is critical. Access to the Kasai region by humanitarian actors is challenging - therefore identifying, testing and validated remote information sources (including SM) is a potential way forward in terms of a better understanding of the context.

The monitor was initially set-up in French and English; initial tests have begun in Lingala and Swahili (please see further details below).

ITERATIONS TO THE APPROACH

- Originally the monitor was established with relatively large geographical searches (including DRC / #DRC); this proved to be too wide in scope for the monitor - with many posts linked to violence in the Kivus. To better refine the monitor, the taxonomy was revised to a smaller geographic scope based on current arrivals data from the Angola Operation. This immediately filtered out the 'noise' of posts not related to violence in the Kasais.
- Given the common use of variances in spelling and use of accents for place names, there was an identified risk of missing some posts. In order to address this, the taxonomy was adjusted to include geographic spelling variances for each location.
- Based on an initial review of the identified posts, it was clear that retweets were 'creating noise' - the timing and scale of the incident could not be determined. To address this, retweets were filtered out, the monitor now only returns original tweets, however this does not filter out directly copied tweets (those without RT @). As this is the first time a retweet filter has been applied, several attempts to remove 'RTs' was required before this was successfully established.
- From 21st June onwards an in-depth analysis of the filtered 'specific' category was undertaken; this has improved the accuracy of the machine-led categorization.



Social Media Monitoring: Big Data Analytics
The Innovation Service

NEW ARRIVAL PREDICTIONS: DRC SITUATION, ANGOLA

May 1st, 2017 at 11:00 CET and onwards

Source: Crimson Hexagon

Language(s) monitored: English, French

Geography monitored: Global

Type: Opinion Monitor

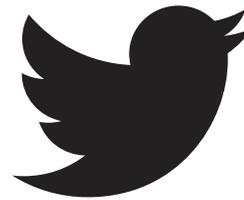
Posts Analysis

- Total posts: 11,416 (May 1st - July 11th 2017)
- Specific Incidents: 39% of relevant posts
- Changes in post volume (how the overall volume trends are changing over time): 110% increase
- Total posts during reporting period: 820 (June 22nd to June 30th) of which 100% were Relevant Posts.
- Specific incidents during reporting period: 64%

Posts Data Analysis

- In the period from 1st July - 11 July (since the last update on 30th June), the vast majority of specific incident posts (totaling 515) were related to: a court case ruling convicting a group of soldiers [or militants] for killings in Kasai.
- In the reporting period from 1st July - 11 July, there was a notable absence of violent/security incidents being reported.

This summary is for posts in French and English only.



CATEGORIES:

- General commentary: this category refers to general updates related to the violence in Kasai, for example detailing cumulation of incidences, deaths and events to date.
 - Specific Incidents: this category refers to a time and location bound report of an incident.
-

MONITOR ASSUMPTIONS:

- Specific incident reporting is 'relatively real-time' as indicated by the inclusion of date/times within Tweets, which correlate to the date and time stamp on posts.
- Many reports do not originate from eyewitnesses themselves, but from media agencies (incl. Radio Okapi, Actualite CD) - the assumption is that these agencies are sharing information received from primary sources/stringers.
- Due to the proliferation of copied tweets (not RTs) - either human or bot driven - it is necessary to drill down into the data to identify the initial tweet (by time-stamp) and source; in some instances this is not possible as the time stamp is limited to minutes as the smallest unit of measurement. Simultaneous posting has been observed by media agencies (e.g. Actualite CD and Afrique Live, Le Voix de l'Afrique au Canada), the assumption is these agencies share a common primary data source.
- There is no certainty on how representative social media posts are in terms of socio-economic diversity and demographics (e.g. twitter represents those with access to connectivity)
- The georeferenced methodology is tied to Crimson Hexagon software queries (10 last tweets = geography). Some of the users might be located elsewhere but using a foreign VPN for connectivity; this has been observed for DRC based news reporters whose geo-location indicates they are in France.
- The algorithm is biased to the proprietary software (Foresight) parameters for searching and sampling tweets. The sample is a minimum of 10,000 posts containing the taxonomy words.



Social Media Monitoring: Big Data Analytics
The Innovation Service

ARRIVALS DATA:

Arrivals data has been provided by the Angola operation (from beginning of 2017 to present). The reporting methodology was adapted from 20th June onwards, due to a decrease in new arrivals and changing arrival points. Since this date, pre-registration is not being conducted at the border and individual registration is taking place - with continuous registration due to start from July. The number of arrivals to Angola from DRC has significantly dropped.

