NatCen
Social Research that works for society

SOCIAL DATA SCIENCE LAB.

DFID Practice Note

Using Social Media Data in International Development Research, Monitoring & Evaluation

Authors: Professor Matthew L. Williams, Dr Pete Burnap, Dr Luke Sloan & Curtis Jessop
Date: August 2016
Prepared for: Department for International Development

At **NatCen Social Research** we believe that social research has the power to make life better. By really understanding the complexity of people's lives and what they think about the issues that affect them, we give the public a powerful and influential role in shaping decisions and services that can make a difference to everyone. And as an independent, not for profit organisation we're able to put all our time and energy into delivering social research that works for society.

NatCen
Social Research that works for society

NatCen Social Research
35 Northampton Square
London EC1V 0AX
T 020 7250 1866
www.natcen.ac.uk

A Company Limited by Guarantee
Registered in England No.4392418.
A Charity registered in England and Wales (1091768) and Scotland (SC038454)

At the **Social Data Science Lab** we blend innovative technology with cutting edge research techniques to study the methodological, theoretical and empirical dimensions of big social data in social, policy and operational contexts. The Lab's mission is to democratise access to big social data among the academic, private, public and third sectors, and to support real-time social data analytics for research, policy & practice.

Scalable real-time social data analytics for research, policy & practice

Social Data Science Lab
Data Innovation Research Institute
Social Science Research Park
Cardiff
http://socialdatalab.net

Contents

TABLES

FIGURES

Overview

The offer of social media data

The global adoption of social media over the past half a decade has seen the user base expand to an unprecedented level. Estimates put social media membership at approximately 2.5 billion non-unique users globally, with Facebook, Google+ and Twitter accounting for over half of these. These online populations produce hundreds of petabytes (one billion megabytes) of information, with Facebook users alone uploading 500 terabytes (five hundred million megabytes) of data daily. Social media data can add value to international development research, monitoring and evaluation in several ways.

These data are 'transformative' as they are user-generated in real-time and produced in large volumes, in contrast to the necessarily retrospective snapshots of social trends provided by conventional means such as household surveys and administrative data. As such, they can provide insight into the behaviour and opinions of specific populations that are often unreachable by conventional methods where social media uptake is high. The examples below show cases where social media data were available in high volume in development contexts. However, it is important to note that for some situations and regions social media data may not be available in such volumes, precluding their use to gain near real-time insights (see section on Challenges of Using Social Media Data).

Examples of social media data used in an international development context

There are now many examples of research using social media data in international development contexts:

Disaster relief
In the area of disaster relief, humanitarian organisations are increasingly using social media data to assess impact during, and immediately after, events. These data, for example, were adopted by the public to communicate information regarding the South East Queensland floods in 2011, increased the accuracy of disaster impact assessment during the 2013 floods in Colorado, supported the fine tuning of emergency response in relation to the floods in Indonesia in 2013, assisted in identifying the needs and concerns of those impacted by the Nepal earthquake in 2015, and allowed analysts to identify topics of discussion amongst migrating Syrian refugees in 2015.

The DFID case studies 'Early Flood Detection for Rapid Humanitarian Response' (Jongman et al. 2015) and 'Inclusiveness in Crowdsourced Disaster Response' (Weber 2012) showed how in the Philippines and Pakistan, Twitter-based analytics platform Floodtags supported disaster monitoring, and how in Haiti, social media data facilitated a grassroots approach to digital humanitarianism by giving local actors the ability to voice their needs.

Monitoring and evaluation
The data from social media offers potential to be used in the monitoring and evaluation of development programmes. For example, DFID funded research into the potential of social media data to inform elections support in Nigeria (Bartlett et al. 2015, see Section 7), found that these data could assist in monitoring the issues identified during the election, and in evaluating the impact of different actors in the process. Ongoing

DFID research in the middle east seeks to identify whether social media analysis can be used to analyse how beneficiaries reacted to a cut in their assistance package, whether impacts were felt differently by different groups, and how communications and media can have an effect.

Disease outbreaks
Social media data have also been used to assist in the identification of disease outbreaks. Some digital platforms (e.g. HealthMap) mine social media datasets and search trends for keywords and have been credited with helping to detect outbreaks of influenza (Nagar 2014). However, these systems, in particular Google Flu Trends, have come under criticism due to inherent biases present in social media data (Lazer et al. 2014).

Challenges of using social media data

Despite the successes outlined above, social media data are not a panacea for international development research and evaluation. Key challenges to using these data are outlined below.

Knowledge of and tendency to post on social media platforms
Most social media datasets are not representative of the population as a whole. We know that propensity to use the Twitter varies by socio-demographic and economic factors. In particular, previous work shows that younger people and higher income earners are more likely than older people and lower income earners to use the platform (Sloan et al. 2015). Analysis of social media data needs to be based on a clear understanding of how the population in the region of interest use various platforms, and which elements of the population will be overrepresented, and which will be absent.

There are also variations in how people use social media platforms. For example, we know that of those that do use Twitter, the propensity to geo-locate is also influenced by various factors such as age (Sloan and Morgan 2015). Furthermore, changes in technology, such as the release of new mobile phone handsets and software updates of the Twitter app, have also been shown to impact the number of users including geo-location data in their tweets (Swier et al. 2015). Using social media alone to study mobility patterns after a disaster may therefore give a biased picture.

There are techniques to calibrate for sampling bias, which use standard statistical models and methods to control for mobile or internet penetration rates in, for example, a given area or age group. However, even with calibration, the ability to generalise the models and their results to other times and places is limited.

Perceptions of offline phenomena
Concern over disasters, violence during elections, cuts to international aid etc. can vary by citizen. This could shape their responses and actions in relation to reporting on these topics on social media.

Attrition
Attrition relates mainly to monitoring communication following disasters. For example, in the aftermath of an earthquake, more social media posts are likely to come from less affected areas than from areas that have been devastated due to network disruption and difficulties in access to working technology. Assessing need based on the number of social media posts may send aid to the wrong places. In these situations, it is important to verify any online patterns with ground personnel and other forms of data (see Section 7 for further discussion on the use of social media following the Nepal Earthquake).

Tendency to broadcast issues and concerns related to international development on social media

The propensity to use social media to report on different types of offline phenomena related to international development is perhaps the most challenging as no weighting or calibration is likely to succeed in adjusting the sample to generate reliable results. For example, research suggests social media is used in different ways by different users. While some users may take to Twitter to voice an instance of violence witnessed at a voting station, another Twitter user may instead decide to report such violence to an official offline. In summary, not every Twitter user who witnesses violence will use the platform to report it.

The remainder of this practice note further outlines the challenges of social media data and provides an overview of the main types of analysis that are currently available and are being used in international development contexts. The practice note ends with details of several case studies (Section 7).

1 The key benefits and challenges of social media data: The 6 Vs

Researchers in international development are being challenged by new forms of socially relevant data produced in large volumes on social media networks. The exponential growth of social media uptake and the availability of vast amounts of information from these networks has created benefits for research, monitoring and evaluation, but also fundamental methodological and technical challenges. These can be summarised as the 6 Vs: **volume**, **variety**, **velocity**, **veracity**, **virtue** and **value**.

1.1 Volume

Volume refers to the vast amount of socially relevant information uploaded on computer networks globally every second. Within the UK alone there are 15 million registered Twitter users (Rose 2014), posting on average 30 million tweets per day. Of these online social interactions, a sizable portion are relevant to social and government research, including international development research, monitoring and evaluation.

This volume creates technical and methodological challenges. The technology to collect, store, search and retrieve such vast amounts of data is rarely available to researchers, meaning the insights contained within these datasets often remain undiscovered. Furthermore, our existing modes of analysis (qualitative and quantitative) may not be appropriate for such sizable datasets. For example, manual qualitative analysis of millions of tweets following a disaster is simply too time consuming, while statistical modelling on such large datasets remains largely uncharted[1].

Finally, while big dataset sizes are vast compared to what researchers are used to dealing with, social media does not produce a census of offline populations, and research is still ongoing with regards to mapping the coverage of the various platforms in operation.

However, despite these limitations, the volume of data allows researchers to potentially reach populations that are inaccessible using conventional methods, especially in relation to disasters in remote regions. Large volume also means multiple topics can be mined, providing opportunities for research, monitoring and evaluation on a wide array of international development programmes.

1.2 Velocity

Velocity refers to the speed at which these new forms of data are generated and propagated by social media users. This velocity can create technical challenges in terms of presenting the most useful information to an analyst in the moment of an event. For example, data may be produced at such speed that important information is quickly overtaken by new social media posts.

Despite this challenge, the rapid and continual production of these naturally occurring data means researchers can observe events, such as disasters and elections, as they unfold, as opposed to retrospectively gathering data months or even years after. Recent social unrest illustrates how social media information can spread over large

[1] For example, existing big data approaches tend to produce models and algorithms that are over fit to the idiosyncrasies of a particular data set, leading to spurious results that often do no not reflect reality.

distances in very short periods of time as evidenced by the Tunisian and Egyptian Revolutions (Choudhary 2012; Lotan 2011).

1.3 Variety

Variety relates to the heterogeneous nature of social media data, with users able to upload text, images, audio and video, and the array of networks available to users and researchers. Multimedia datasets obtained from platforms are rich in meaning that can be harnessed by researchers. However, unlike qualitative and quantitative data that are often labelled, coded and structured within matrices and ordered transcripts, social media data are messy, noisy, complex and unstructured making it difficult to manage and analyse. Social media networks also produce different forms of data, some favouring images over text for example. The provision of data also varies by network, with some providing data for free, while others charge a fee (see Section 4 on data access).

1.4 Veracity

Veracity relates to the quality, authenticity and accuracy of social media data. Tweets collected during or after an emergency may be deliberately misleading or false. It is also sometimes difficult to verify who is producing social media posts. For example, accounts can be controlled by Bots (automated accounts produced to spread particular messages) that masquerade as real users.

However, these data can be considered as 'naturally occurring' (albeit mediated by technology), reflecting the opinions and actions of particular populations in real-time, unmediated by researchers who shape research questions to collect data.

1.5 Virtue

Virtue relates to the ethics of using social media data. It is practically difficult to seek informed consent from social media users in research, and many Terms of Service require users to consent to share any content posted with third parties. We may argue therefore that researchers in this field must accept that consent has been provided, as long as researchers adhere to basic principles of social science ethics, while ensuing results are presented at an aggregate level. The issue of ethics and social media research is discussed further in Section 6.

1.6 Value

Finally, **value** links the preceding five Vs – only when the volume, velocity and variety of these data can be handled, and the veracity and virtue established, can researchers begin to extract meaningful information.

The large scale of social media data and their 'real-time' nature allows them to fulfil a role that data generated via conventional methods cannot. However, conventional methods retain their primacy in research given their robustness. Social media based analysis techniques cannot act as a surrogate for more established terrestrial methods. Instead, they should augment them, triangulating online data with terrestrial sources, such as curated and administrative data.

2 Types of Social Media Analysis

Social media researchers have both adapted existing methods and created new tools to analyse these new forms of data. With Twitter data it is currently possible to conduct, at the individual tweet level, language, geospatial, sentiment (positive, neutral, negative), topic (such as threats of violence and hate speech) gender, age, and occupation/social class classification. At a corpus (dataset) level it is possible to conduct keyword/hashtag/tweet frequency analysis, topic frequency (often visualised via a word-cloud or cluster) social network analysis, and information flow analysis (indicating what features lead to virality). These forms of analysis can be combined into workflows to create models that allow for the visualisation and prediction of an array of phenomena.

2.1 Geospatial Classification

From an international development research, monitoring and evaluation perspective, location data is incredibly valuable as it enables analysts to establish the geographic context in which the tweeter is immersed at the point of data creation. Having geospatial data enables researchers to position social media posts within existing geographies to which demographic and contextual data from traditional curated or administrative sources can be linked, improving the research value and veracity of the social media data. For example, in the run up to an election, tweets containing mentions of witnessing violence can be located proximate to voting stations allowing action to be taken swiftly.

2.1.1 Identifying location of Twitter users

Twitter provides several opportunities for collecting geographical information about Twitter users: from the user profile and from geo-tagged tweets[2]. Given the paucity of research conducted on rates of Twitter use in some parts of the world we cannot provide accurate statistics on all regions at this point in time. Given these limitations, the sections below outline what we know about Twitter geolocation data at a global and UK level. These sections can be read as a baseline for what *can be known* about this dimension of Twitter data in other world regions.

User profile

Geolocating a social media poster can be achieved by examining the location field in the Twitter user's profile. This is the location where Twitter users say they live. When analysing user-supplied locations, researchers face a number of challenges:

- Missing data – not all users provide location data
- Messy data – users may write a place name, but it may be misspelt or include extra punctuation/symbols rendering its interpretation difficult
- False data – users may lie about where they live (e.g. humorous/wishful references to real or fictional places)

Despite these challenges, it is possible to identify the country for over half of the Twitter users from the profile location field using software such as the Yahoo! PlaceFinder geographic database. Using this service it is possible to locate the country for 52% of Twitter users, the state for 43% of users, the county for 36% of users, the city for 40%

[2] Location can also be determined from the content of tweets and from users' own networks, but this requires specialist computational expertise.

of users and the postcode for 10% of users. These figures relate to Twitter users globally, and it must be noted that there is likely significant variation by region.

Geo-tagged tweets

Beyond including geolocation information in the profile, Twitter users have the option to enable location services on their account. This feature is off by default and requires users to opt in, but once it is enabled users can geotag their tweets[3] with precise location data in the form of latitude and longitude coordinates.

In most cases geo-tagging is performed when tweets are sent from mobile devices such as smart phones, tablets and laptops. Despite the ubiquity of mobile devices the proportion of geo-tagged tweets is very small - typically 1-3% of all tweets are geotagged, meaning that the exact position of where the tweeter was when the tweet was posted is recorded using longitude and latitude measurements (although this percentage can vary depending on the tweeters activity)[4]. There are two main reasons for such a low geo-tagging rate. First, geo-tagging is turned off by default on most mobile devices and many people do not know how to activate geo-tagging or even that their mobile device is capable of geo-tagging their tweets. Second, there is increasing concern over privacy issues.

