"Social Media and Forced Displacement: Big Data Analytics & Machine-Learning"

White Paper
September 2017

UN GLOBAL PULSE | UNHCR INNOVATION SERVICE

TABLE OF CONTENTS

ACKNOWLEDGMENTS

_____

SUMMARY

This white paper summarizes the initial findings and lessons learned from a project conducted by UNHCR's Innovation Service and UN Global Pulse[1] to inform on the viability and value of social media analytics to complement understandings of the Europe Refugee Emergency.

Ongoing conflicts and violence around the world[2] led over 1.4 million people to seek refuge in Europe between 2015 and the first part of 2017.

Data from social media offers a wealth of information that can be parsed to better understand what people think, and how people feel about things affecting their lives, such as the displacement and movement of large volumes of people. Researchers in turn, can use this data to inform topics of interest; decision makers can use such data as evidence on which to inform for example, programmatic responses and alterations.

The paper outlines the process, questions and methodology used to develop the project and presents preliminary observations on how aspects of the Europe Refugee Emergency are related on Twitter. The paper describes ten *quantitative social media mini-studies* that were developed as part of the project.

The project team initially set out to explore the value of social media both for monitoring Persons of Concern's (PoC), sentiment towards the provision of services, and their interactions with service providers[3]. However, based on inconclusive initial results and anticipating an increase in negative public views towards PoC following the 2015-2016 terrorist attacks in Europe, the project refocused on the analysis of host communities' sentiment towards PoC in reaction to incidents taking place in different European countries. Findings revealed that within local active Twitter communities, a small number of people connected PoC and the different terrorist attacks.
Being able to assess peoples' views in real-time provides a unique opportunity for UNHCR to counter non-conducive behaviour online. It also allows the Agency to better understand generalized perceptions vis-a-vis longer-term solutions for PoC.

The processes detailed herein are intended to serve as examples and to inspire other agencies looking to use social media and data analytics to inform on decision-making processes, operational responses and policy development in emergency-related contexts.

BACKGROUND

The Europe Refugee Emergency was a constantly, and rapidly-changing context. Ongoing conflicts and violence around the world[4] led over 1.4 million people to seek refuge in Europe between 2015 and the first part of 2017. This included increasing numbers of families, women, and unaccompanied and separated children—some seeking to reunite with other family members already in Europe. This new movement was challenging for many organizations, including UNHCR; people moved quickly, often across several international boundaries in very short periods of time; sometimes encountering changing protection risks, particularly when legal practices evolved, when borders closed, or when alternative routes begin to develop.

According to a report released by Social Media for Good[5], social media monitoring can provide significant value to decision makers in such dynamic contexts, where humanitarian access is poor, the information landscape fragmented, and social media widely used. For example, UNHCR's report "From a Refugee Perspective" portrays the discourse of refugees and migrants and the use of social media.[6] Social media platforms are powerful communications tools for humanitarian organizations, both at a strategic corporate level and an operations level to directly interact with affected communities. They also contain a wealth of information that can be parsed to measure and monitor conversations and emerging narratives.

Further, sentiment analysis of social media content can be used to capture public perceptions of an organization and its activities in a particular context to not only help develop new strategies, but also to ensure that existing programmes and projects are re-aligned and course-corrected in real-time. Several pilot and research projects have shown the feasibility of using social media data to crowdsource topics of relevance to sustainable development and humanitarian action. However, there has been little effort in extending the quantification of online sentiment to inform on interactions between PoC and services providers. Similarly, organizations would benefit from understanding how host communities view PoC on social media to inform their decision-making processes.

UNHCR currently uses social media for two main purposes[7]: 1) to publicly portray the Agency's work, and digitally engage with public audiences; and 2) to communicate with affected communities (CwC)[8]. UNHCR has a strong influence among different audiences on platforms, such as Twitter, Facebook, and Instagram, and it has clear guidelines on the use of these platforms for communication purposes. While the use of social media for CwC-supported activities is relatively new to the Agency, there are many promising efforts underway, both at Headquarters, and in field operations.

PROJECT OVERVIEW

The work described in this paper was initiated by the UNHCR Winter Operations Cell[9] and UNHCR's Innovation Service in November 2015. UNHCR recognized that big data analytics could provide additional insights into understanding the protection environment within the Europe Refugee Emergency. However, it did not have vast in-house knowledge, skills, or the necessary tools to conduct large-scale analyses. Therefore, it was limited in its ability to feed potentially valuable information contained in big data into operational responses.

To validate the value of social media data in emergency situations, UNHCR's Innovation Service partnered with UN Global Pulse in January 2016. UN Global Pulse provided technical guidance, coaching and tools for the project. The joint-collaboration explored how alternative sources of data can and should play a role in pursuing humanitarian outcomes.

The project team identified two **opportunities** in which social media could be harnessed to better understand the Europe Refugee Emergency:

O₁: Monitor interactions between PoC, and between service providers and PoC, in an aggregated form, and;
O₂: Understand the sentiment of PoC, host communities, and communities through which PoC have transited, in aggregated form.

The project envisioned a near-real-time monitoring system that could inform operational responses in support of the Europe Emergency Regional protection strategy. This system would have a two-tier architecture, with a machine learning component in the

backend that would process and classify social media posts according to predefined categories (e.g., posts related to abuse, or of xenophobic nature), and an information visualization interface that would enable UNHCR staff to routinely monitor, and analyze relevant social media feeds in six different languages: Arabic, Farsi, English, Greek, German and French.

To inform the feasibility of this system, and to ensure the opportunities identified were substantial, the project team iteratively conducted a series of ten *quantitative mini-studies* using Twitter posts.

A COMPENDIUM OF MINI-STUDIES USING SOCIAL MEDIA DATA

The ten studies were divided into two main iterations. For each of the iterations, a methodology largely inspired by the Harvard Data Science curriculum[10] was used. The following sections detail the data and tools that were employed, discuss the main hypotheses, and share the general iterative procedure

Data and Tools

Twitter posts, or *tweets*, are mostly public expressions of ideas and opinions[11]—as opposed to Facebook posts, which are mostly private. As of 2014, only 5.1% of Twitter accounts are protected[12]. Therefore, given that the majority of Facebook posts are private, and potentially a PoC could be the one expressing an opinion, the project chose tweets as the main source of data, complying with UNHCR's data protection policy[13].

UN Global Pulse has a long-term research partnership with Crimson Hexagon that allowed the project team to use the company's ForSight tool to access and analyze social media posts. Crimson Hexagon provides an online social media monitoring platform that enables users to create *monitors*, which have built-in machine learning capabilities to semi-automatically classify and extract sentiment from posts. These capabilities are based on algorithms that are iteratively improved using a *training dataset*, i.e., a curated collection of posts that helps *train* the monitor to correctly interpret any new incoming post[14].

The three main steps in setting up and training a monitor are: 1) *defining a taxonomy* to identify the keywords, hashtags, and phrases that will help retrieve the most relevant posts from the social media platform of interest (e.g., Twitter), 2) *formulating a query* using those terms to retrieve the posts; and 3)

manually classifying an initial subset of the retrieved posts to establish the training dataset. Once a monitor is trained, it provides different views of automatically classified posts, as well as of the sentiments extracted from them, which enable users to conduct a variety of quantitative analyses. While the project described in this paper was implemented with the ForSight tool, the methodology is generic and can be executed with other technological solutions.