Looking at the arrivals data from 1st May - 20th June: there is a weak *negative linear relationship* (-0.34) when correlating the daily values of both variables (# of incidents & # of arrivals). As the Angola team had reported a lag of a few days (2-4 days) between reported incidents and an increase in arrivals, a correlation between the variables with a three day lag was also completed. This involved off-setting the arrivals to three days *after* the incident reporting and re-running the correlation. A weak *negative linear relationship* was also noted (-0.30). Further information on this correlation is available in previous updates (1 & 2).

At this stage, this correlation exercise does not provide conclusive evidence to inform arrival predictions.

Due to the significant reduction in arrival numbers, using the SM monitor for arrival predictions is no longer a priority. As the testing phase has generated some interesting *qualitative* information related to how incidents are reported, and by whom. It also provides an additional source for data related to reported protection concerns; with regular monitoring and analysis this can complement existing protection monitoring data. UNHCR Innovation is coordinating with IM colleagues in DRC and the Global Protection Cluster (GPC) to determine if the data generated is of use for staff and partners responding in Kasai.

PRELIMINARY RESULTS & RECOMMENDATIONS:

Social Media Usage:

Less than 1% of relevant posts were sourced from Facebook - there could be a number of factors contributing to this: either Facebook is not the preferred information sharing channel for this type of information or, information is not being shared within the public forum. This needs to be further examined to understand if/how such information sources could be accessed.

A social media monitor was established in Swahili and Lingala in collaboration with Translators Without Borders who provided translation of the 20 key search words. Excluding RTs there were only 53 posts for the 1st May - 11th June period. On initial analysis only 3 were in Lingala (with French) and a couple of Italian posts (anomalies) were returned. The 53 posts are currently being translated by Translators Without Borders. An additional monitor in Kasai language is planned. Despite the low use of Twitter in local languages further analysis is required to determine the relevance and application of *qualitative* information available in local languages.

Recommendations:

- UNHCR Innovation will run a monitor in Kasai language and coordinate with Translators Without Borders to ensure that Swahili, Lingala and Kasai translations are included in information shared with Angola and DRC operations.
- Further analysis/understanding of the use of Facebook for sharing sensitive, incident related information needs to be conducted; potentially through outreach to security analysts to better understand their data sources, and in consultation with communities (a rapid Information and Communications Needs assessment) to better understand Facebook usage.

Primary and secondary sources:

The majority of posts - particularly those that are 'copied' - are generated by media agencies and/or journalists affiliated with these agencies. These journalists/agencies will have a comprehensive information-gathering network, which appears to operate relatively 'real time'. To understand more about the information eco-system and incident reporting in DRC, UNHCR Innovation sought to engage these key sources by researching further contact details (Direct Messaging, emails, phone numbers and websites). Engagement included following on Twitter (from UNHCR Innovation account) and reaching-out through direct messages on Facebook. To date, these sources have provided no response.



Social Media Monitoring: Big Data Analytics
The Innovation Service

- **@soniarolley** (The permanent special envoy for RFI, based in Kinshasa); Sonia is often tweeted at with information and updates, she has also been 'copied' by several other news agencies and individuals. 177,000 Twitter followers to date: <https://www.facebook.com/sonia.rolley>
- **@actualitecd** (A specialist DRC news agency); often copied and tweeted at - in addition **@bujakeratshiams** a journalist with Actualite CD can be considered as a key source. 210,100 and 826 Twitter followers respectively, to date. <https://www.facebook.com/actualiteCD/> +243 999 136 373 & email info@actualite.cd
- **@radiookapi** (The Radio for Peace, developed in partnership with MONUSCO and Fondation Hirondelle); the frequency of Radio Okapi's posts is notable - 310,000 Twitter followers to date. <https://www.facebook.com/radiookapi/> & +243 818 906 316 & email contact@radiookapi.net
- **@AfriqueLive** (A 24 hour English and French Language news service) - 417 Twitter followers to date. <https://www.facebook.com/Afrique-Live-448790305184106/>
- **@jeanluckienge** (an Ontario based reporter with La Voix de l'Afrique au Canada) - 7,110 Twitter followers. Email: jlkiengenews@gmail.com
- **@InfoDiasporaRDC** (Limited information, no geolocation information available) - 3,029 Twitter followers.
- **@RadioCongoShare** (An Internews supported radio, no geo-location available) - limited number of follows (104). <http://radio-congoshare.net/>

During the research a relevant FB group was also identified:

<https://www.facebook.com/groups/773178522719044/permalink/1295160833854141/>

Discussions in this group are in French (majority), Swahili with the occasional Lingala post. Further analysis of the information being shared within this group is required. UNHCR Innovation will explore ways to systematize this analysis; participation in the group has recently been approved.