Figure 2:1 shows the distribution of 966,082 geo-tagged tweets from a random sample of 113 million. The dotted areas follow closely the population of the globe whereas the dot-free areas follow the regions of the globe known to have little or no population.

Figure 2:1 Global distribution of geo-tagged tweets

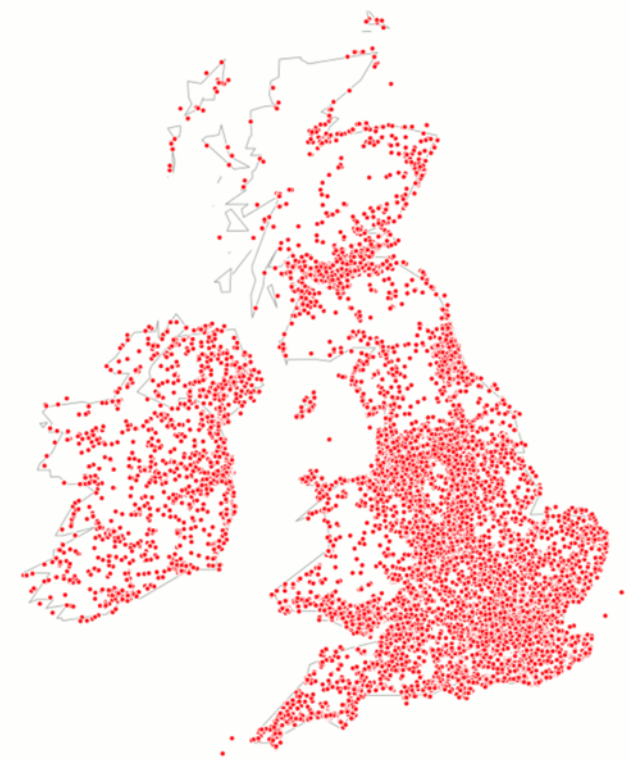
Base: All geo-tagged tweets from a 1% subsample of 113M tweets (N=966,082)

The scale of the map in Figure 2:1 makes it difficult to see how closely the geographic distribution of tweets mirrors the geographic population distributions within individual countries. Zooming in to show just the geo-tagged tweets in a sample sent from the United Kingdom and the Republic of Ireland, Figure 2:2 shows that the Twitter users send tweets in proportion to the population densities.

[3] Only tweets with original content can be geotagged. Retweets generated by invoking the retweet command in the Twitter user interface are not classed by Twitter as original content and are never geotagged. However, retweets generated by copying and pasting the content of a tweet into the tweet-composition box are classed as original content and can be geocoded (if the user chooses).

[4] For example, Twitter users tend to geolocate content more frequently when at events or on vacation.

Figure 2:2 Distribution of geo-tagged tweets in UK & Ireland



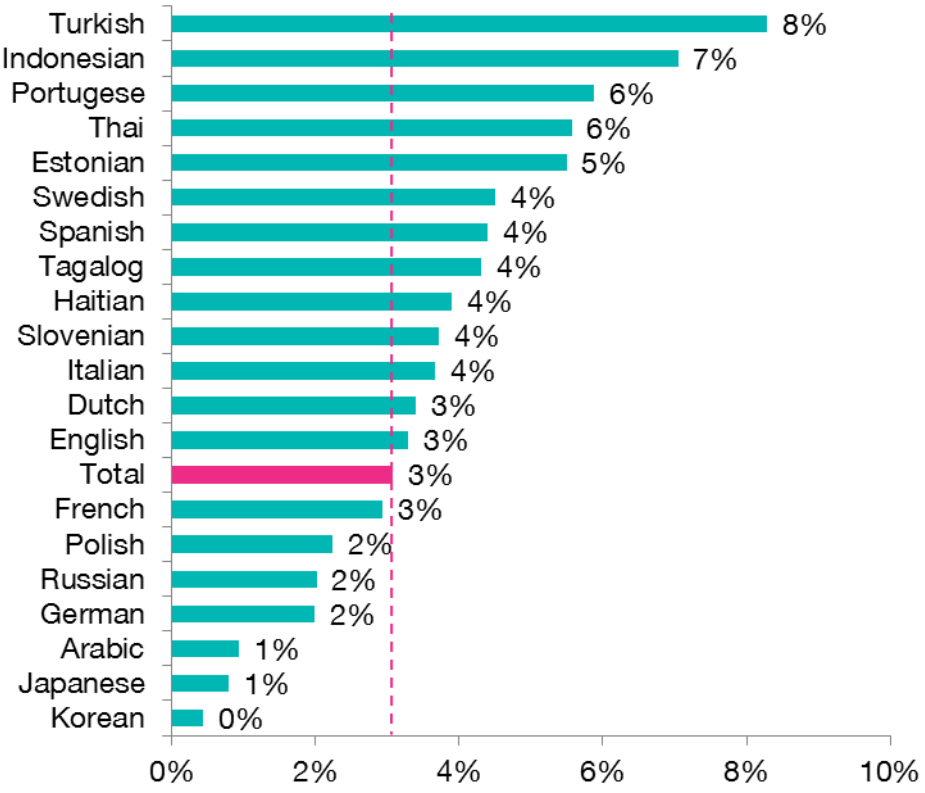
Base: All geo-tagged tweets from a 1% subsample of 113M tweets (N=966,082) – only tweets in UK shown

The representativeness of geo-tagged tweets

There is a conceptual difference between geotagging and other forms of location tagging (profile-based, text-based and network-based). Geotagged data tell researchers where a person is when they publish the tweet, whilst the other forms of location tagging could tell researchers any number of things including where people were born, lived, employed, are passing through or simply identify with. For these reasons, geotagged tweets have become the gold standard. They contain the most information in the most useful and accurate format for international development research, monitoring and evaluation.

However, it is unlikely that the small proportion of users with geocoding enabled are representative of the wider Twitter population. Analysis of the demographics of who geotags their tweets suggests there is little difference from the Twitter population in terms of sex, age or occupation (Sloan & Morgan 2015). However, the proportion of tweets geo-tagged varies significantly by language (Figure 2:3).

Figure 2:3 Proportion of tweets geo-tagged by language



Base: All tweets: Total (21975361), Korean (257154), Japanese (4801053), Arabic (1364826), German (116989), Russian (734154), Polish (43591), French (440241), English (8358844), Dutch (102381), Italian (165954), Slovenian (27684), Haitian (41596), Tagalog (311788), Spanish (2451251), Swedish (41934), Estonian (38117), Thai (195789), Portuguese (989437), Indonesian (874075), Turkish (618503)

User language is not a proxy for location so these cannot be dubbed as country level effects, but perhaps there are cultural differences in attitudes towards Twitter use and privacy for which language acts as a proxy. Researchers may make some tentative observations about technological infrastructure and levels of smartphone use and it may be the case that decisions about behaviour on Twitter are primarily cultural for some groups but a function of technological necessities for others, or even a mix of both. Regardless of the cause, clear differences exist based on language that demonstrates inconsistent adoption of geotagging Twitter content.

Those who enable the location setting and, perhaps more importantly, those who geotag their tweets are not representative of the wider Twitter population. For international development researchers, the impact of these biases will differ in magnitude depending on the topic and region being studied. For those using geotagged data from the 1% Twitter API the gender, age and class differences may be tolerable but careful consideration of the heterogeneity apparent in geotagging based on language (perhaps as a function of cultural and technological factors) is essential (Sloan & Morgan 2015).

2.1.2 Examples of using geolocated tweets in a development context

The display of *topic related* geolocated tweets has also been used in several international development contexts, especially in relation to disease spread and response to disasters.

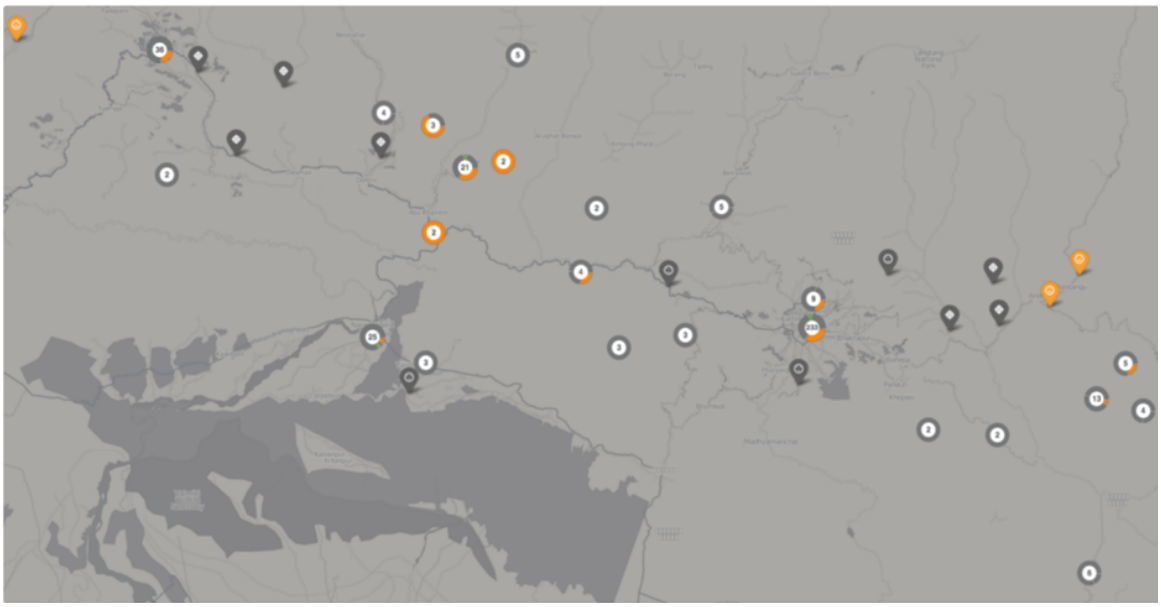
A DFID case study into the social media response to the Nepal earthquake in 2015 serves as a good example (Ripjar 2015). The epicentre of the earthquake was close to rural Gorkha and the affected area included the Nepalese capital, Kathmandu. DFID in the UK and Nepal were interested in seeing what signals, if any, there were in social media that could help them to understand the situation and needs in the most impacted districts in Nepal[5].

A Twitter key term search generated over 1.5 million tweets related to the event, far too many to read in a short space of time. Reducing the dataset to those accounts identifying as being inside Nepal reduced the number of tweets to 38,000, and further focus on those containing GPS coordinates resulted in a more manageable dataset of just 480 tweets.

Figure 2:4 shows a plot of these geocoded tweets relating to the earthquake. The majority of the data points are focussed on the population centres of Kathmandu and Pokhara with a smaller cluster close to the epicentre of the earthquake. The content of the majority of these tweets related to requests for aid to the regions between Kathmandu and Pokhara. Some of these regions waited 10 days to receive aid due to accessibility reasons. This analysis was conducted after the event and it is unclear whether such information should be acted upon in future crises given the biases inherent in social media data.

The Assessment Capacities Project on the Nepal Earthquake Disaster Response (ACAPS 2015, see Section 7) found that social media monitoring was not useful in breaking down needs geographically. Social media users in Nepal were overwhelming concentrated in Kathmandu meaning data generated by these users were more suited to analysing issues that directly affected people in the capital than in rural areas. The project concluded that the Twitter population is not representative of the national population in Nepal, and that there were clear biases in relation to the use of geo-location services following the earthquake. Therefore, extreme caution is required when interpreting data from geo-located Tweets, and any information garnered should be cross-referenced with on the ground knowledge and other forms of data.

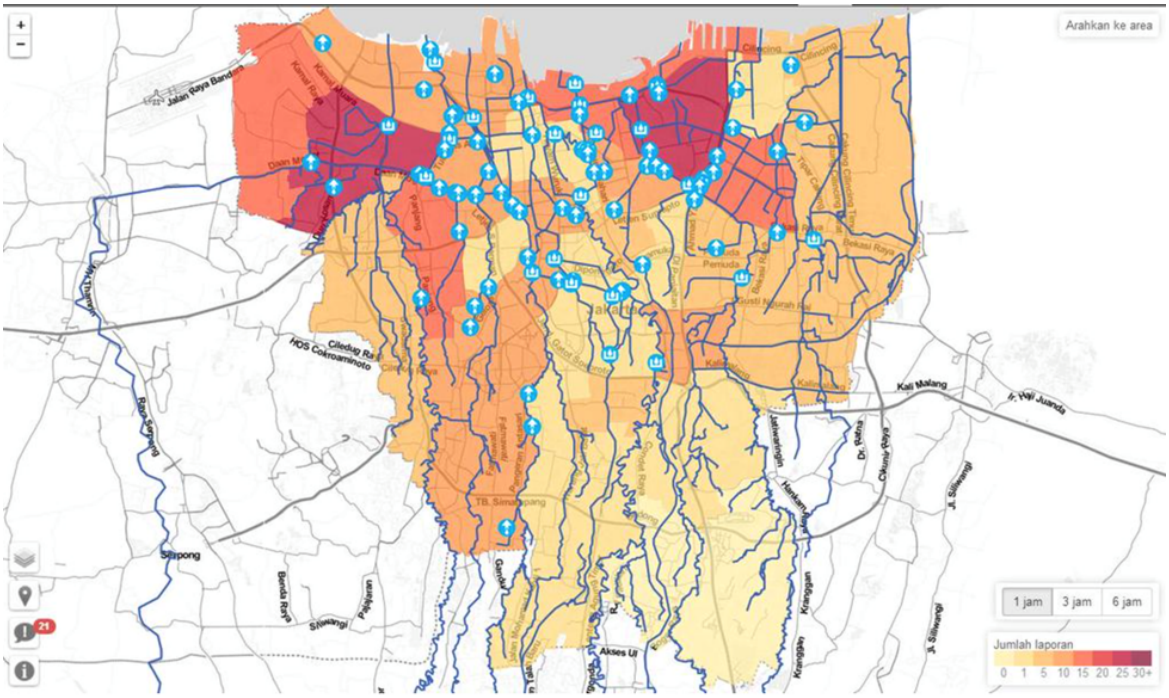
Figure 2:4 Geolocated tweets about the Nepal earthquake



[5] It is important to note that social media reported the earthquake before the international media

Figure 2:5 shows the PetaJakarta system that uses crowdsourced topic related geolocated Twitter data to respond to flooding in Jakarta. The PetaJakarta project has mapped 8 million flood-related tweets throughout the region over the past two years. The system builds a real-time map of areas affected by floods, based on geo-tagged tweets directed to the project using a specific hashtag. This technique involves the active participation of the general public to use a specific hashtag to assist the emergency services. The goal is to help emergency workers and citizens understand how floods are moving and what areas have been hit the hardest.