Queries and Taxonomies

Formulating appropriate queries is not always straightforward, and a significant amount of effort can be put into training a monitor before an inadequate query is detected. This is an iterative process that involves a certain degree of trial and error. For example, selecting the appropriate vocabulary can be difficult. Tweets abound with colloquial language and "internet-speak"—Arabic slang typically varies across countries and regions, and can be written in either Arabic *abjad* (which is *dextrosinistral*), or using the Roman alphabet (e.g., ArabEasy, which is *sinistrodextral*). The 140-character restriction on Twitter also encourages word abbreviation. As a rule of thumb, a lack of relevant posts can be indicative of a poor query, or a very restrictive combination of keywords based on the use of logical operators (AND/OR/NOT).

In addition, certain assumptions must be made regarding *tweeters'* specific knowledge of the topic of interest. For example, a premise of this project is that, in general, people who tweet have little to no knowledge of the legal and protection differences between *migrants* and *refugees*. Both terms were used as synonyms in the queries, even though they have different implications for UNHCR. Contrary to migrants, refugees are specifically defined and protected by international law, particularly regarding refoulement[15].

Finally, queries can be restrained in space and time. These two dimensions can be helpful for further bringing out the voice and opinions of for example, PoC vs. host communities. Geo-referencing of social media posts[16] can be done based on a combination of the location declared by the user in his/her profile, and the latest location(s) from where s/he posted.

Classification

The classification process first requires determining a set of relevant categories, in which the queried posts will be filed. In initial explorations, the project team found that simple dichotomous categories are most effective, like *racist–non-racist*, or *positive–negative*.

Categories for *irrelevant* and *neutral* posts are also useful, since all posts may not fit into the dichotomous pair, either because the content is inapplicable, or because it is incongruous[17]. All the categories determined in the mini-studies are presented in Annex I, along with their respective queries.

An initial subset of posts must then be filed manually to create the training dataset (or *training tweets*), which the monitor's underlying algorithms will use to automatically classify new incoming posts. This involves personal judgment as to whether content is relevant or not, and can turn out to be a lengthy procedure. Typically, the more categories there are, the more posts there are that need to be read, and manually sorted.

Iteration 1

The project team conducted six mini-studies in this iteration. For each study, a unique monitor on Crimson Hexagon was created (see Table 1).

Table 1: Initial Monitors Overview

Monitor	Unit of Analysis: Geography	Unit of Analysis: Timeframe	Language	Number of Posts*	Identified Opportunity
1. Interactions Arabic	Greece	February 1, 2015– April 18th, 2017	Arabic	6,341	O_1
2. Interactions Farsi	Greece	February 1, 2016– April 18th, 2017	Farsi	1,483	O_1
3. Xenophobia Greek	Greece	June 1, 2015– April 18th, 2017	Greek / Greeklish (latin chars.)	248,691	O_2
4. Xenophobia English	Greece	June 1, 2015– April 18th, 2017	English	26,466	O_2
5. Xenophobia Arabic	Greece	February 1, 2015– April 18th, 2017	Arabic	196	O_2
6. Xenophobia Farsi	Greece	February 1, 2016– April 18th, 2017	Farsi	160	O_2

* Number of post analyzed by the machine to the date: April 18th, 2017.

Source: Crimson Hexagon ForSight tool

Hypotheses

The assumption for O_1 was that analyzing social media posts could provide insights into, for example, altered routes, or the conversations PoC are having with services providers, including smugglers; and that this could provide better situational awareness for decision making, and thereby better inform the orientation of resource allocations, and advocacy efforts.

The **hypothesis** for O_1:

O_1H: Monitoring interactions will reveal behavioral patterns and intent of PoC with regard to service provision and access to territory and asylum. This can inform UNHCR programme design and planning strategies for Strategy Objective 1 of the Europe Emergency Regional Protection Strategy: *Access to territory and asylum is safe*.

The assumption for O_2 was that understanding the sentiment that host communities express on social media could help identify pockets of, for example, xenophobic attitudes towards PoC; and that this could help UNHCR to improve the conditions in which durable solutions may occur, by better targeting corporate communications, and advocacy-related actions around legislation in specific countries.

The **hypothesis** for O_2:

O_2H: Understanding PoC and host communities' mutual sentiments will reveal how both groups view and react to asylum conditions and protection. This can inform programme design and planning strategies for Strategy Objective 3 of the Europe Emergency Regional Protection Strategy: *Access to protection systems and durable solutions are reinforced*. This will also provide a baseline for responses, their adjustment, and their possible improvement.

10   Blitzstein, J. & Pfister, H. (2015). The Data Science Process. Harvard Data Science
11   Page, C. (2014). Twitter has almost 430 million inactive users. The inquirer.
12   Idem
13   UNHCR (2015). Policy on the Protection of Personal Data of Persons of Concern to UNHCR. RefWorld.
14   SAS (2016). Machine Learning. SAS Institute.

15   Expel or return a refugee to the territories where her/his life or freedom would be threatened on the account of race, religion, nationality, membership of a particular social group or political opinion. UNHCR (1977). Note on non-refoulement.
16   Crimson Hexagon FAQ: How does Crimson base its geographical data.

17   See Annex II for a detailed classification of racist, non-racist, neutral, and irrelevant tweets.

Setup

The flexibility of language use on social media requires native speakers to query and classify the training tweets. Native speakers alone can understand the semantic nuances, colloquial, and even unusual uses of local language, as well as typical abbreviations discussed above. The project team relied on a group of English, Farsi, Arabic, and Greek native speakers, all of whom have basic knowledge in computer programming. This paper will refer to the team of native speakers as the *monitor trainers.* The monitor trainers were coordinated by UNHCR's Innovation Service's Data Scientist, and overseen by UN Global Pulse's technical team.

The project team concentrated on Greece for the first iteration, as the country has played many different roles throughout the Europe Refugee Emergency—it is now a host country for a static population of refugees and migrants. Based on UNHCR data from the Europe Regional emergency—which includes demographic population data—it was assumed that PoC would largely be Arabic or Farsi speakers, and that posts from Greece in these languages would likely be those of PoC.

To address **O$_1$H**, the trainers set up two specific monitors to track interactions between PoC, service providers, and the general public regarding access to territory, asylum conditions, shelter conditions, transportation, and movement in Greece. The first monitor was set up for Arabic, and was called **Interactions Arabic**. The second monitor was set up for Farsi, and was called **Interactions Farsi**. A full description of the categories and queries used is provided in Annex I. Tweets were monitored from February 1st, 2015 or February 1st 2016 to April 18th, 2017 (end of the study). February 2015 corresponds to a period of major influx of PoC in Europe. February 1st, 2016 is approximately one month prior the EU-Turkey Agreement.

To address **O$_2$H**, the trainers set up four monitors to track negative sentiment and perceptions, like xenophobic, discriminatory, or racist sentiments, of host communities towards PoC for Greek, English, Arabic, and Farsi. The monitors were respectively called **Xenophobia Greek**, **Xenophobia English**, **Xenophobia Arabic**, and **Xenophobia Farsi** (see Annex 1). Tweets were monitored for the period February 1st, 2015 or February 1st, 2016 to April 18th, 2017.

To build the Xenophobia monitors (**O$_2$H**), the project used the following categories to classify the tweets:

Xenophobic: tweets that express negative attitude, prejudice, or hostile sentiment that vilifies PoC;
Non-Xenophobic: tweets that express explicit support, positive attitude, or friendly sentiment towards PoC;

Neutral: tweets that describe facts about PoC (for example, news articles) but that do not express a strong sentiment or any sentiment at all;
Irrelevant: tweets that are not related to PoC.