Recommendations:

- Given the access these agencies/individuals have to direct primary data of incidence reporting in the Kasai region, follow-up (continued engagement) with these sources to best understand their information eco-system is recommended.
- Monitors for these accounts to be established, and further investigation into additional key, trusted sources.
- Outreach to the DRC UNHCR operation and GPC to understand if/how they have previously engaged with these sources.

Protection Monitoring Information

The information being shared (including media links/images) details significant protection concerns in DRC. This qualitative information can complement UNHCR's own protection monitoring, so the response can be better tailored to meet the protection needs of new arrivals in Angola. This information can also be used to support UNHCR's external advocacy and response planning inside DRC. As detailed above, this week saw a marked reduction in the frequency of incidents reported in Kasai. The operation in DRC confirmed during an internal Task Team call that there has been a reduction in security related incidents recently. That this is 'corroborated' by SM reporting helps build confidence in the SM data set - there were no reports of 'false' incidents during the reporting period (1st July - 11th July). Of the specific reports, the following topics were included:

- The conviction of 8 [on occasion 7] soldiers [on occasion militants] related to violence in Kasai. This was reported in English and French: Eight DR Congo soldiers get stiff jail terms for Kasai killings <http://ow.ly/i7sD50cDrGt> & #lemondefr: En #Rdc, huit militaires condamnés à lourdes peines pour un massacre au Kasai.
- Reports on internal displacements in DRC: #RDC #Kasai À Idiofa (Kwilu), entre 80 et 90 déplacés fuyant les Kamwina Nsapu arrivent chaque jour. Ils sont maintenant plus de 3000.

Recommendations:

- Information is frequently shared (weekly) with protection colleagues in Angola, DRC and the UNHCR Africa regional bureau for operational follow-up.



Social Media Monitoring: Big Data Analytics
The Innovation Service

DEMOGRAPHICS:

Of all posts, 4152 (36%) have an identifiable gender, 23% of posts were identified as female while 77% of total number of post were identified as male. Of the total, 1,675 posts (15%) had an identifiable age; 93% of which (1,562 posts) were in the '35 and above' age group. 6,847 posts have an identifiable geographic location with the following geographical distribution: 1) France (19.8%), 2) DRC (16.71%), 3) United States (11.22%), 4) Canada (6.65) and 5) Belgium (6.24%).

As of July 11th, 2017 the affinities of posts differed from the rest of the Twittersphere, with affinities in the rest of Twitter strongly with the soccer and fashion; whereas within this specific monitor, strong affinities were made with Africa (39.3%) Rwanda (29.4%) and Afrique (28.4%). This may or may not be a signifier of influential words, or organizations that change, or could be used to change opinions.

EXAMPLES OF TWEETS FOUND:



Specific Incidents - Tweets

@RadioCongoShare: #Voxcongo du 07/07, Condamnation des militaires accusés de massacre au #Kasai saluée par le Gvt, le BCNUDH et la SC <https://soundcloud.com/voxcongo/vox-congo-edition-du-07-07-2017-swahili>
...

@CongoActu: Massacre au Kasai: Sept Militaires Condamnés à Des Peines Allant De 15 Ans De Prison à la... <https://congoactu.net/massacre-kasai-sept-militaires-condamnes-a-peines-allant-de-15-ans-de-prison-a-perpetuite/> ...

General Commentary - Tweets

@TSHIKAPA01: Le corps de Chebeya, Les fausses communes du kasai et Les massacres de Béni hantent jkabila jusqu' à la folie <https://m.youtube.com/watch?v=k3yTgD-VoLY> ...

@Africannews7: The 21st Century and the Devil's Cruelty. DR Congo: UN Experts to Investigate Kasai Region Violence - <http://tinyurl.com/ycy2ffku>

@Nkumunue: The situation of D.R.Congo is not good at this time, we should act to end the instability and atrocities in the country specially in Kasai ...