Figure 2:5 Map of flood-related Twitter activity across Jakarta in February 2015



2.2 Sentiment Analysis

Sentiment analysis is a text analysis technique that provides a numerical measure of the overall emotional content in a piece of text. It generally requires:

- The identification of an entity on which the opinion is focused (e.g. a person, event, product)
- Views, attitudes or feelings towards the entity and its attributes (commonly defined as sentiment)
- An opinion holder
- A time at which the sentiment was expressed

Sentiment analysis can be used in both real-time and on historical social media data. In a real-time setting a researcher might use a hashtag search on an unfolding political protest in a middle-east region using the Twitter Streaming API. As tweets are produced sentiment scores are computed and displayed in a line chart (or alternative visualisation) in real-time in an aggregate fashion, allowing for the identification of peaks in positive or negative content. Repeated peaks in negative content might indicate a pattern is forming, which may prompt the researcher to visually inspect the content of these highly negative tweets, and maybe invoke other tools, such as network analysis to identify key thought leaders in the negative social media discourse.

2.2.1 Examples of using sentiment analysis in a development context

Figure 2:6 and Figure 2:7 show the sentiment of tweets related to the Ebola outbreak in West Africa in 2014 (example from COSMOS data archive). Figure 2:6 plots average sentiment (scored on a range of -5 to +5) by gender, against time, while Figure 2:6 incorporates geotagging information contained within tweets to plot sentiment (size of circle) by gender (colour) on a global map.

Figure 2:6 Twitter sentiment towards Ebola outbreak in West Africa in 2014 over time by gender

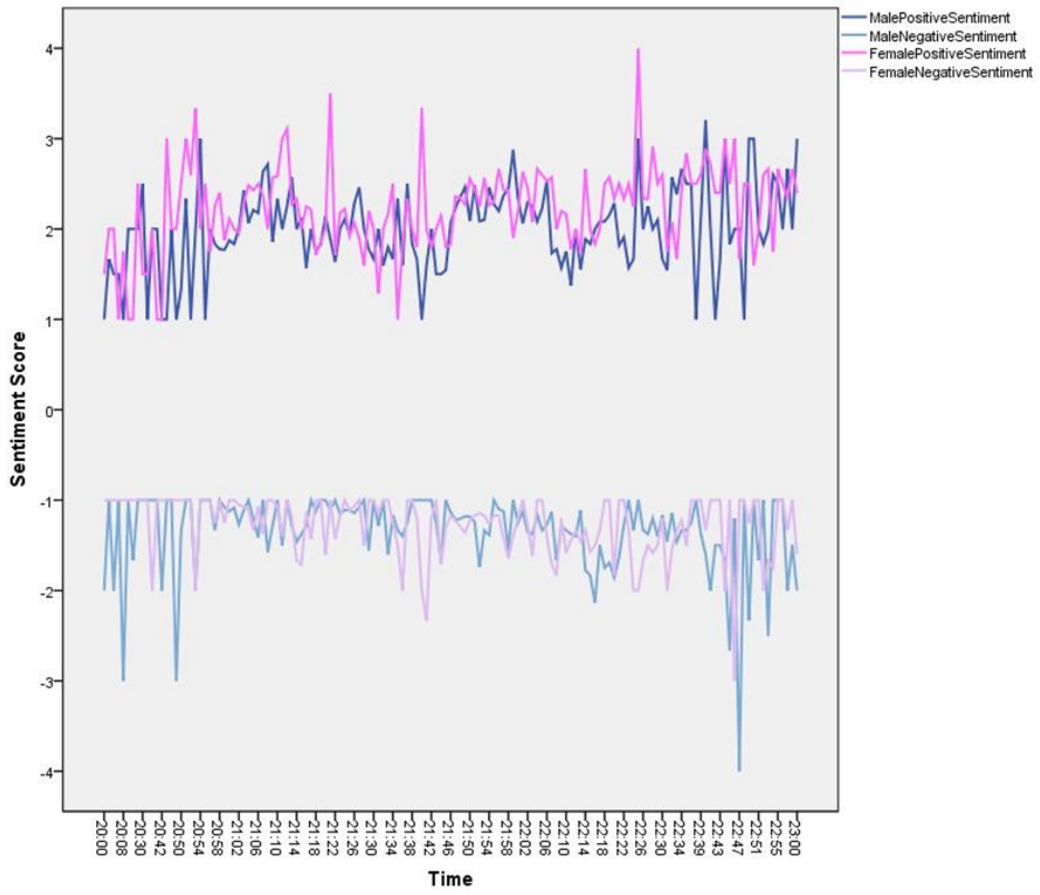
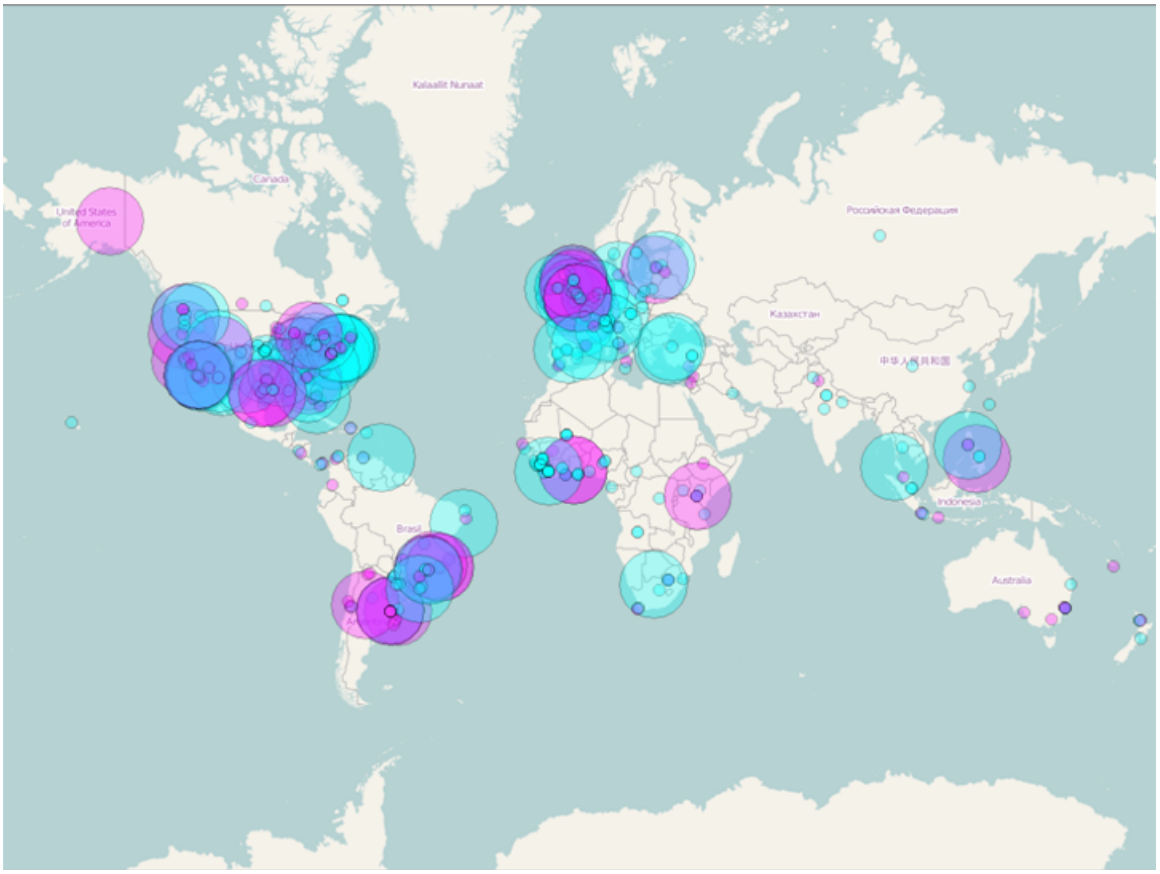


Figure 2:7 Twitter sentiment towards Ebola outbreak in West Africa in 2014 by location and gender



2.2.2 Validity of sentiment analysis

The outcome of sentiment analysis is often subjective and based on the existence of a list of keywords in a message, and it is important for researchers to be aware of the problems associated with employing generic sentiment analysis tools to specific contexts.

Generic tools are trained and tested using terms that are classified as either positive or negative. Many social media software platforms have validated the 'ground truth' of their sentiment analysis tools with human coders and compared the human annotated sentiment score with results returned from the sentiment algorithm. This is a semi-automated approach where human input is used to tailor a machine's interpretation of what is positive or negative and can dramatically increase the speed at which a general opinion on a topic can be obtained relative to using solely human coders.

However, this approach means sentiment analysis is not sensitive to certain forms of communication, such as sarcasm. As a result, tweets can be misclassified resulting in erroneous results. If the sentiment of content is of importance to a research project, adequate resources should be set aside to manually code tweets or develop bespoke algorithms designed around the research problem.

2.3 Tweet and Keyword Frequency Analysis

Tweets can be analysed by the occurrence of specified keywords (or tweet frequency if no keywords are specified) over time. This allows a researcher to visually identify points of high and low activity in relation to an event or topic.

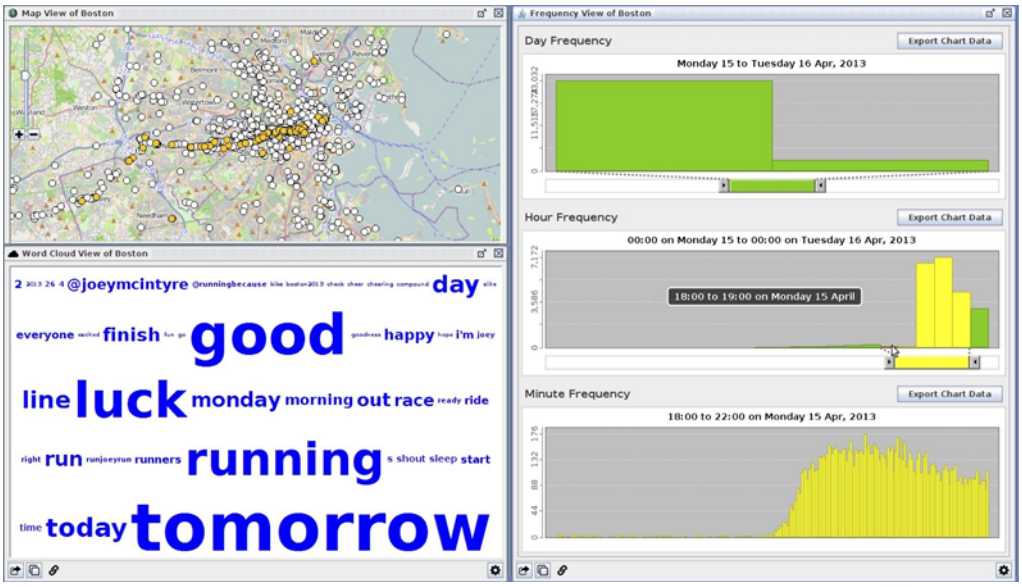
2.3.1 Example of tweet and keyword frequency analysis in COSMOS software

For example, Figure 2:8 shows the COSMOS software visualises frequency by day, hour and minute, each visualised on its own time line. Each bar in the chart represents a period of time (i.e. day, hour and minute) and users can mouse over each bar to display the text or geolocation of the tweets posted during that period. This provides an 'at a glance' view to enable the visual identification of key attributes of tweets. Sliders can then be used to select a range of bars in the chart, thereby refining the dataset to extend and contract the time frame as a scopic tool to visualise and analyse only those tweets posted within the selected range.

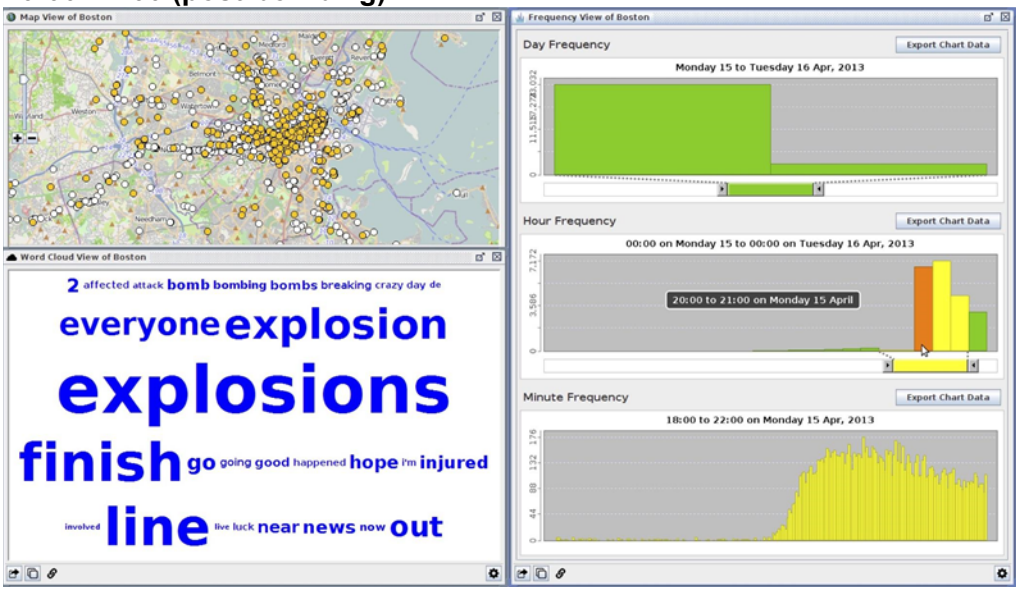
Figure 2:8 shows tweets collected in real-time on the keywords "Boston" + "Marathon". The first image shows the 18:00-19:00 time range selected in the hour bar chart, which presents a summary of all tweet content in a wordcloud at the bottom left of the display (see Section 2.5 for details of topic classification), and a geographic representation of the tweets (tracing the marathon route) in the top left of the display (see Section 2.1 for details of tweet geolocation). The second image shows the 20:00-21:00 time range selected. This range shows a significant spike in traffic, indicating that something anomalous may have occurred. The resulting visualisations show that the content of the tweets for this period have changed from the previous period. In place of well wishes for the marathon runners, terms relating to a bomb explosion appear, indicting a likely attack near the finish line. Further, the geolocation display shows tweets are now more evenly distributed over the city as the news of the bombing propagates beyond the marathon route.