The monitor trainers identified posts as belonging to the xenophobic category based on the UNESCO[18] definition of xenophobia: "*xenophobic behavior is hostility based on existing racial, ethnic, religious, cultural, or national prejudice*"; and the UN Fund for Contemporary Forms of Slavery, OHCHR, declaration definition of xenophobia: "*attitudes, prejudices, and behavior that reject, exclude and often vilify persons, based on the perception that they are outsiders or foreigners to the community, society or national identity*"[19].

The trainers also distinguished between factual, opinion-driven, rumor-driven, and breaking news tweets, in order to adequately train the machine for the neutral category. They further subtracted re-tweets (RT) from certain queries, following the findings of Mendoza *et al*[20], especially for the monitors related to **O$_2$H**, to avoid 'inflating' the number of xenophobic posts.

Insights

The **Interactions Arabic** monitor was successfully trained but did not retrieve a large number of relevant tweets (<7,000—see Table 1). Annex II portrays a subset of these posts. The **Interactions Farsi** monitor, however, could not be trained, due to an apparent lack of tweets in Farsi (<1,500) regarding access to territory, shelter conditions, and transportation. These results did not provide enough data to confirm or refute **Q$_1$H**, and could indicate that PoC—assumed to be either Arabic or Farsi speakers—a) simply do not use Twitter to inquire about, complain, or request services; b) do not have access to Twitter; or c) prefer other communications channels. The latter two possibilities seem further supported by the **Xenophobia Arabic,** and **Xenophobia Farsi** monitors, which also retrieved a very low number of tweets (<200 and <160, respectively).

The analysis also showed it is difficult to systematically separate tweets coming from PoC, host communities, and the general public for further analysis. Only few tweets described access to territory in Europe—including closing borders and entry restrictions—asylum conditions, and the economic challenges encountered during, and at the end of their journey, while many expressed the sentiment of host

18   UNESCO (2016). Xenophobia. Learning to Live together. International Migration.
19   OHCHR (2011). Declaration on Racism, discrimination, Xenophobia and Related Intolerance against Migrants and Trafficked Persons. Asia-Pacific NGO Meeting for the World Conference Against Racism, Racial Discrimination, Xenophobia and Related Intolerance. Teheran, Iran
20   Mendoza, M., Poblete, B. & Castillo, C. (2010). Twitter Under Crisis: Can we trust the RT? 1st Workshop on Social Media Analytics (SOMA '10), July 25, 2010, Washington, DC, USA.

communities towards PoC. In hindsight, this could have been caused by improper querying and training of the monitors. Based on these early insights, the project decided to concentrate on **Q2.**

The analysis of the **O$_2$H** monitors found few online signals for the Arabic and Farsi monitors. For English and Greek however, the number of posts was much bigger, in the order of thousands. Interestingly, only **5%** of the tweets retrieved by the **Xenophobia Greek** monitor (12,423 out of 248,691—see Table 1) were classified as xenophobic, compared to **15%** (3,969 out of 26,466) in the **Xenophobia English** monitor. Although the monitors (queries) retrieved a larger number of posts in Greek, the analysis did not reveal the absolute number of tweets. However, with the sample retrieved, there were more xenophobic posts in English than in Greek for this particular geographic location. See Annex 3 for a summary of the main topics discussed in tweets retrieved by the **Xenophobia Greek** monitor.

Iteration 2

Four follow up studies were conducted in the second iteration. The project created a unique monitor for each study using Crimson Hexagon (see Table 2).

Table 2: Situational Awareness Monitors Overview

Monitor	Unit of Analysis: Geography	Unit of Analysis: Timeframe	Language	Number of Posts*
7. Situation Awareness Nice	Worldwide	Date of event (14 July 2016) – April 18th, 2017	English, French, Greek, German	3,748,198
8. Situation Awareness Munich	Worldwide	Date of event (22 July 2016) – April 18th, 2017	English, French, Greek, German	58,815,918
9. Situation Awareness Saint-Étienne	Worldwide	Date of event (27 July 2016) – April 18th, 2017	English, French	28,884,522
10. Situation Awareness Berlin	Worldwide	Date of the event (18 December 2016) – April 18th, 2017	English, French, Greek, German	353,580,956

Source: Crimson Hexagon Forsight tool   * Number of post analyzed by the machine to the date: April 18th, 2017.

Hypotheses

Based on the first iteration, and in reaction to inconclusive results as well as in reaction to a number of terrorist attacks which occurred in Europe—resulting in refugees being mentioned in various media, including social media, in potentially concerning ways—, the project refocused to explore whether social media could provide a way to:

O$_3$: Monitor the general public's opinion on possible mislead relations between PoC and terrorist attacks, in aggregated form.

The project team focused on measuring the volume of posts that either blamed, or defended PoC to gauge public opinion and understand whether opinions were generally in favour or against PoC.

The **hypothesis** for **O$_3$**:

O$_3$H: Host communities and the general public may make a link between PoC and terrorist attacks.

Setup

To address **O$_3$H**, the monitor trainers created four additional monitors that covered the unforeseen incidents in Nice (FR), Munich (DE), Saint-Étienne (FR), and Berlin (DE), which occurred on the 14th, 22nd, and 27th of July, and on the 18th of December 2016, respectively.

Each was intended to gauge responses to the terrorist attacks, and how these might be related to PoC in the global Twittersphere. They were respectively called **Situation Awareness Nice**, **Situation Awareness Munich**, **Situation Awareness Saint-Étienne**, and **Situation Awareness Berlin**. All were trained in English, French, Greek, and German—except the **Situation Awareness Saint-Étienne**, which was only trained in English and French—using

almost exactly the same query—only some local references, and particular hashtags specific to each incident varied. Particular attention was given to employing the same vocabulary for each language to enable a relative degree of comparison between monitors. For example, "attack" in English was translated to "attaque" in French, "επίθεση" in Greek, and "Anschlag" in German.

The monitors were not restrained to specific geographic boundaries, but rather looked to understand global reactions and opinion. Nevertheless, the choice of language did concentrate the tweets

retrieved to areas where those languages are spoken (see Annex I for details on the categories and queries that were used for each language). Tweets in each monitor were tracked onwards from the date of the incident covered, i.e., in the aftermath of the terrorist attack, until April 18th, 2017.

Categories

The following categories were used for the situation awareness monitors (O_3H).

Blame: tweets that explicitly blame PoC for the incident;
Don't Blame: tweets that advocate for not blaming PoC for the incident, or at least that attempt to de-attach them;
No reference to PoC: tweets that describe facts about the incident, but that do not mention PoC;
Irrelevant: tweets that mention PoC, but that are not related to the incident;
Off-topic: tweets that are neither related to PoC, or the incident.

Insights

While the **Situation Awareness Saint-Étienne** monitor gathered a significant number of tweets (>28M—see Table 2), the religious nature of the incident tended to skew the results: PoC were rather linked with fundamental Islam, than with the event itself. As such, this monitor was discarded, and the project team further inspected only the incidents that did not specifically target religion.

Only **6%** of the tweets retrieved by the **Situation Awareness Nice** monitor, **11%** by the **Situation Awareness Munich** monitor, and **5%** by the **Situation Awareness Berlin** blamed PoC for the incident. There were also more don't blame tweets in the **Situation Awareness Berlin monitor** than in the other two, with 7% of posts expressing explicit support for PoC in German, condemning racism and xenophobia, and stating that terrorism and violence are the main reasons why PoC flee their homes in the first place. It is also important to note that while the percentages of posts blaming refugees for the incidents are small, they still represent between hundreds of thousands to several millions of spontaneous messages of this direction: **0.2M** tweets (Nice), **6.4 M** tweets (Munich), and **17.6 M** (Berlin).

The Situation Awareness Berlin monitor retrieved a significantly higher absolute number of tweets connecting PoC with the attack. These results could be attributed to several instances. First, because the police quickly identified and arrested a Pakistani asylum seeker as the perpetrator of the attack.

Although he was later released when found inno-cent[21], Twitter users pursued the discussion on a possible relationship between the incident and PoC. Some even continued to associate the incident with the Pakistani suspect, even after the police had clarified there was a Tunisian suspect. Secondly, this was the third in a series of recent incidents in Germany, after the Munich attack, and the Würzburg train incident, the latter carried out by a 17-year old Afghan asylum seeker. Third, on September 15th, 2016, Angela Merkel made some remarks regarding the integration process of PoC in Germany, in an interview aired on RBB-Inforadio, one of Berlin's main radio stations. In one comment, she stated that "drivers are needed everywhere" in Germany[22]. This comment fueled negative posts on social media, which drew a link with the lorry truck attack, and blamed Germany's Open Door Policy.

LIMITATIONS AND LESSONS LEARNED

While the results of the second iteration provide some interesting insights into the way people perceive issues related to the European Refugee Emergency, they should be considered cautiously, as social media alone can seldom provide a comprehensive overview of needs and opinions. For example, tweets are generally not representative of socio-economic diversity and age. Only people with access to connectivity, and who have an account, can post, or respond on Twitter. In addition, although Crimson Hexagon's machine learning and geo-referencing capabilities are advanced, they may not always be entirely accurate. This means that the geo-based queries may have retrieved additional tweets posted from outside the determined geographic boundary (*false positives*) while omitting others posted from inside the geographic boundary (*false negatives*). Furthermore, machine classification is not always accurate. At the same time, the project assumed that the general public would not make the legal difference between *refugees* and *migrants*, and used both terms interchangeably in the queries.

Also, the project did not establish a clear, systematic distinction between *host communities* and *the general public,* in terms of language vis-à-vis geographical location. This means that there could be, for example, a French person posting opinions in German language, currently residing in Germany; or a German person, posting in German language, currently residing outside Germany. Either example could be categorized as both host community, or general public, depending on different perspectives and proximity to the community.

Furthermore, the automatic classification and sentiment extraction from tweets may have missed some

21   The Guardian (2017). Berlin Truck Attack: first suspect released as drive thought to be still at large - as it happened. February 17, 2017.
22   Der Zeit (2016). Angela Merkel: Flüchtlinge sollen schnell in Arbeitsmarkt integriert werden. September 16, 2016.

important contextual cues, as both procedures use only the posts' textual content. Tweets may contain links, or allude to other information. They may also be part of a broader, ongoing conversation. Overlooking, or omitting these contextual cues may result in a misinterpretation of certain uses of language, such as irony, satire, or metaphoric speech. This was observed in the **Situation Awareness Berlin** monitor, where there were many sarcastic references to the radio interview with Angela Merkel.

The project initially imagined it would be possible to filter out the voices of PoC by collecting posts in Arabic and Farsi, and by geo-locating their point of origin. However, it turned out to be extremely complicated to determine whether a person tweeting in these languages was indeed a migrant or refugee, or simply a person from the host community or local diaspora—especially seeing that the Arabic and Farsi monitors in the first iteration did not retrieve many tweets. This would induce a high degree of uncertainty in any attempt to address hypotheses related to PoC vs. host communities, which is why a deeper analysis of O_1 was not employed. However, extending this research to other social media platforms like Facebook, which PoC use extensively[23], might facilitate the distinction, and help better understand interactions.

The second iteration also showed the extent to which the comprehensiveness of the vocabulary used in a query could both increase the volume of retrieved posts, and their overall accuracy. Firstly, several iterations are needed to capture all the information. Secondly, the messages should be manually scanned to be able to re-classify and re-train the machine to include relevant words. Although it was generally similar, the query used for the **Situation Awareness Berlin** monitor was more sophisticated, and better tailored, than those used for the other situation awareness monitors (see Annex I). This was the result of immediate feedback received from end-users within Germany on specific language nuances. The project used colloquial which were also used by local media and the general public when referring to PoC. UNHCR's Innovation Service's Community and Content Manager also assisted, pointing to specific hashtags and keywords that were being used.

More generally, findings showed that working with social media requires a dynamic mindset. The project had to adapt and iterate rapidly. The hypotheses from the first iteration proved too broad, and for them to be of any tangible use to UNHCR, they had to be adjusted. The project also required more resources than initially identified within UNHCR, as linking social media monitoring with operational responses and planning is a new concept for the Agency—the mini-studies were

23   European Commission (2016). Effective use of technology and social media for refugees' labour market integration.

typically inaccurately conflated with a range of other social-media-driven projects, including CwC efforts, Information Management work, Communication and Public Information (PI) activities, and UNHCR digital brand marketing.

In addition, while this white paper refers to the current work as a series of "mini-studies," it should be emphasized this was a labor-intensive process. UNHCR's Innovation Service had to resolve to recruit monitor trainers, all of whom had both relevant native language skills, and basic knowledge in computer programming. UN Global Pulse also invested the equivalent of a full-time staff member, in addition to providing the necessary partnership agreements, which enabled access to the data and tools. In the future, it will be important for UNHCR to have qualified and dedicated personnel to develop similar approaches to collecting, processing, and analyzing big data sources. These efforts should be integrated across different units, to further the Agency's understanding—as a whole—of the potential of social media and big data to inform operational decisions, advocacy activities, and strategic communications, as well as to improve listening to different affected communities, in order to demystify non-accurate information.

The limitations of the initial interactions and the lessons learned throughout the course of the project helped reshape its initial scope. The project set out to use social media posts to build a better, more nuanced understanding of different complex aspects of the Europe Refugee Emergency, that are otherwise difficult to assess with traditional tools—such as surveys. However, it now sees added value in trying to use social media to detect unexpected signals of ongoing events that could put PoC at risk, and that UNHCR may need to quickly respond to, or act upon.

The streaming nature of social media posts affords the detection of such signals in near-real-time, which could be useful in cases similar to the aftermath of the Berlin terrorist attack, where more than 17.6 million tweets linked the incident with PoC. There are few data sources that can facilitate such in-depth, rapid response mechanisms, and the project intends to continue exploring their potential.

THE WAY FORWARD

UNHCR routinely collects massive amounts of data, through, for example, registration and information management exercises, programme and project implementation, and financial activities. The main challenge, and therefore an important opportunity for the Agency, is to find ways of accompanying the integration of new data sources into this culture, and to bring more data-driven evidence into decision-making processes and advocacy efforts, particularly in developing an institutional policy against xenophobia,

discrimination, and racism against PoC. The current project intends to continue exploring this integration with the development of a social media monitoring system (an early snapshot of which is presented in Annex IV), which will use streamed posts as a way to detect signals of ongoing events, which the Agency may need to act upon.

Beyond this, there are several other opportunities for UNHCR in the future like:

- Defining clear, rigorous methodologies and protocols to distill relevant information extracted from biased data sources like social media. Interpreting social media using quantitative methods and machine intelligence is complex, particularly when the context of the composite data[24] is nuanced and sometimes unclear from individual pieces of information;
- Integrating these new types of insights into operational workflows. Social media posts can typically feed into operations, policy or advocacy, and communications. This is another opportunity for UNHCR;
- Adopting the relevant ethical and privacy frameworks relating to data protection, privacy, anonymity, and security;
- Building internal data literacy and specialized capacities within the Agency. This last point should further help improve UNHCR's capacity to make data-driven decisions.