Figure 2:8 Tweets collected on the keywords “Boston” + “Marathon” visualised using geolocation, wordcloud and frequency.

18:00-19:00 (pre-bombing)



20:00-21:00 (post-bombing)



2.3.2 Examples of using tweet and keyword frequency in a development context

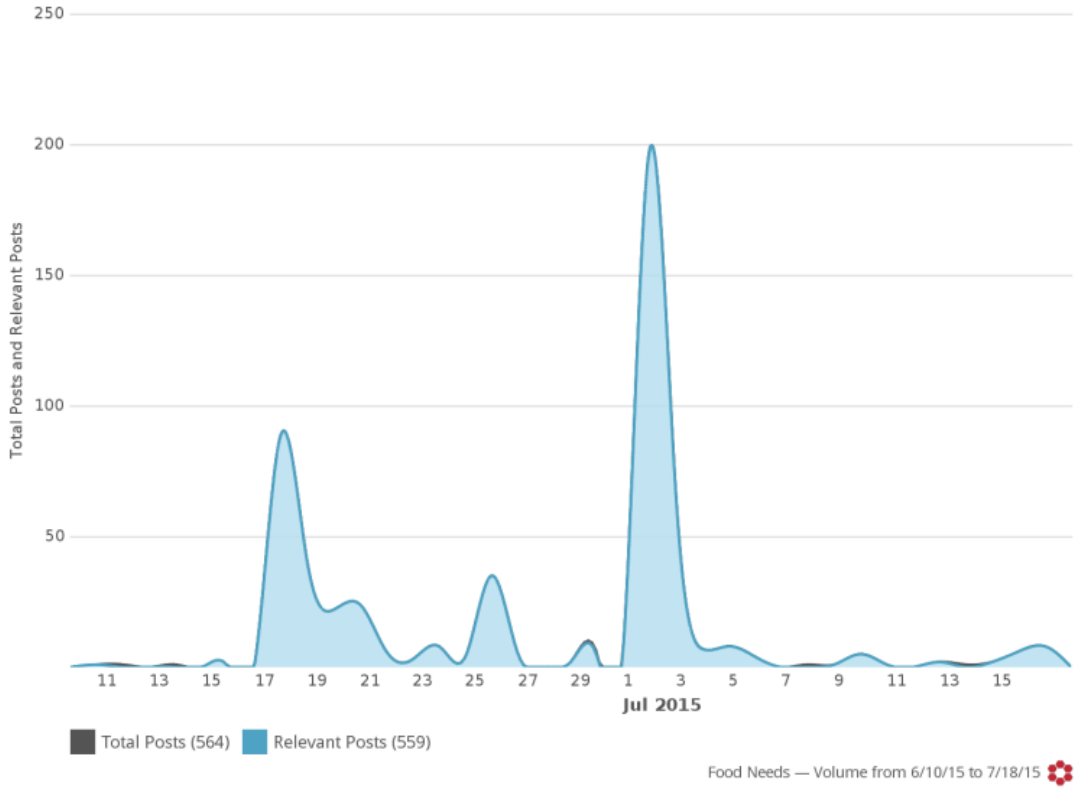
Several DFID case studies have used tweet frequency analysis to visualise spikes in traffic before, during, and following crises and elections.

The project ‘Early Flood Detection for Rapid Humanitarian Response’ (Jongman et al. 2015) studied the temporal trend in the frequency of tweets related to the floods in the Philippines and Pakistan. The top image of Figure 2:9 shows frequency of flood related tweets in a timeline, with example tweets highlighted. The bottom image is a geographic representation of frequency of flood related tweets.

Figure 2:9 Schematic display of a typical Twitter post frequency pattern leading up to a flood event (above) and heat map of flood related Twitter frequency (below)

Figure 2:10 and Figure 2:11 are derived from the ACAPS (2015, see Section 7) project supporting the Nepal Earthquake Assessment Unit. Figure 2:10 shows an increase in frequency of social media posts following media reports on June 18th alleging that a UN agency had distributed substandard food to areas impacted by the earthquake. These frequency-based visualisations are useful as they show reactions to media reports are short-lived.

Figure 2:10 Changes in volume of a query that monitored food-related issues



Frequency of communications can also be compared to historic data, and Figure 2:11 shows a comparison of the volume of discussions related to shelter following the Nepal earthquake with the year before. This visualisation helped analysts to more clearly establish a baseline for what level of conversations related to certain topics could be considered “normal” at that time of the year and what can most likely be attributed to extraordinary events like the earthquakes. Figure 2:12 shows the frequency of tweets relating to the Nigerian election, classified by topic – general tweets and tweets containing references to violence (Bartlett et al. 2015, see Section 7).

Figure 2:11 Volume of discussions related to shelter in June 2014 (purple) and June 2015 (orange)

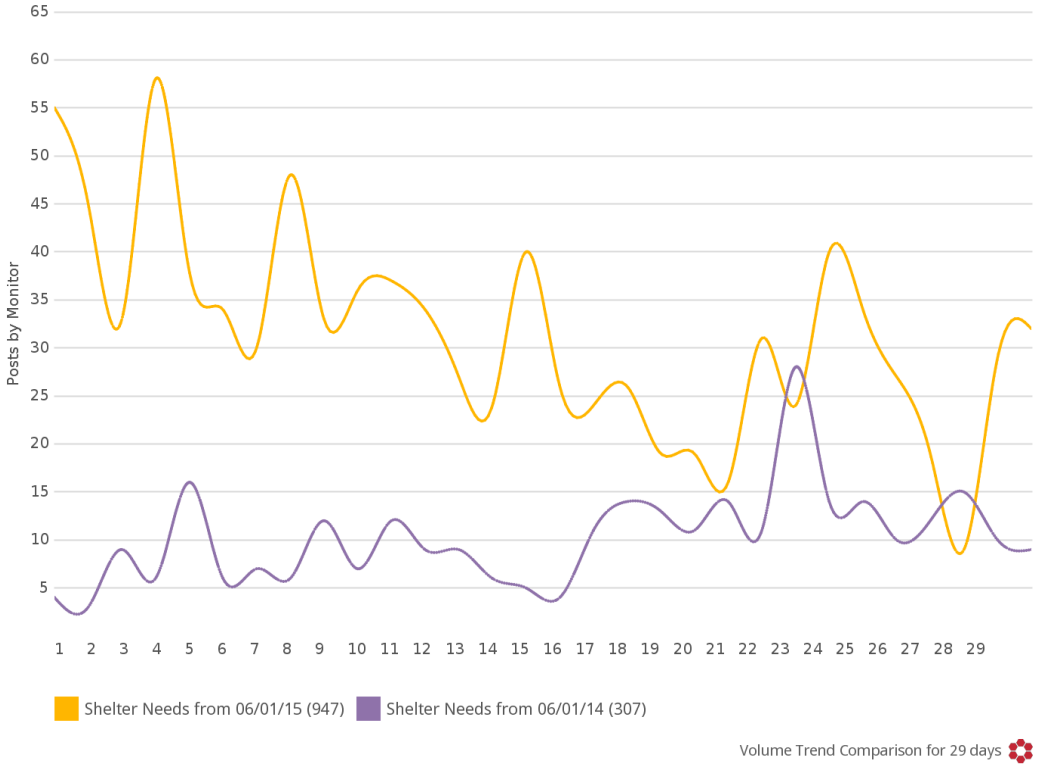
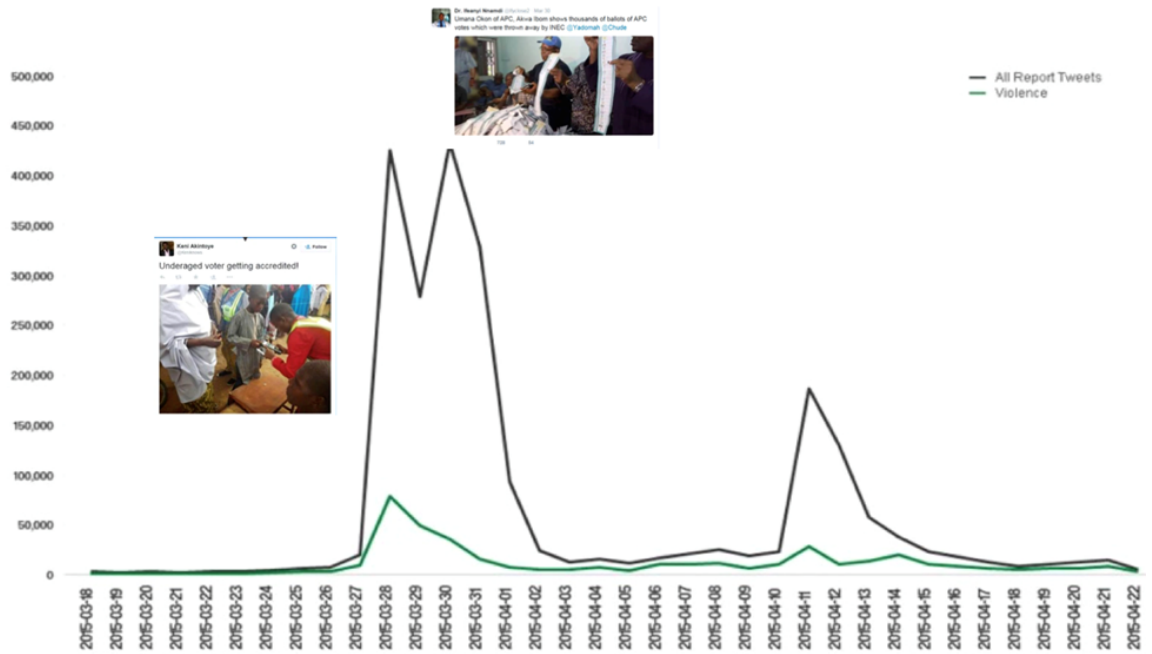


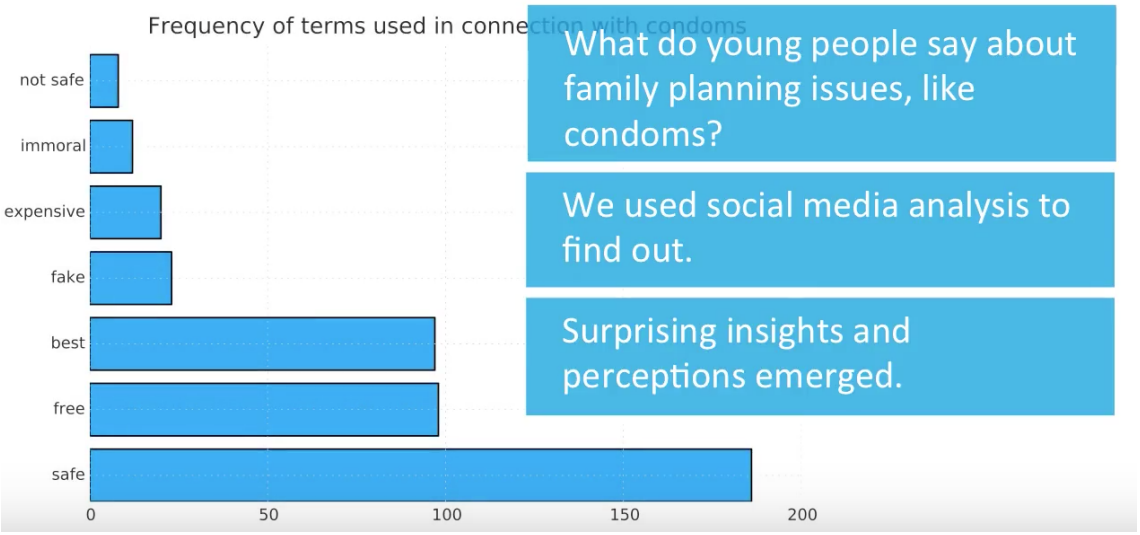
Figure 2:12 Frequency of tweets about the Nigerian elections (including tweets containing violent terms)



Social Media and the Nigeria Elections: Demos 2015

Lastly, Figure 2:13 shows an example from Pulse Lab Kampala where social media content was mined to identify what young people were saying about family planning issues. An analysis of the frequency of narratives related to condom use on social media included some counter-intuitive results, such as the presence of the term 'fake'.