During the *Committee on the Elimination of Racial Discrimination* in March 2011, UNHCR's Senior Legal Coordinator explained that "*Combating racism, xenophobia and related forms of intolerance against refugees, asylum-seekers and stateless persons is one of the principle objectives of UNHCR, and these forms of discrimination are one of the greatest threats to the rights of refugees and asylum-seekers, in Europe and elsewhere*"[25]. From impacting the right to seek asylum, to better understanding how xenophobia is related to the primary root causes of persecution or negatively affecting integration opportunities, this is an area of work UNHCR must be more proactive in. In fact, not addressing xenophobia towards PoC in a strategic way would constitute a shortcoming of UNHCR's overall protection mandate as an agency The 2009 "Combating Racism, Racial Discrimination, Xenophobia and Related Intolerance through a Strategic Approach" along with the 2015 evaluation of UNHCR's Southern Africa Programmes "Protection from Xenophobia" layout specific guidelines on how the agency is addressing the issue.

However, confronting growing intolerance and xenophobia are just some of the many challenges that may lie ahead for UNHCR, in a world that is more connected, and where ideas and words can be shared across many channels, including digital channels. The European Network Against Racism (ENAR) published a study that highlights an increase in protests, political/elections rhetoric, and formation of structured groups against refugees and asylum seekers in Europe[26]. They mention that "*social media is becoming increasingly crucial in forming opinions about migrants, and there has been a growing dissemination of fake ethnicity-related news about migrants with alarming and sensationalist headlines.*"

---

24   Composite data or compound data is any data type which can be constructed in a program using the programming language's primitive data types. In summary, is any language data type that isn't a machine number

25   OHCHR (2011). Committee on the Elimination of Racial Discrimination. Thematic discussion: "Racial discrimination against People of African Descent". UNHCR DIP

26   ENAR (2016). Racism and Discrimination in the Context of Migration in Europe. ENAR Shadow Report.

ANNEXES

———————————————

Annex I: Data Query Taxonomies per Hypothesis

1.1 O₁H: Monitor Interactions

- **Negative perception:** bad conditions in access to services or to territory of asylum, police brutality, closed border, means of transportation
- **Taxonomy** for link**:** basic neutral, basic positive, basic negative
- **Geography**: Greece, national level

Machine learning query: *untrained/discarded monitors*

PoC Farsi:

اروپا OR کوچگر OR (جوی AND پناه) OR پناهنده OR مرز OR شهریانی OR پلیس OR
یونان OR
دستگیری OR بازداشت OR ثب OR
بهداشت OR حال OR وضع OR پناهندگان AND اردوگاه) OR اردوگاه OR
بلوک AND بزرگراه) OR اتوبوس OR قایق OR
ریای (رفتاری AND بد) OR خشونت OR ویزا OR قیمت OR جاده OR راه
قاچاق

Translation: Police Police OR border refugee OR (harbor AND barley) OR migrant Greece OR OR OR Europe Registration arrest OR OR OR arrested Camp OR (AND refugee camp) OR conditions are OR OR OR Health BOAT OR BUS OR (Highway AND Block) OR The price OR VISA OR OR OR roads violence OR (bad behavior AND) OR kidnapping OR trafficking

General Public: Arabic

محاجر OR لاجئ OR لاجئين OR محاجرين OR مهاجرين OR أوروبا OR الحدود OR حدود OR وضع OR
الوضع OR اليونان OR يونانيين OR اليونانيين

Translation: Mohajer (migrant), laje', (refugee) OR Laj'een (refugees) Mohajeryn (migrants) Europe OR borders border OR situation OR the situation OR Greece OR Greeks OR the Greeks

General Public: English
"refugee" AND ("move" OR "movement" OR "move" OR "boat" OR "plane" OR "relocation" OR "resettlement" OR "removed" OR "returned" OR "reintegrated" OR "walk" OR "road" OR "bus" OR "train" OR "money")

1.2 O₂H: Understanding sentiment

- **Negative perception:** racists, extremist or xenophobic comments from host communities in their native language, negative sentiment and feelings towards refugees and migrants.
- **Taxonomy**: racist, non-racist, neutral, irrelevant
- **Geography:** Greece, national level

Machine-learning query

A) Xenophobia English
((migrant OR refugee OR refugees OR immigrants) AND (Greece OR Greeks OR fear OR hatred OR racism OR xenophobia OR foreigners OR arrivals OR Syrians)) AND ((migrant OR refugee OR refugees OR immigrants) AND -(RT OR US OR America OR UK OR Trump OR Brexit OR Merkel))

B) Xenophobia Greek
((μετανάστης OR πρόσφυγας OR πρόσφυγες OR μετανάστες) AND -(RT OR Βρυξέλλες OR Τσίπρας OR Μέρκελ OR Brexit OR Γερμανία)) OR ((μετανάστης OR πρόσφυγας OR πρόσφυγες OR μετανάστες) AND (φόβος OR μίσος OR ρατσισμός OR ξενοφοβία OR ξένοι OR αφίξεις OR Σύριοι)) OR ((metanastis OR metanasths OR metanastes OR prosfugas OR prosfuges) AND -(RT[1] OR Merkel OR Tsipras OR Brexit OR Germania)) OR((metanastis OR metanasths OR metanastes OR prosfugas OR prosfuges) AND (fovos OR fobos OR misos OR ratsismos OR xenofovia OR xenophobia OR afiksi OR xenoi OR afiskeis OR Syrioi OR Surioi))

C) Xenophobia Arabic
محاجر OR لاجئ OR لاجئين OR محاجرين

Translation: migrant OR migrants OR refugee OR refugees

D) Xenophobia Farsi
پناهنده OR پناهندگان OR خوشامد OR کوچگر OR یونان OR رسیدن OR خارجی OR خارجیها

Translation: refugees, refugee, welcome, migrant, Greece, to arrive, foreign, foreigners

1.3 O₃: Incidents Linkage

- **Linking incidents:** blame refugees for attacks/incidents, terrorism activities in Europe, Munich, Nice, St. Etienne, #donotblame refugees, #PrayforMunich, #offeneTür, Bastille, #BerlinAttack
- **Taxonomy:** blame refugees, do not blame refugee, neutral, irrelevant
- **Geography:** Worldwide

Machine-learning query

A) Situation Awareness Munich
(Munich OR MunichAttack OR PrayForMunich OR offeneTür OR Beschuldige OR Flüchtlinge OR Flüchtlingen OR Schuld OR Attacke OR Tod OR Töten OR Opfer OR Schießen OR Schiessen OR Attentäter OR Gewehr OR Pistole) OR (attack OR killer OR kill OR killed OR dead OR deadly OR death OR shooting OR gun OR bullets OR victims OR killing) OR (Μόναχο OR Μόναχο OR επίθεση OR PrayForMunich OR πρόσφυγες OR κατηγορούν OR πρόσφυγες OR ένοχος OR επιθέσεις OR θάνατοι OR θάνατο OR θυμάτων OR όπλο) OR (attaque OR attaques OR attentat OR attentats OR tué OR tueur OR assassin OR mort OR morts OR tournage OR fusillade OR pistolet OR fusil OR balles OR victimes)