Figure 2:13 Frequency of family planning topics on social media in Uganda



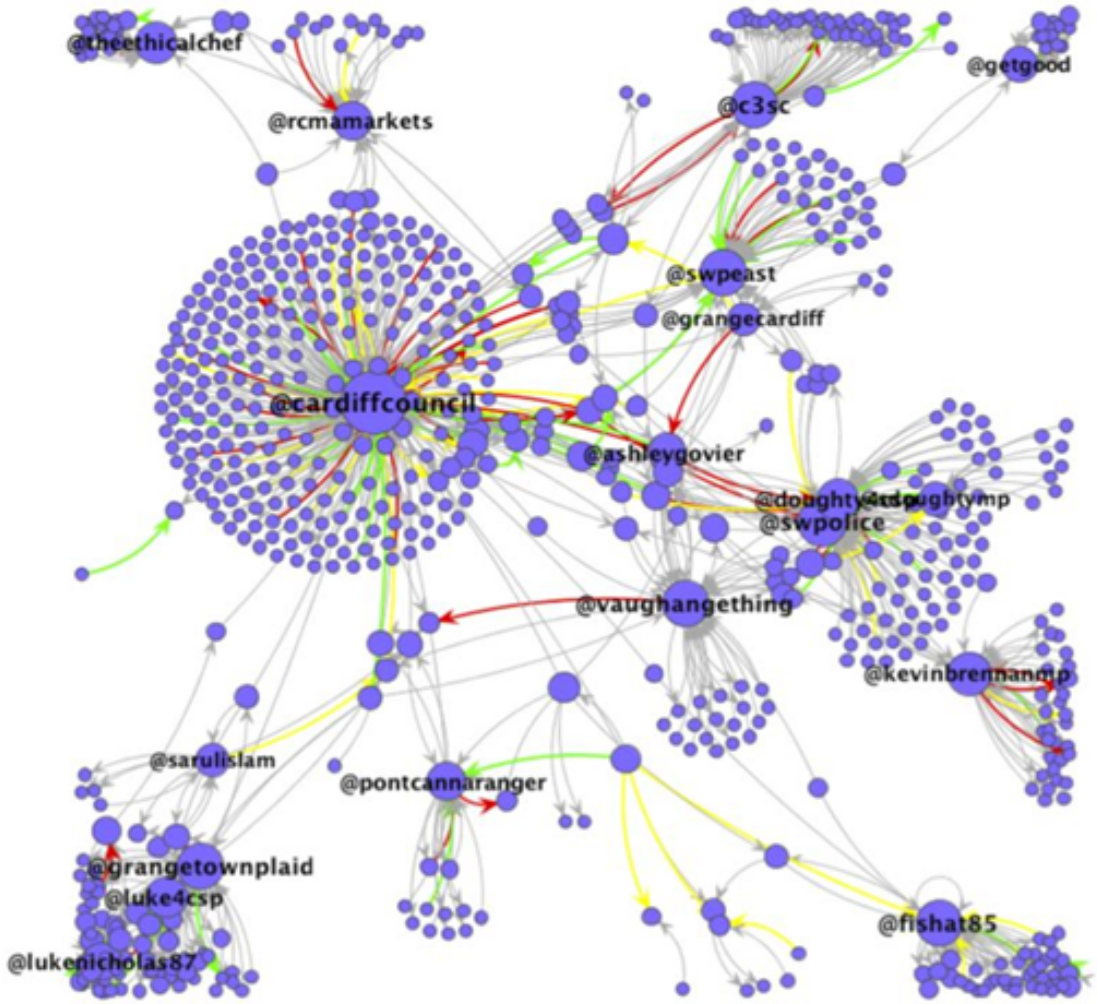
2.4 Social Network Analysis

Social network analysis is useful for visualising and describing the interactions of social media users. Each tweet is analysed to determine whether it contains any interaction with another Twitter user. This can include a direct mention or a retweet. The number of these interactions between users can be quantified and visualised to indicate the level of influence an individual has in a network and to identify important connecting users between groups.

2.4.1 Example of Social Network Analysis in COSMOS software

Figure 2:14 shows a network of over 17,000 tweets collected over a period of one month from the city of Cardiff (example from COSMOS data archive). The purpose of the network was to highlight prominent user accounts in the city. This can be used, for example, to identify users who are highly active in a political discussion so that researchers can analyse their political position, or to identify users who provide a link between influential politicians from different parties. In this example the colour coded edges (lines between nodes) indicate the level of sentiment expressed in the tweet text, allowing a researcher to identify the overall level of sentiment expressed in a given network.

Figure 2:14 Retweet network of tweets collected from the most active 30 Twitter accounts in Cardiff over one month



2.4.2 Examples of Social Network Analysis in a development context

Figure 2:15 shows a Twitter re-tweet network of users discussing the Ebola outbreak in West Africa in 2014 (example from COSMOS data archive). What is interesting are the differences in the network by gender - while the World Health Organisation feature as central node in both networks, the next most influential nodes are different for male and female tweeters.

Figure 2:15 Twitter network of the use of the keyword 'Ebola', split by gender

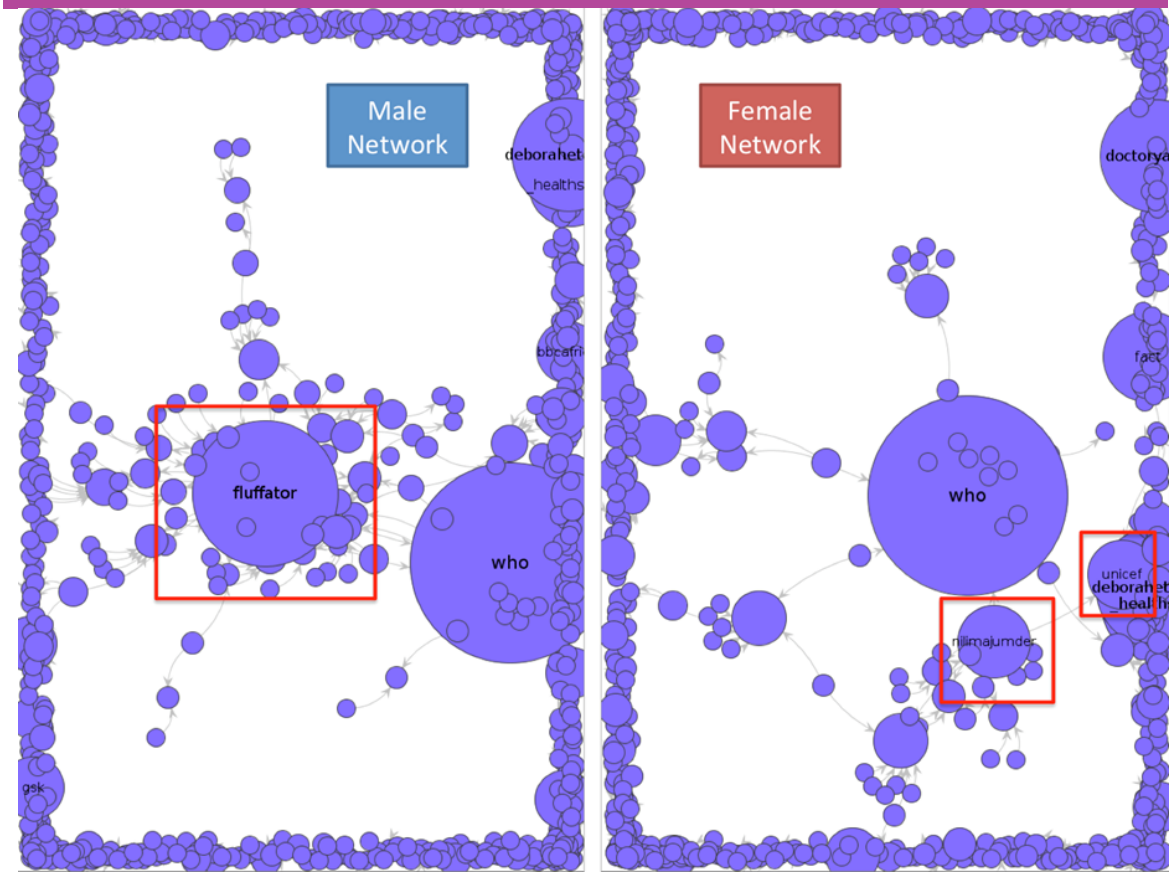
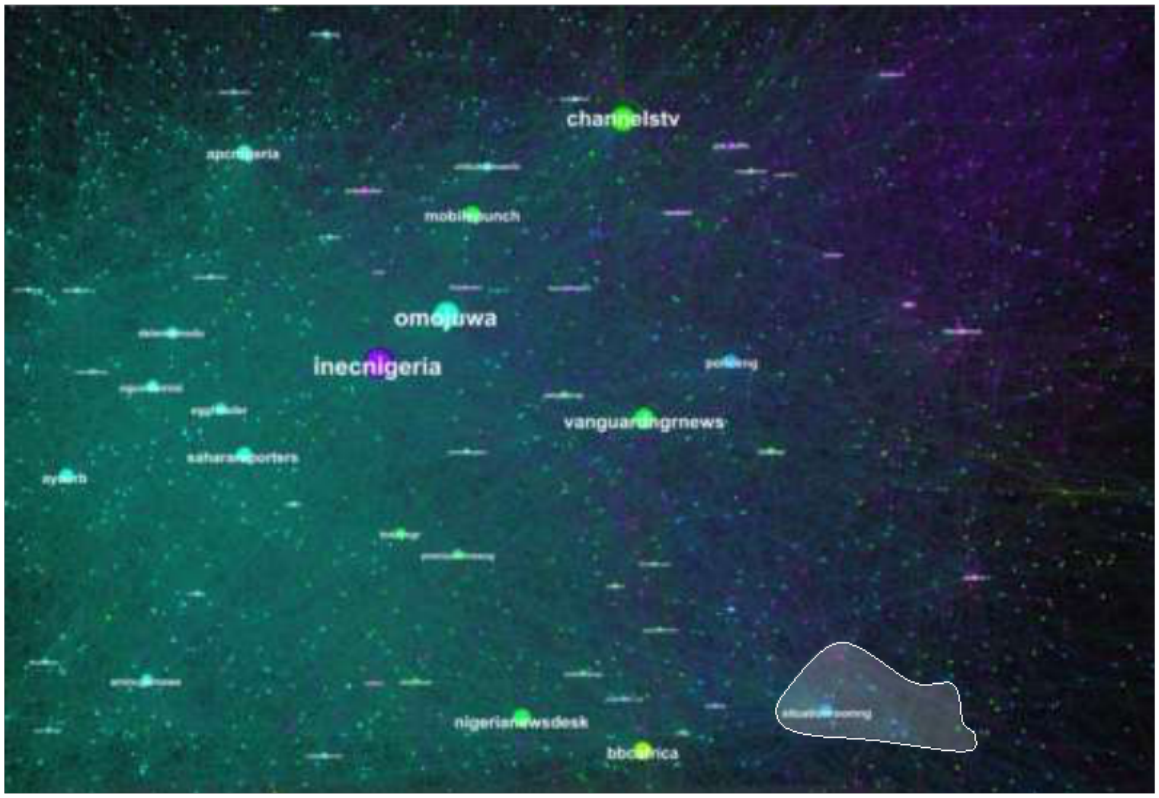


Figure 2:16 is visualisation of a Twitter mentions (when a user mentioned another user's Twitter handle in their tweet) network of the Nigerian elections (Bartlett et al. 2015, see Section 7). This network represents attempts from users to send public messages to others in the network. The network shows that the account 'Inecnigeria' is mentioned most frequently by a diverse array of users (indicated by their size and centrality in the network). Conversely, DFID related accounts, highlighted in white, are densely clustered and located in the bottom left of the network, indicating these accounts often mention each other, or are all mentioned by similar accounts. This suggests a limited reach in the spread of messages from these accounts in relation to the election.

Figure 2:16 Twitter network of Nigeria elections mentions



2.5 Topic Classification

Many social media analysis platforms provide topic classification tools. Topic classification can utilise an array of algorithms to reduce vast amounts of textual information into summaries of the main themes of the tweet corpus. In its most simple form, the frequency of terms is visualised in a wordcloud, with more frequent terms appearing larger in the visualisation. In platforms such as COSMOS words can be removed from view with a right mouse click, allowing the researchers to refine the visualisation by removing irrelevant content.

More sophisticated general topic classification involves attempts to summarise text based on the relationships between terms in strings of text, thereby providing the researcher with a more nuanced overview of the data. Researchers have begun to develop topic classification tools for social media to identify specific types of language use, such as hateful, antagonistic and threatening speech. These classifiers remove all non-relevant tweets and only present content that is deemed to be of the class of interest by the algorithm, allowing researchers to study specific types of language on social media in isolation.

2.5.1 Examples of topic classification analysis in a development context

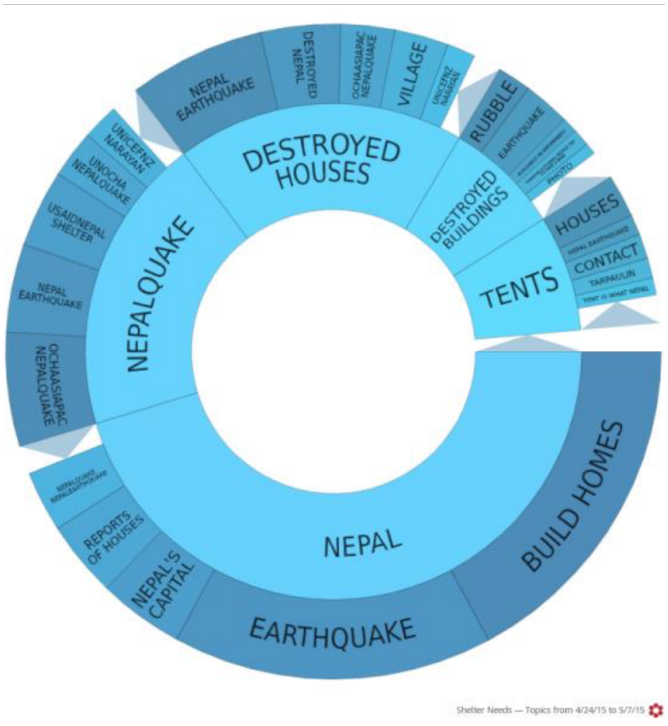
Figure 2:17 shows wordclouds for tweets sent by male and female users containing the keyword 'Ebola' (example from COSMOS data archive). We can see that while there are many similarities, '#fightebola' only appears in the wordcloud for female users.

Figure 2:17 WordCloud (frequency of terms) of tweets containing the keyword Ebola by male and female users

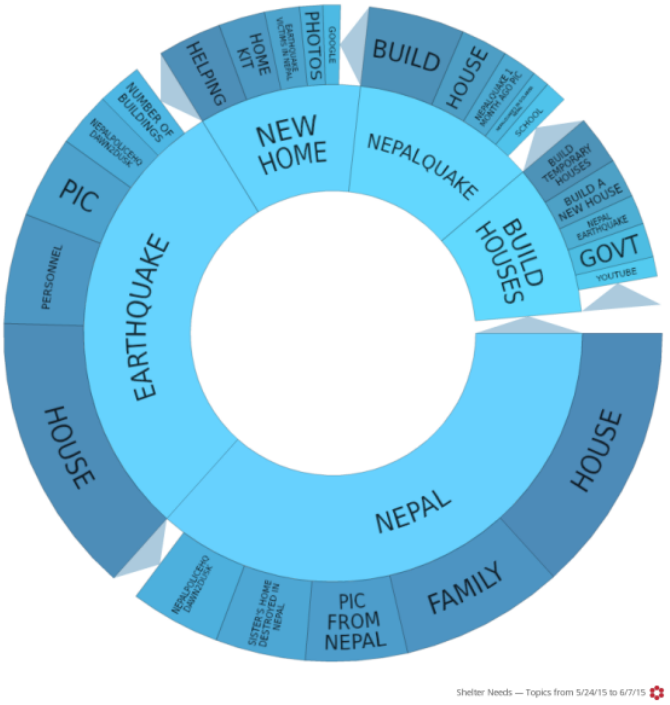
The ACAPS (2015, see Section 7) project supporting the Nepal Earthquake Assessment Unit found that topic analysis of social media enabled analysts to demonstrate shifts in conversations over time. The assumption was that topics of high importance were being discussed more frequently on social media than topics of low importance. A change in volume and topics was therefore an indicator of which topics were more important to the demographic that is using social media. Figure 2:18 shows how, in the first two weeks after the 25th April earthquake, conversations on social media relating to shelter focused mainly on destruction and emergency shelter solutions like tents. Four weeks later, the conversation had shifted to issues related to reconstruction.