B) Situation Awareness Nice
(Nice AND (terrorist OR attacks OR France OR dead OR (Bastille AND Day) OR terror OR deaths OR blame OR refugees OR refugee OR deaths OR attack OR victims OR assassins OR gun)) OR (Νίκαια AND (τρομοκράτης OR τρομοκρατική OR Γαλλία OR νεκρός OR νεκροί OR (Bastille AND Day) OR τρόμος OR θάνατοι OR επίθεση)) OR (Nizza AND (terroristischen OR Attacke OR Frankreich OR Tot OR (Bastille AND Tag) OR terror OR Tötten OR Beschuldige OR Flüchtlinge OR Flüchtlingen OR Schuld OR Töten OR Opfer OR Schießen OR Schiessen OR Attentäter OR Gewehr OR Pistole)) OR (Nice AND (terroriste OR attaque OR attaques OR attenat OR faute OR attentats OR France OR mort OR morts OR (Jour AND de AND la AND Bastille) OR terreur OR mortes OR blâme OR réfugiés OR réfugiés OR blâmer OR attaque OR mort OR victimes OR assassin OR pistolet OR (14 AND juillet) OR terreur))

C) Situation Awareness Saint-Etienne:
otage OR armés OR (Saint AND Etienne AND du AND Rouvray) OR mort OR morts OR (prise AND d'otage) OR église OR prêtre OR assaillants OR tué OR blessé

d) Situation Awareness Berlin
(Berlin OR BerlinAttack OR BerlinTerrorAttack OR (Berlin AND Terrorist AND Anschlag) OR (Berlin AND Terroranschlag) OR Breitscheidplatz OR merkeldeutschland OR Weihnachtsmarkt OR (Weihnachts AND Markt) OR Anschlag OR offeneTür OR Beschuldige OR Flüchtlinge OR Flüchtlingen OR Schuld OR Attacke OR Tod OR Töten OR Opfer OR Weihnachten OR Attentäter OR Gewehr OR LKW OR Islam OR Pakistaner OR Pakistanisch OR Islamophobie OR Liberale OR Immigrant OR Asyl OR Lastwagen OR Asylant OR Asylanten OR Fluechtlingsbewerber OR Asylbewerber OR Lastkraftwagen OR Migranten OR Rassismus OR Fremdenfeindlichkeit OR (Beschuldige AND Flüchtlinge AND nicht) OR (Beschuldige AND Flüchtlingen AND nicht) OR Einwanderer OR vorwerfen OR (scheiß AND Flüchtlinge) OR (scheiss AND Flüchtlinge) OR (scheiße AND Flüchtlinge) OR (scheisse AND Flüchtlinge) OR anschuldigen OR anklagen OR Vorwürfemachen OR Muslime OR (Die AND Schuld AND den AND Flüchtlingen AND zuschieben)) OR (attack OR blamerefugees OR (blame AND refugees) OR terror OR terroristattack OR terrorist OR killer OR Merkel OR (open AND door) OR opendoor OR kill OR killed OR dead OR deadly OR death OR ISIS OR islam OR Pakistani OR Christmas OR christmasmarket OR truck OR victims OR killing OR RefugeesWelcome OR liberal OR immigrant OR migrant OR asylum OR lorry OR Afghan OR jihad OR islamophobia OR racism OR (don't AND blame AND refugees) OR dontblamerefugees OR Asylmafia OR xenophobia OR thanksMerkel OR ThankyouMerkel) OR (Βερολίνο OR επίθεση OR τρόμος OR τρομοκράτης OR προσφύγων OR πρόσφυγας OR πρόσφυγες OR κατηγορούν OR πρόσφυγες OR ένοχος OR επιθέσεις OR θάνατοι OR θάνατο OR θυμάτων OR Χριστούγεννα OR φορτηγό OR (χριστουγεννιάτικος AND αγορά) OR Αφγανός OR Πακιστανός OR τζιχάντ OR ισλαμοφοβία OR ρατσισμός OR ξενοφοβία OR Μουσουλμάνος) OR (attaque OR attaques OR attentat OR attentats OR tué OR tueur OR terreur OR assassin OR mort OR morts OR tournage OR victimes OR camion OR Natale OR (marché AND de AND Noël) OR pakistanais OR asile OR (porte AND ouverte) OR refugie OR réfugié OR xénophobie OR MerciMerkel OR Musulman OR (ne AND blâmez AND pas AND les AND réfugiés))

Annex II: Tweets found and catalogued by AI

O₁H: Monitor Interactions

Translation: You are frustrated by all the refugees dying in the sea but words don't do us much, open the borders

Source	Post	Date
	أنت المحرض على قتل كل لاجئ في بحر إيجه. الكلام لا يفيد، افتح الحدود الآن #OpenEUBorders#refugeesGr #OpenEvrosBorder #OpenEvros View post	October 30, 2015

Translation: The governor of Greek Central Macedonia: There are about 13,000 refugees are swarming to the Greek-Macedonian borders in miserable conditions

	حاكم محافظة مقدونيا الوسطى اليونانية : هناك نحو 13 ألف لاجئ يحتشدون على الحدود اليونانية المقدونية في ظروف بائسة View post	March 6, 2016

	ولأن اليونان تعيش أزمة اقتصادية كبيرة .. وأحوال اللاجئين تعد فيها الاصعب View post	June 22, 2016

Translation: Greece is currently facing a huge economic crisis.. and the circumstances for the refugees are even more difficult

O₂H: Understanding sentiment

Xenophobia English Monitor
Category: Xenophobic

Child abuse rife in Greece's #Muslims migrant camps #islamisSICK #DeportMuslims #StopIslam https://t.co/c7UNAeggmC https://t...

Apr 27, 2016

Klout score:
50

RACIST

Child abuse rife in Greece's #Muslims migrant camps #islamisSICK #DeportMuslims #StopIslam thetimes.co.uk/article/greece...
12:08 PM - 27 Apr 2016

Non-Xenophobic

Neutral

Substantial drop in migrant arrivals in Greece https://t.co/Pqz5BYyQO0

Mar 28, 2016

Klout score:
44

NEUTRAL

Substantial drop in migrant arrivals in Greece
thetoc.gr/eng/news/artic...
11:49 AM - 28 Mar 2016

Irrelevant

Angelina Jolie in Greece for the refugees https://t.co/dwZbXrz1gK https://t.co/vctyWvuOW

Mar 17, 2016

Klout score:
26

IRRELEVANT

Angelina Jolie in Greece for the refugees
nikisreviews.com/2016/03/17/ang...
10:01 AM - 17 Mar 2016