Figure 2:18 Frequency and proximity of terms identified as part of the shelter query two weeks and four weeks after the Nepal earthquake.

Two weeks after earthquake



Four weeks after earthquake



3 Adding depth to social media data analysis

One of the criticisms of using social media data for social research is that although they are 'big', they are also 'light', meaning that they have little analytical power. One of the ways that researchers can improve the value of social media data is to create contextual information by linking them to external datasets or deriving characteristics from the content of social media text.

3.1 Data linking

Linking social media data with curated and administrative data allows researchers to build a multi-layered picture of a given situation. For example, tweets that contain comments indicating rising tensions between groups could be mapped using geospatial classification, and areas characterised by proportionally high levels of social tension identified. This could then be cross-referenced with official statistics on, for example, the religious and ethnic population composition of those areas.

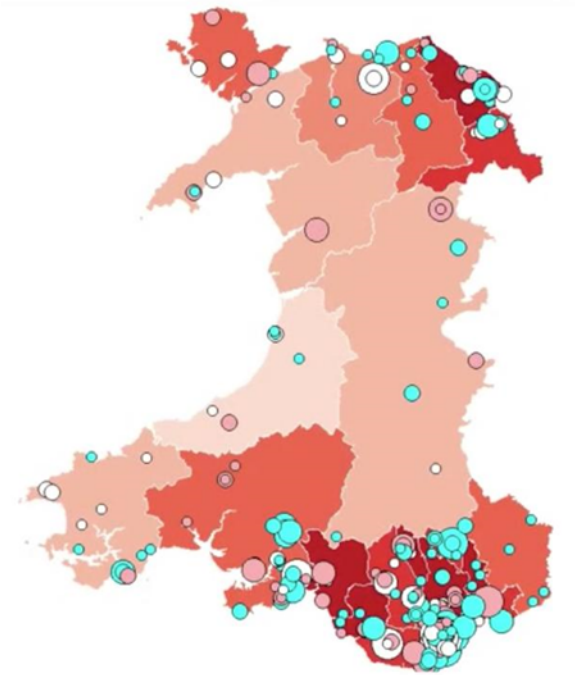
Linking to social media data using COSMOS

Currently, COSMOS is the only free to use platform that allows researchers to link these multiple data forms in a single software environment. Within the UK, the platform can link data from Twitter to the Office for National Statistics API (providing access to Census data as well as other national datasets), while internationally it can be linked to RSS feeds (providing access to rolling news and other feeds, such as earthquake alerts). Researchers can also load bespoke datasets into COSMOS using its JSON and CSV import feature.

All these data formats can be linked if they have a common geographic reference format (e.g. lat/long coordinates for each data entry or a reference to common geographical region). Currently COSMOS presents the user with an outline map of the UK broken down by local authority (LA) area (with an option for isolated views of England, Scotland, Wales and Greater London). Each LA can then be coloured according to an aggregated demographic variable from Census data, for example.

Figure 3:1 shows the percentage of the population that is in full time employment in Wales, with a layer of geolocated tweet data collected around the UK general election presented by gender (colour coded) and political sentiment expressed (size coded).

Figure 3:1 Map of Wales overlaid with Census data (proportion in full-time employment) and geolocated general election tweets by gender (colour coded) and sentiment (size coded)



For social and government researchers, this ability to link geospatial, Twitter user characteristics (e.g. age, gender, occupation), tweet text classification (e.g. political sentiment, social tension, hate speech) and census data via an interactive map allows for the real-time formulation of hypotheses for future testing.

3.2 Demographic classification

As well as linking external sources to social media datasets, demographic information can be extracted, or derived, from the social media content itself. This section focusses on classifying gender, age, social class/occupation and language using Twitter data, but it may be possible to extract/derive other characteristics (e.g. ethnicity, religion, sexual orientation). However, attention should always be paid to the validity of these classifications where they are derived.

3.2.1 Gender Classification

Gender classification is used to derive the portrayed gender of the person who posts a tweet (the tweeter).

Many social media software platforms perform this type of classification in an attempt to improve the usability of these data for research. Twitter gender detection algorithms work by analysing the content in the name field of the Twitter profile, by analysing the text of a user's tweets to identify gender specific language, or a combination of both.

For speed, most software adopts the first method. The first name is extracted from each Tweeter's profile metadata and is mapped onto a database of names that have been manually classified as male, female and unisex. Most software classifies names into one of four categories: male, female, unisex and unknown. Typically software is able to determine the gender of 50% of users.

The COSMOS software uses the 40,000 Namen (40N) database to identify the gender of Twitter users. Table 3.1 shows the results of the COSMOS gender classification tool run on a random sample of Twitter users. Those unclassified will be partially due to users not using their real name, or using names not part of the 40N database, which is mostly based on European names. There is currently no similar database for regions or countries within DFID's remit.

Table 3.1 Gender classification of sample of Twitter users

Gender	Proportion of Twitter users
Unclassified	52%
Male	22%
Female	23%
Unisex	4%

Base: random sample of 13M Twitter users

Within the UK, COSMOS typically classifies 48.8% of Twitters users as male, and 51.2% as female (excluding unisex/unclassified), matching the 2011 Census Statistics population estimates for England and Wales (ONS 2011), indicating that the 1% random sample supplied by Twitter is representative of this population in this regard (Sloan et al. 2013).

3.2.2 Age Classification

Although Twitter profiles do not have an age field which can be extracted via the API, researchers can identify the age of a user from their profile description. Some software, such as COSMOS, performs this classification task automatically.

Often profile descriptions in Twitter accounts contain details that can be used to estimate the age of the account holder (for example where profile information includes a number followed by the word 'years'). Age pattern matching is limited by language and can only identify the age of users who have English language profiles.

Based on the age classification of UK users, the population of Twitter users is much younger than the UK population as a whole. Table 3.2 shows that almost two-thirds of UK Twitter users are aged 20 or under, and over 90% are aged 30 or under. The fact that younger people are more populous on social media as a proportion of the user population is well established, but even the 1.1% of users between the ages of 51 and 60 could account for approximately 165,000 people in the UK (Sloan et al. 2015).

Table 3.2 Age classification of sample of UK Twitter users

Age group	Proportion of UK Twitter users	Estimated number of users
13-20	59.4%	8,955,000
21-30	31.6%	4,680,000
31-40	4.4%	630,000
41-50	3.4%	510,000
51-60	1.1%	165,000
60+	0.3%	45,000

Similar analysis can be conducted for regions and countries within DFIDs remit, but no studies currently exist.

3.2.3 Social Class & Occupation Classification

The profile description allows users to make a statement about hobbies and information relating to employment, allowing for data to be extracted to classify the occupation and social class of users. By using the SOC2010 look-up table provided by the Office for National Statistics, researchers can cross-reference all text from this tweet field with the occupations list for classification purposes. Some social media analytics platforms, such as COSMOS, automate this process (Sloan et al. 2015).

Occupation classification of Twitter users remains in its infancy and is mainly constrained to Twitter users in the US and UK. More research is required before this feature can be used in other regions and in social research with confidence.

3.2.4 Language Classification

Language classification is used to determine the language used in the text of the tweet. Language classification is important in international development research, monitoring and evaluation for a number of reasons, including the analysis of language use (e.g. proportion of languages mentioning keywords, and geospatial distribution), and the pre-processing of data to remove irrelevant content.

As well as providing an important demographic characteristic for Twitter users, language detection enables researchers to improve the efficiency of subsequent tweet analyses. For example, one of the techniques that can be applied to each tweet is sentiment analysis (see Section 2.2). Most current sentiment-analysis tools are built to process English-language text[6] so by detecting the language in which a tweet was written, researchers can efficiently skip non-English-language tweets when performing sentiment analysis and other English-language analyses.

Methods of classifying language of tweets and Twitter users
The language preferred by the Twitter user can be determined with two methods:

The first method takes the language that Twitter users set in their profile. This language setting specifies in which language Twitter users prefer to interact with the Twitter website. For example, the default language of the Twitter website is English, but French-speaking Twitter users may prefer to see a French-language version of the website. This language setting gives researchers a strong indication of each user's preference for and proficiency in a particular language.

The second method of identifying the language used by Twitter users is to analyse the text of the tweets, for example by using the Language Detection Library for Java (LDLJ). The LDLJ software recognises a comprehensive subset of the languages in which the worldwide Twitter community writes their tweets. In a sample of 113 million tweets (supplied by COSMOS), 99.3% were written in a language identifiable by the software. Table 3.3 details the number of tweets written in languages identified by the software in the sample (Sloan et al. 2013).

[6] Although also available for Arabic Twitter data

Table 3.3 Number of tweets written in each of the 53 languages identified by the LDLJ software

Language	Number of Tweets	Language	No. of Tweets
English	45,594,240	Hungarian	235,894
Japanese	12,738,687	Slovak	219,654
Spanish	10,136,337	Lithuanian	200,237
Indonesian	9,142,131	Albanian	178,708
Portuguese	6,991,330	Vietnamese	162,587
Arabic	3,172,589	Czech	108,080
Somali	2,553,774	Persian	105,789
Dutch	2,240,281	Latvian	95,477
Tagalog	1,899,788	Simplified Chinese	86,284
French	1,767,104	Bulgarian	80,332
Italian	1,705,202	Greek	78,015
Turkish	1,536,013	Traditional Chinese	52,063
German	1,446,948	Urdu	40,736
Korean	1,337,590	Macedonian	37,081
Afrikaans	1,307,274	Ukrainian	27,455
Estonian	1,223,220	Hebrew	12,827
Thai	836,832	Tamil	4,933
Finnish	833,097	Hindi	1,942
Russian	728,551	Nepali	1,420
Swahili	721,658	Bengali	936
Norwegian	709,519	Malayalam	898
Slovene	547,050	Punjabi	688
Danish	521,356	Marathi	442
Swedish	477,127	Telugu	220
Polish	395,858	Kannada	183
Romanian	384,550	Gujarati	183
Croatian	319,283		

Base: All tweets from a 1% subsample of 113M tweets where a language was identified by the LDLJ software

4 Data Access

Social media researchers have experimented with data from a range of sources, including Facebook, You Tube, Flickr, Tumblr and Twitter to name a few. Twitter is by far the most studied of all these networks. It differs from other networks such as Facebook, in that it is public and the data (in part) are freely available to researchers. Twitter also has an open friendship network (non-reciprocal linking between users means that the followed are not required to follow their followers) resulting in a digital 'public space' that promotes the free exchange of opinions and ideas. As a result Twitter has become the primary space for online citizens to publicly express their reaction to events of national significance. A hashtag convention has emerged amongst Twitter users that allows tweets to be tagged to a topic which is searchable. The term 'trending' is used to describe hashtags that become popular within the tweet-stream, indicating a peak or pulse in discussion usually surrounding an event. Hashtags and keywords therefore make it relatively straightforward for researchers to identify reactions to major incidents, news stories and events and to quickly collect data in real-time via the Twitter Application Programme Interface (API).

4.1 Application Programme Interfaces

4.1.1 Real-time data

The Twitter Streaming API is more open and accessible compared to other social media platforms. Twitter provides three levels of data access (the lowest of which is free) and the data can be obtained using an online query or dedicated software. The free random 1% of the Twitter stream is dubbed the 'spritzer'. Both the 'garden hose/deca hose' (providing access to a random 10%) and the 'fire hose' (providing 100% access) can be obtained via agreement with Twitter for research or can be purchased via their data-resellers e.g. Gnip (https://gnip.com). For most social research projects the 1% feed is usually sufficient and provides access to approximately 3.5-5 million tweets per day free of charge. Several social media data collection and analysis software platforms make accessing the 1% stream of data straightforward (see Section 5).

4.1.2 Historic data

The Twitter Search API provides free access to historical data up to 7 days into the past from the day of the query. It is important to know that the Search API is focused on relevance and not completeness. This means that some Tweets and users may be missing from search results. For research where match for completeness is important researchers should consider using the Streaming API (above) instead. For social media posts over 7 days old data can be purchased via Twitter's chosen data-resellers e.g. Gnip (https://gnip.com). The cost of data is dependent upon volume of returned posts and the length of the query. For example, a query spanning 1 week that returns 1M posts may cost the equivalent of a query spanning 1 year that returns 10K posts.

5 Tools for social media analysis

5.1 Free-to-use social media analysis tools

Software tools have been developed that provide access to social media data and suitable forms of analysis, including topic detection, sentiment classification and network analysis.

Table 5.1 includes only the tools that are free to use (most of which have been developed in an academic environment). Paid for software and services (such as Radian6, ripjar, Pulsar and DataMinr) are available but tend to be tailored to commercial solutions (e.g. brand tracking) and provide metrics that are uninspectable (e.g. classification of users) due to intellectual property issues.

Table 5.1 Free to use social media data collection, visualisation and analysis tools.

Tool	Operating System	Download from	Data sources
COSMOS	Windows Linux (Ubuntu) Mac OS X	www.cosmosproject.net	Twitter RSS feeds Survey data via ONS API
Webometric Analyst	Windows	lexiurl.wlv.ac.uk	Twitter YouTube Flickr
NodeXL	Windows	nodexl.codeplex.com	Twitter YouTube Flickr
Netlytic	Web based	netlytic.org	Twitter Facebook YouTube Instagram
Mozdeh	Windows	mozdeh.wlv.ac.uk/installation.html	Twitter
Twitter Arching Google Spreadsheet (TAGS)	Web based	tags.hawksey.info/	Twitter
Chorus	Windows	chorusanalytics.co.uk/chorus/request_download.php	Twitter
Visibrain Focus	Web based	www.visibrain.com/en	Twitter

5.2 Cardiff Online Social Media Observatory (COSMOS)

COSMOS is a social media data collection, analysis and fusion platform. It programmatically collects data from a number of sources using publicly accessible APIs – with a particular focus on Twitter.