O$_2$H: Understanding sentiment

Xenophobia Greek Monitor
Category: Xenophobic

Translation: *They treat Greek people Bad, to make space for 'refugees'.*

Non-Xenophobic

Translation: *Humanitarian help for the refugees in Heraklio.*

Neutral

Translation: *More than 53.900 refugees and immigrants in the country.*

Irrelevant

Translation: *On monday the first Syrian refugees will move from Turkey to Germany.*
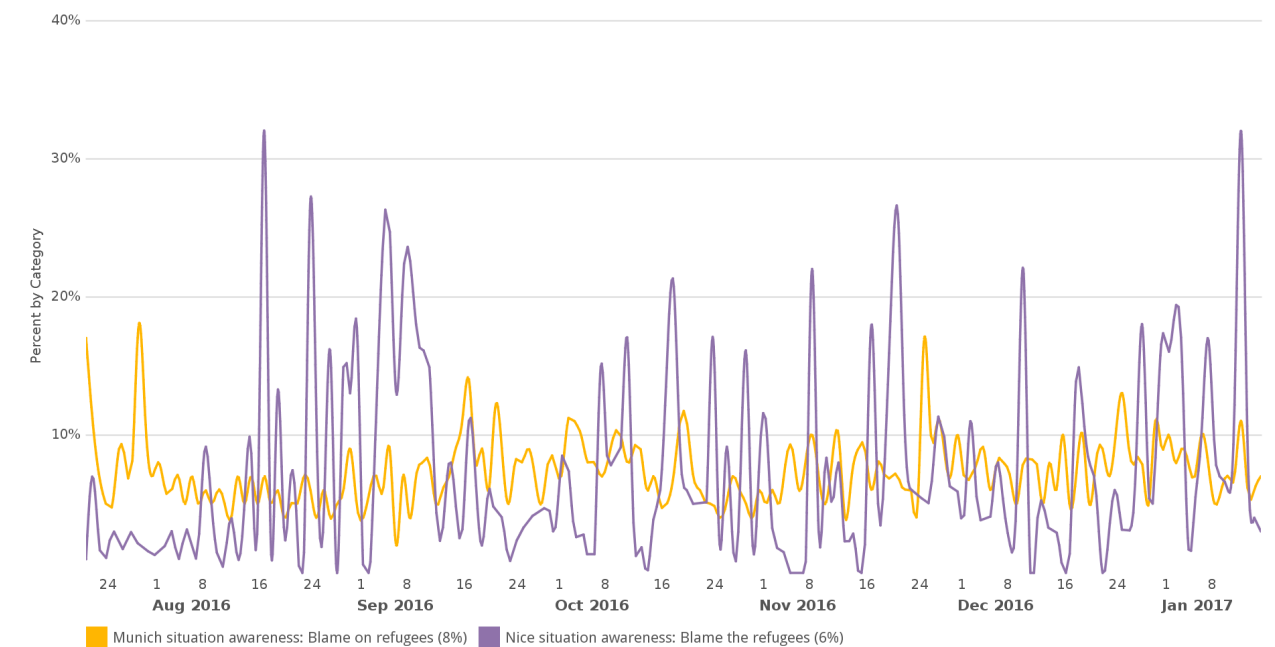
Annex III: Data Visualizations (Quantitative inputs)

O$_3$H: Incidents Linkage
Total Number of tweets analyzed: 3,433,800 (Nice) + 297,506,445 (Munich) = 300,940,245 posts up to Jan 10th, 2017.
Xenophobic: Munich (8%) in yellow and Nice (6%) in purple
Geography: Worldwide

Xenophobic

Category Trend Comparison from 7/22/16 to 1/16/17

Nice and Munich (January 10th, 2017)

O$_3$H: Incidents Linkage
Total Number of tweets analyzed: Munich (58,815,918 posts), Nice (3,748,198 posts) and Berlin (353,580,956 posts) = total 416,145,072 posts
Xenophobic: Munich (8%) in yellow, Nice (7%) in green and Berlin (5%) in purple
Geography: Worldwide
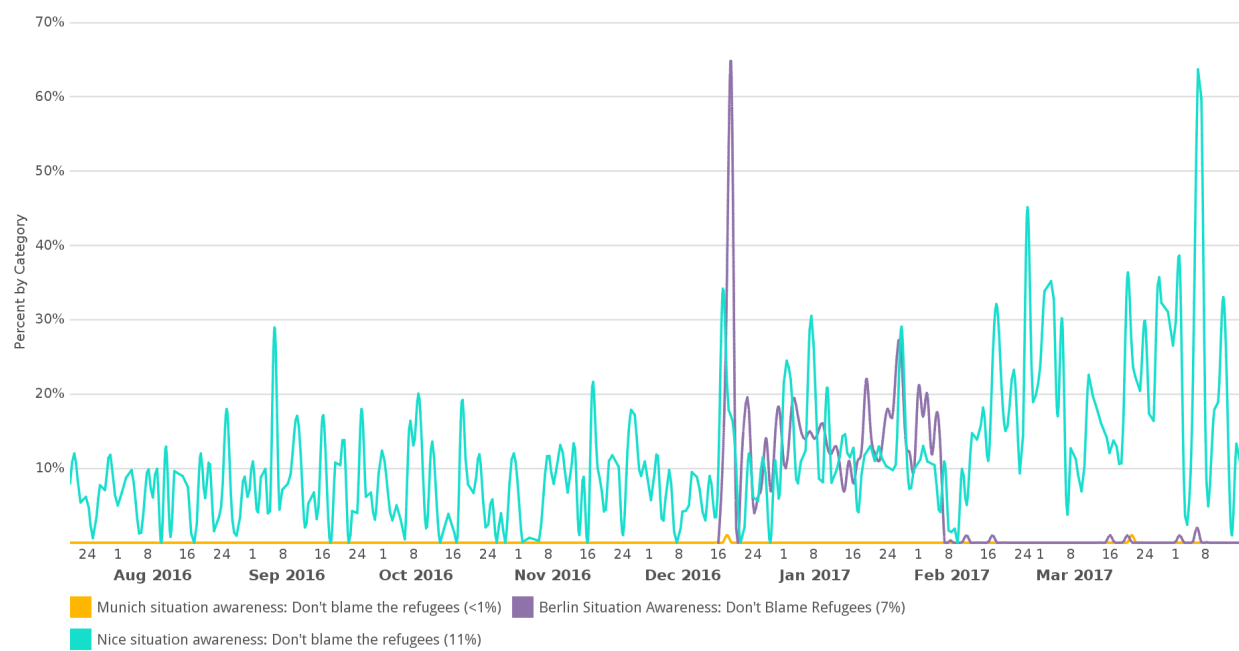*Xenophobic*

Including Berlin (April 18th, 2017)

Not-Xenophobic

O₃H: Incidents Linkage
Total Number of tweets analyzed: Munich (58,815,918 posts), Nice (3,748,198 posts) and Berlin (353,580,956 posts) = total 416,145,072 posts
Non-Xenophobic: Munich (<1%) in yellow, Nice (11%) in green and Berlin (7%) in purple
Geography: Worldwide