COSMOS has been collecting a random 1% sample from the Twitter API (commonly referred to as the 'spritzer') since 2012 and the database currently holds over 4 billion tweets. It is also possible to collect 10% from the API ('gardenhose') or the full 100% of all tweets ('firehose') with permission from Twitter. However, the data storage

requirements for 10% and 100% are impractical for many social research establishments.

COSMOS also has a persistent connection to the Office for National Statistics API allowing access to all national curated datasets with the capacity for linking (data fusion) these with social media data geographically. Data import also allows for the loading of CSV files and RSS feeds, meaning almost any quantitative and qualitative data source can be subject to the analytical tools within COSMOS.

COSMOS provides a single interface to a number of tools, with no data collection overhead and automated translation of input files from one format to another. For example, a user can extract data from the COSMOS archive of Twitter data, containing the term “Election” and associated political party names and voting stations posted during a voting period in a developing country, and pose a number of research questions such as:

- Is the keyword “violence” being used in combination with these search terms?
- How does sentiment change over time and in relation to news reports?
- Who is talking to whom about the election and are there any leading actors or bridges between communities in the social network?
- Are there gender-based differences in opinion at certain points in time?
- Can we identify clusters of geo-located data in different regions?

Each of these questions can be interrogated by using the various big data inspection tools within COSMOS, including sentiment tracking, geospatial visualization, social network analysis and demographic classification. COSMOS has an extendable architecture, allowing users to develop new analysis tools specific to their analysis needs. For example, users have developed classifiers for suicidal ideation, community tension, hateful sentiment and counter hate speech.

6 Ethics & Social Media Research

Ethics has emerged as a contentious area of debate in the use of social media data in social and government research. For example, a recent study conducted on Facebook that claims to have altered the emotions of users via tailored content did not obtain informed consent from participants (Kramer et al. 2014). Facebook and the researchers received wide-spread criticism in the international media. This section outlines some of the key issues.

6.1 Informed consent to research

The issue of informed consent is complicated in Internet based research, especially with social media data.

Many social media terms of service specifically state that users' data will be sold to third parties, and it can be argued that by accepting these terms users consent to research. These terms of service need to be interpreted and engaged with through the principles and practice of social and government research, and it can be contested whether this truly counts as *informed* consent – even if they read the terms of service, do users truly understand how their data may be used?

6.2 Anonymity and publishing tweet content

The publication of quantitative findings from social media research is largely ethically unproblematic as the data are aggregated. However, it is likely that the publication of qualitative findings will be enhanced by the inclusion of full tweet text, yet problems arise here as this makes users identifiable. While Twitter do not provide guidance for social researchers, they do provide a set of Best Practices for Media (static uses and publication):

- Show name, @username, unmodified Tweet text and the Twitter bird nearby, as well as a timestamp
- If displaying Tweets, make sure they are real, from legitimate accounts and that you have permission from the author when necessary
- Display the associated Tweet and attribution with images or media
- If showing screenshots, only show your own profile page, the @twitter page, the Twitter "About" page or a page you have permission from the author to show

Twitter provide additional information for developers on the maintenance of their products. Violation of these guidelines can result in Twitter taking action, such as preventing access to their data.

Maintain the integrity of Twitter's products:

- @username must always be displayed (and name if possible) with tweet text
- Respond to content changes such as deletions or public/private status of tweets
- Do not modify, translate or delete a portion of the content

Respect Users' Privacy and get the user's express consent before you do any of the following:

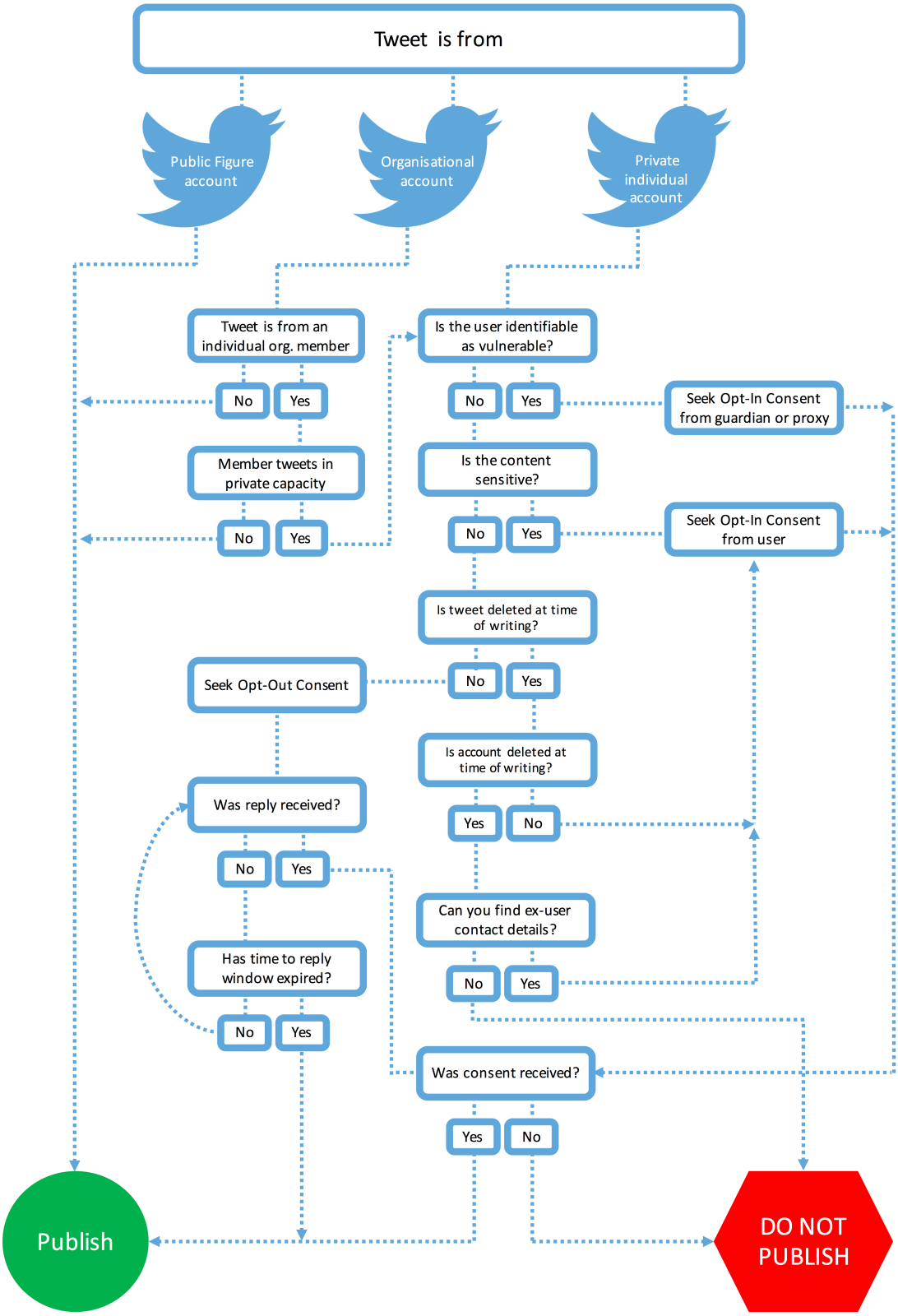
- Take any actions on a user's behalf, including posting Content and modifying profile information
- Store non-public Content such as direct messages or other private or confidential information
- Share or publish protected Content, private or confidential information

Twitter's Terms of Service require users to provide their consent for Twitter to share any content posted with third parties. While it might be acceptable for social researchers to accept users have provided informed consent for their data to be shared with them under these T&Cs, they should not accept that this provides them with informed consent to publish the content of individual tweets (anonymised or not). Doing so could put users at risk of harm, including reputational, personal and/or physical.

For example, in a project examining the spread of hate speech on Twitter, it would be irresponsible to directly quote a tweet using extreme racist language without consent from the user for two reasons:

1. Tweets cannot be anonymised by removal of the username because Twitter guidelines suggest that all reproduction of content should be accompanied by the username (see above); and even if the username was removed from the content, the text of a tweet is Internet searchable.
2. Publication of content renders the user discoverable, opening up the possibility for harm via the targeting of counter hate speech, or at the extreme extent, the targeting of physical violence (possible if the user enables geolocation).

The decision flow chart below has been informed by Twitter guidelines and assists researchers in choosing whether or not to publish the original content of tweets.



Deciding on the status of tweeters (e.g. public, organisational, private, vulnerable) and their tweets (e.g. organisational, private, sensitive) is at the discretion of the researcher and/or the ethics review board. In seeking to reach these decisions researchers should consult existing ethical guidelines (see Section 6.4) that provide definitions of public figures (e.g. politicians and celebrities who aim to communicate to a wide audience), vulnerable individuals (e.g. children, learning disabled and those suffering from an illness) and sensitive content (e.g. posts about criminal activity, financial problems, mental health issues and feelings of suicide, extramarital sexual activity, controversial political opinions and activism). As social media accounts can lack personal details, and it is difficult to find additional identifying details, researchers and ethics review boards may be satisfied with the use of the information presented on the profile and in posts alone to reach decisions on the status of users.

6.3 Sharing Data

Twitter data are collected from the API using the researcher's Twitter account credentials. Datasets larger than 50,000 tweets cannot be shared for any purposes beyond the single user (or their immediate research team or government department).

In order to make datasets larger than 50,000 tweets sharable it is possible to provide tweet IDs and user IDs to other researchers outside of the immediate team or government department. Those who wish to reproduce research results or conduct secondary analysis can then use these data to query the API to re-vivify the IDs into full tweet objects (including the text). However, it is important to note that the revivified dataset will contain any deletions implemented by Twitter users or Twitter themselves.

6.4 Existing guidelines

The Association of Internet Researchers (AoIR) ethical guidelines highlight three key areas of tension (AoIR 2012):

The question of human subjects online
The notion of the 'human subject' is complicated when applied to online environments. For example, can we say semi-automated 'Bots' (hybrid human-machine social media accounts that post and retweet) are human subjects? Does digital representation and automation of some online 'behaviours' call into question the definition of human subjects in Internet based research?

Data/text and personhood
The Internet complicates the conventional construction of 'personhood' and the 'self', questioning the presence of the human subject in online interactions. Is digital information an extension of a person? In some cases this may be clear-cut: emails, instant message chat, newsgroup posts are easily attributable to the persons that produced them. However, when dealing with aggregate information in social media datasets, such as collective sentiment scores for sub-groups of Twitter users, the connection between the object of research and the person is more indistinct. Attribute data on very large groups of anonymised Twitter users could be said to constitute non-personalised information, more removed from the human subjects that produced the interactions as compared to, say, an online interview. In these cases, the AoIR (2012: 7) guidelines state 'it is possible to forget that there was ever a person somewhere in the process that could be directly or indirectly impacted by the research'. Anonymisation procedures for big social data and linked data are in their infancy and researchers are not yet fully aware of the factors that may result in the disclosure of identity and subsequent potential harms.

The public/private divide

It is accepted that people who use online 'public' spaces can perceive their interaction as private. This can question the use of APIs that make accessible to researchers communications that were intended for private consumption. The AoIR (2012: 7) guidelines state that social, academic and regulatory delineations of the public-private divide may not hold in online contexts and as such 'privacy is a concept that must include a consideration of expectations and consensus' within context.

Cabinet Office Data Science Ethical Framework

The Cabinet Office Data Science Ethical Framework (Cabinet Office 2016) applies to all government research. It includes the following ethical guidelines that are relevant to research and evaluation and in development contexts:

> "The Data Protection Act requires you to have an understanding of how people would reasonably expect their personal data to be used. You need to be aware of shifting public perceptions. Social media data, commercial data and data scraped from the web allow us to understand more about the world, but come with different terms and conditions and levels of consent." (p.4)

> "…public attitudes to data are changing. Working with data in ways which makes the pubic feel uneasy, without adequate transparency or engagement, could put your project at risk and also jeopardise other projects... Consideration of public attitudes and communication with them is key". (p.3)

> "You should always use the minimum data necessary to achieve the public benefit. Sometimes you will need to use sensitive personal data. There are steps you can take to safeguard people's privacy e.g. de-identifying or aggregating data to higher levels, querying datasets or using synthetic data." (p.4)

> "Using data that is voluntarily in the public domain (e.g. social media data) needs careful consideration. Legally it is personal data and needs to be processed fairly (i.e. in line with the T&Cs of the social media provider)." (p.9)

The full Cabinet Office risk assessment checklist is provided in the appendix of this practice note.

6.5 Users' views

Recent work by NatCen and the Social Data Science Lab at Cardiff University shows how users of social media platforms are uneasy about their posts being used without their explicit consent (NatCen 2014, Williams 2015). A recent survey of approximately 600 UK social media users' perceptions of the use of their social media posts found the following:

- 94% were aware that social media companies had Terms of Service;
- 74% knew that when accepting Terms of Service they were giving permission for some of their information to be accessed by third parties;
- 56% agreed that if their social media information is used for academic research they would expect to be asked for consent;
- 77% agreed that if their tweets were used without their consent they should be anonymised;

- 82% were ‘not at all concerned’ or only ‘slightly concerned’ about university researchers using their social media information;
- 50% were ‘not at all concerned’ or only ‘slightly concerned’ about government departments using their social media information.

7 Detailed cases studies

This section outlines four case studies of social media data use in the context of International Development.

7.1 Data-Pop Alliance: Big Data for Disaster Resilience

Based on the findings of 11 case studies and pilot projects, this project concluded that there is mounting evidence that Big Data has the potential for increasing social resilience to disasters.