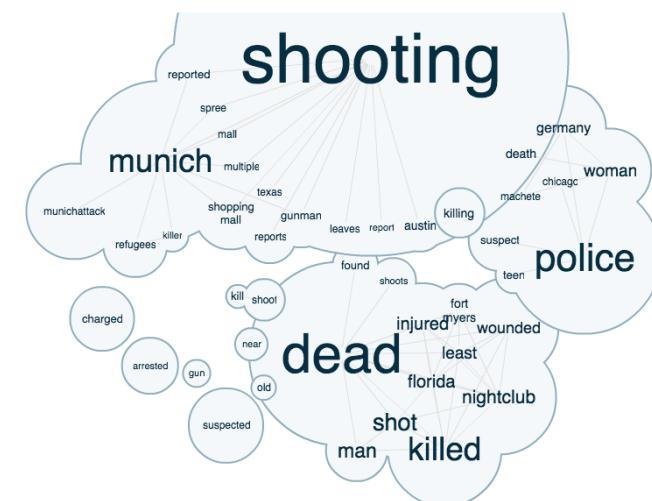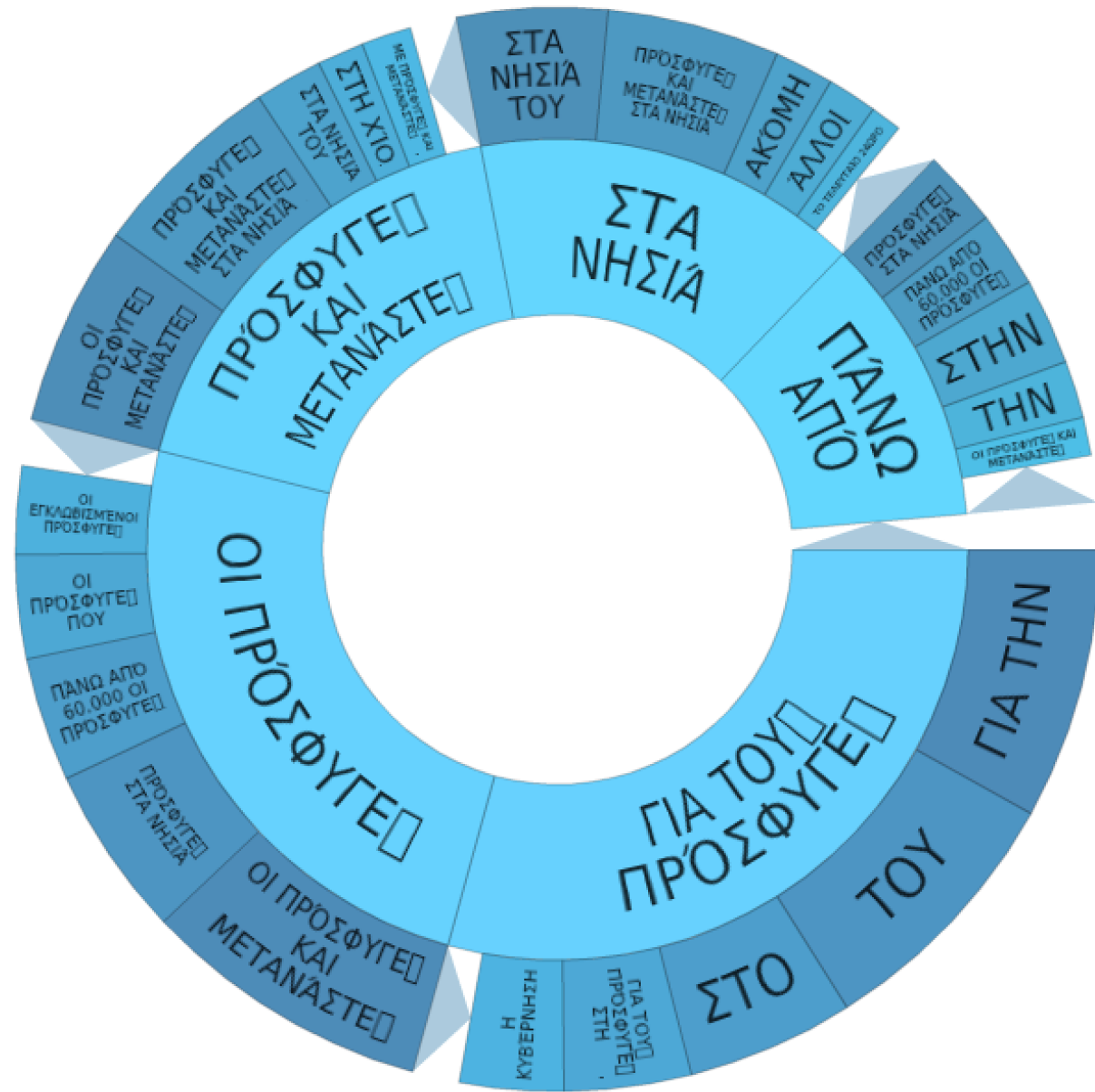Munich situation awareness — Words from 7/22/16 to 8/8/16

Data Visualization type: Cluster, Munich situation awareness monitor
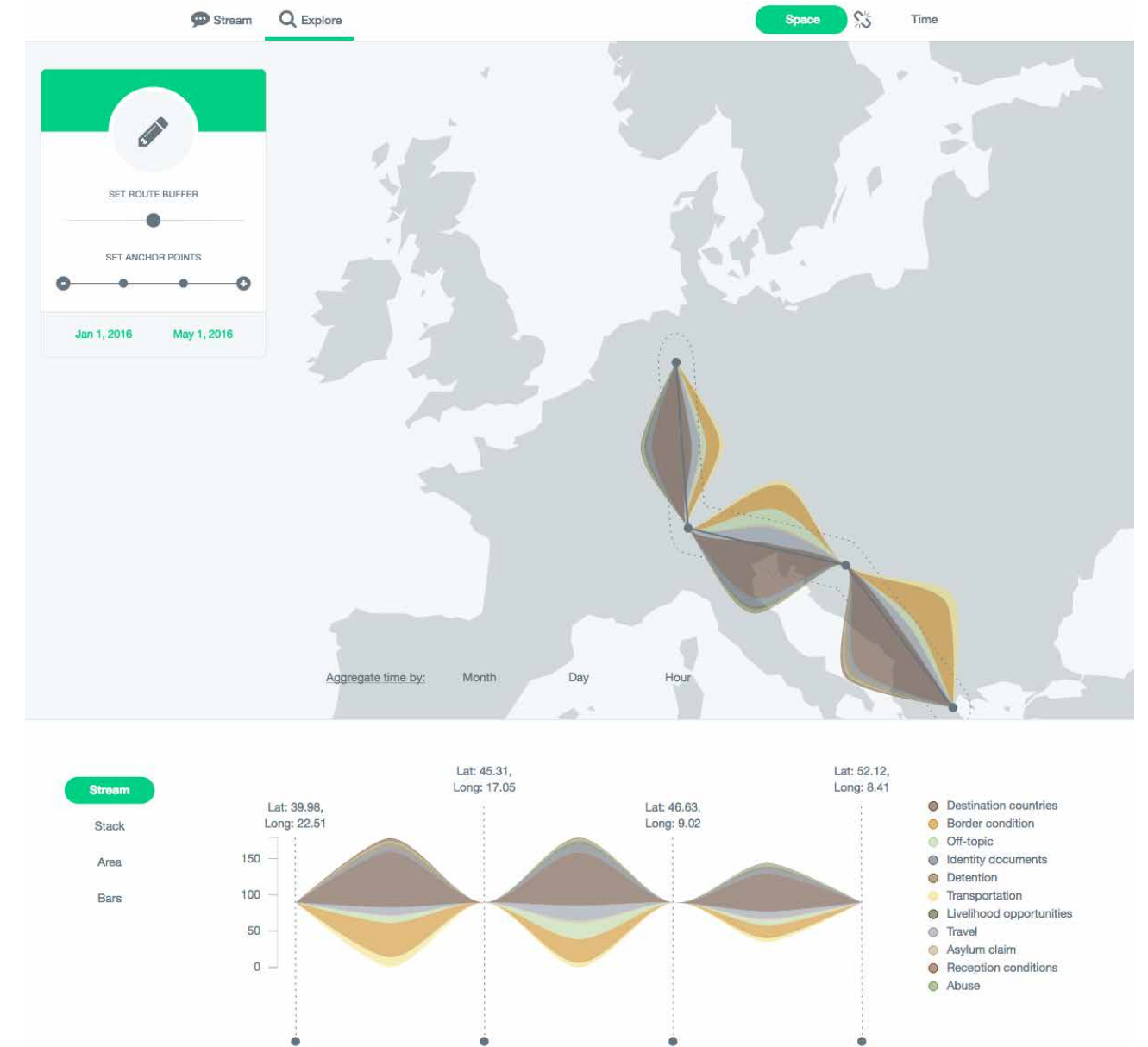

Munich situation awareness — Clusters from 7/22/16 to 8/8/16

20

21

Xenophobia Greece Greek — Topics from 9/9/16 to 10/10/16

Data Visualization type: Topic Wheel, Xenophobia Greece in Greek monitor

Annex IV: Interactive map
(under-construction by UNGP)

Tweets geo-located by route, interactive map (under construction), Python based