The projects highlighted four functions of Big Data for disaster resilience:

1. Descriptive - Involves narrative or early detection such as using data from social media to identify flooded areas or identifying areas in need from crisis maps

2. Predictive - Includes what has been called 'now-casting' - to make real-time inferences on population distribution based on social media activity before, during or after an event, as well as forecasting sudden and slow onset hazards

3. Prescriptive (or diagnostic) - Goes beyond description and inferences to establish and make recommendations on the basis of causal relations, for instance by identifying the effects of agricultural diversification on resilience

4. Discursive (or engagement) - Concerns spurring and shaping dialogue within and between communities and with key stakeholders about the needs and resources of vulnerable populations via social media (for example to assist disaster relief efforts)

The use of Big Data to build resilience generally falls into one of five categories throughout the disaster cycle:

1. Monitoring hazards.

Social media offers a degree of remote sensing capability. For example, adding information from Twitter feeds offers tremendous potential for monitoring hazards such as earthquakes and floods.

2. Assessing exposure and vulnerability to hazards.

Crowdsourcing initiatives using social media can empower volunteers to add ground-level data that are useful notably for verification purposes. Social media posts with geo-location and time stamp data can be used when estimating moving populations.

3. Guiding disaster response.

Social media can be monitored to provide early warning on threats ranging from disease outbreaks to food insecurity. Remote sensing has been used to provide early assessment of damage caused by hurricanes and earthquakes.

4. Assessing the resilience of natural systems.

Citizen science reporting via social media and other platforms can radically expand scientists' observations of ecological systems.

5. Engagement of communities.

Building long-term resilience takes more than enhancing the ability of both external and local actors to react to single events. Resilient communities manage their natural systems, strengthen their infrastructure, and maintain the social ties and networks that make communities strong. The longer-term potential of Big Data lies in its capacity to raise citizens' awareness and empower them to take action. Decisions that facilitate or hinder this capacity are fundamentally political ones.

However, the pilot projects also highlighted the limits of big data for resilience:

Constraints on data access and completeness. For all the talk about the 'data deluge', most Big Data sets are controlled by private corporations, and as of yet no comprehensive frameworks and principles for data sharing exist. The tools to gather and process these data also tend to be difficult to use and expensive.

Analytical challenges to actionability and replicability. Big Data sets and streams face issues of reliability and representativeness that may hamper internal and external validation of findings derived from their analysis. Approaches to mitigate these effects such as verification techniques and sample bias correction methods have been or are being developed.

Human and technological capacity gaps. At present the capacity to gather and analyse data, as well as the ability to integrate it into policy making and programming are still largely lacking – especially among the institutions of the Global South.

Bottlenecks in effective coordination, communication and self-organization. The knowledge people need to inform risk assessment, preparedness and response efforts come from many sources that are rarely coordinated, and socio-cultural and psychological factors are too often ignored, notably the need to build knowledge and exchange networks rather than provide information products.

Ethical and political risks and considerations. The potential for unethical or even dangerous use of Big Data grows exponentially in developing countries and there is an urgent need for developing ethical guidelines rooted in the long history of ethics in social science and medical research. Participation must be voluntary, users' data must be protected, and the needs of people without access to technology must be addressed.

More information on this case study can be found here: http://datapopalliance.org/wp-content/uploads/2015/11/Big-Data-for-Resilience-2015-Report.pdf

7.2 Assessment Capacities Project (ACAPS) - Nepal Earthquake Disaster Response

Following the earthquake on 25 April 2015, the Assessment Capacities Project (ACAPS) was deployed to support the Nepal Earthquake Assessment Unit through data analysis, assessments and identifying key needs.

As part of this role, ACAPS was asked to feed into the "Communication with Communities" (CwC) project with insights gained through social media (mainly Twitter, Facebook, YouTube, Flickr and blogs) and national media monitoring. The social media monitoring pilot was set up to monitor social media conversations related to the earthquake, as well as local media.

The monitoring was done in English and Nepali. Issues of main interest for the social media monitoring pilot were:

- Needs, concerns, developing trends and emerging risks of the effected population as expressed on social media
- Conversations related to the quality and accessibility of aid

In the context and timeframe of the ACAPS project, social media was mainly useful in:

- Analysing public reactions to media reports
- Showing the relative prevalence of topics and identify changes compared to the onset of the emergency, as well as to before the earthquake

However, social media analysis also had a number of limitations in this context:

- Social media monitoring was not useful in breaking down needs geographically
- It was not very useful in gaining insights into issues that are sensitive and generally not discussed publicly, such as protection issues
- Social media users in Nepal were overwhelming concentrated in Kathmandu. In Nepal, social media data are more suited to analysing issues that directly affect people in the capital than in rural areas
- The Twitter population is not representative of the national population, there are biases in relation to region (rural/urban), accessibility (technology, income, education), geo-location (not all users geo-locate), and propensity to tweet about natural disasters
- The Nepali language uses Devanagiri script, an alphabet that is not or not fully supported by most social media analysis tools. This severely limits the range of available software solutions

The project also outlined some recommendations for conducting social media analysis in the future:

- The volume of social media updates related to a rapid onset emergency is largest in the first days of the emergency. Social media monitoring should start as soon as possible to inform situational awareness and provide the most benefit to decision makers
- Product selection and contracting can take multiple weeks. Relevant products should already be pre-identified and framework agreements in place to facilitate a quick start when an emergency occurs
- A social media expert should be deployed on-site during the first phase of the emergency to set up and customize the systems, help train staff and increase awareness of the possibilities and limitations

- Having qualified, computer literate national staff who are familiar with social media, the local media landscape, the local geography and basic information management techniques is key

More information on this case study can be found here:
http://www.acaps.org/sites/acaps/files/resources/files/lessons_learned-social_media_monitoring_during_humanitarian_crises_september_2015.pdf

7.3 Mapping Refugee Media Journeys: Smartphones and Social Media Networks

The project was carried out between September 2015 and April 2016 and involved a range of academic and practitioner expertise. Smartphones and digital connectivity are essential for refugees seeking protection and safety in Europe. This project examined the use of Facebook and Twitter by migrating Syrian refugees, finding:

- Migrating refugees access international news sources via social media and news feed apps shared among friends and family
- Engagement with news is driven by curiosity and need to uncover the facts around major events or news of most direct personal or local relevance
- There is a notable fear of surveillance among refugees which results in them shrouding their identities on social media and online via use of avatars and pseudonyms
- This makes refugees online and on social media, especially those in transit, a particularly difficult group to research
- Despite this, if mobile phones are lost or damaged, Facebook accounts enable a permanent if intermittent perpetual presence
- Refugees connect to Facebook sites mainly in order to communicate with family, friends and influential figures in their social media networks - from prominent and respected activists and NGOs to investigative journalists, political commentators, public intellectuals and participants in controversial debates
- Individuals are perceived to be trustworthy when they give a clear commitment to supporting to the interests and welfare of refugees
- Although it is not easy to discern the identities of refugees online, when refugees were confident about privacy and/or anonymity, they expressed their political views without restraint and often in highly emotional registers and hyperbolic tones
- Relationships in social media networks are shaped not only by kinship and friendship but also by pragmatic and ideological factors. The spaces of social media discussion and debate among those identified as refugees tended to be fractious, intensely politicised and polarised
- Influential figures direct the flow of engagement and information within the social media network and this reinforces and maintains insular ideological enclaves
- Constant commuting between open/public and private/closed Facebook spaces was observed
- The most trusted and influential people on Twitter are those who are close to the ground in Syria and other conflict zones. They have friends, fans and followers who amplify their message content and opinions

- Key influencers can mediate between cultures, languages and groups and perform the role of cultural diplomat and broker

More information on this case study can be found here:
http://www.open.ac.uk/ccig/news/report-launched-for-mapping-refugee-media-journeys-project

7.4 Nigerian Election

The purpose of this research project was to develop an understanding of the effectiveness of social media use for communication and monitoring around the 2015 Nigerian election.

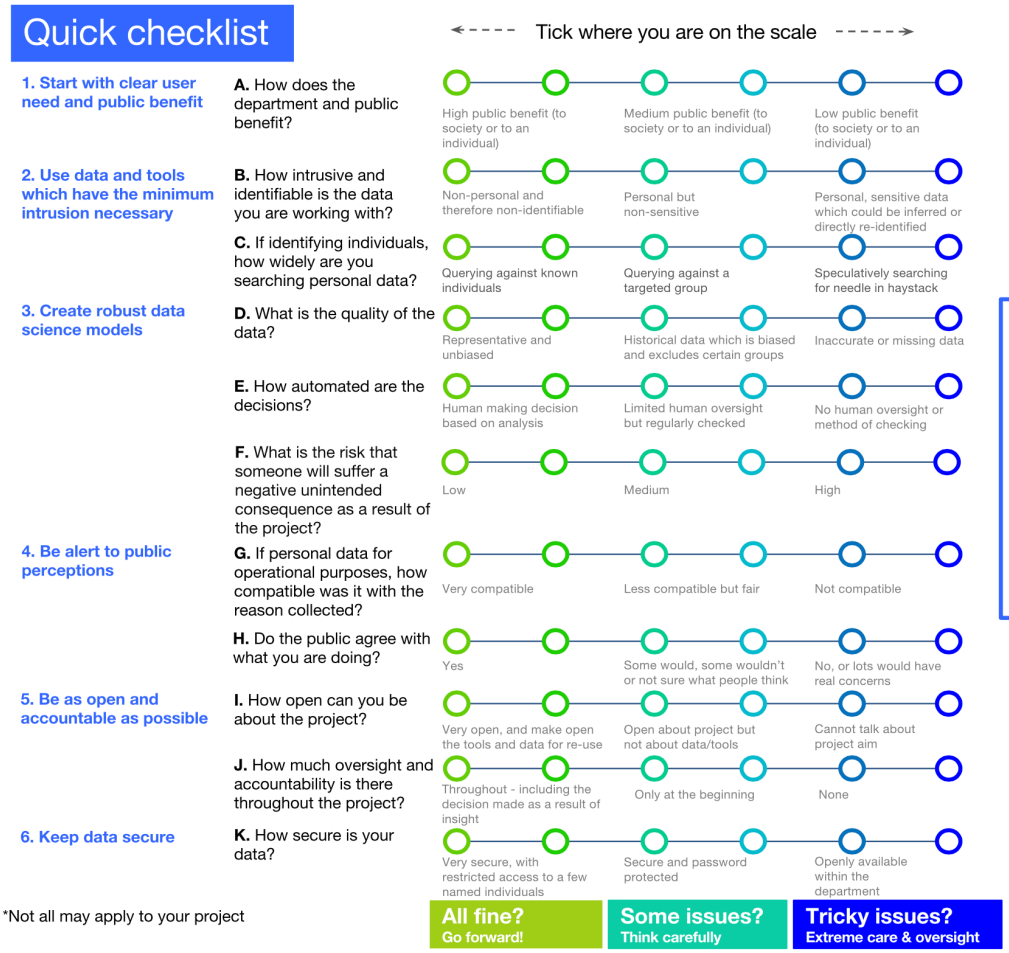
Thirty per cent of Nigerians use the Internet – of which 70 per cent are using social media (Facebook, YouTube and Twitter all count in the top ten most visited sites in Nigeria). The use of social media in Nigerian elections initially became noticeable in the preparations for the 2011 general elections. Social media was used to share information, for campaigning and to improve the efficiency of election observation.

Twitter was the most valuable data source for this project given its open structure and the resulting availability of data. The project identified a number of key points:

- Social media can be used before, during and after an election to monitor and evaluate the electoral process as it unfolds online
- It is useful for monitoring 'events that take place online' especially the spread of misinformation, providing opportunities for organisations to post counter-speech
- It was important to develop bespoke text classification tools to identify election topics and sub-topics. Twitter data tended to be divided into 'reportage' (i.e. people describing events) and 'comment' (i.e. people describing events)
- There was a significant volume of tweets about violence (408k); but a lot of this was about Boko Haram, due to the international interest in the group
- Much of the top content was widely shared news reports or campaigning – which demonstrates the value of more nuanced analysis of less popular content
- There was a significant volume of rumours being spread on Twitter; many of which were not being responded to via counter speech
- Around 4 per cent of the total dataset collected for the project was 'geo-tagged' with lat/long coordinates
- Social media could be used to gain a good understanding of the influencers of debates online. This helped the DFID Nigeria office to reach / follow / understand key influential actors and get a better understanding of the reach of existing / planned partners
- Social network analysis showed that the traditional DFID partners were clustered together meaning they communicated frequently with each other, but DFID communications were not reaching a wider population
- Social media could be used to rapidly identify trouble spots, but there may be biases present in social media use meaning some trouble spots get identified while others do not. Social media data should therefore not be relied upon in isolation, and should be triangulated with on the ground verification
- Social media was useful in picking up citizen claims of electoral misconduct, but biases in the use of social media limit its use in isolation for this purpose
- Social media analysts should be on the ground to set up monitoring systems, to identify, verify and escalate a response to any trouble. The CASE model is an excellent one to follow

More information on this case study can be found here:
http://r4d.dfid.gov.uk/Output/201974/

8 Appendix – Cabinet Office Data Science Ethical Framework Checklist



Some departments might find themselves at the left hand side of the scale, and others more on the right (blue), reflecting the nature of their department's work. This does not mean the project should not go ahead, but think carefully about it, and if possible, bring some elements to the green end of the scale.

5

9 References

ACAPS (2015) *Lessons Learned: Social Media Monitoring During Humanitarian Crises*, Geneva: ACAPS.

AOIR (2012) Ethical Decision-Making and Internet Research: Version 2.0 – Recommendations from the Association of Internet Researchers Working Committee. Available at: <http://aoir.org/reports/ethics2.pdf>

Bartlett, J., Krasodomski-Jones, A., Daniel, N., Fisher, A., Jesperson, S. (2015) *Social Media for Election Communication and Monitoring in Nigeria*, London: Demos.

BIS (2013) *Seizing the data opportunity: A strategy for UK data capability*, London: Department of Business, Innovation and Skills.

Cabinet Office (2016) *Data Science Ethical Framework*, London: Cabinet Office.

Choudhary, A., Hendrix, W., Lee, K., Palsetia, D. and Liao, W. (2012) 'Social Media Evolution of the Egyptian Revolution', *Communications of the ACM*, 55: 5, pp 74-80.

Data-Pop Alliance (2015) *Big Data for Climate Change and Disaster Resilience: Realising the Benefits for Developing Countries*, New York: Data-Pop Alliance.

Gillespie, M., Ampofo, L., Cheesman, M., Faith, B., Iliadou, E., Issa, A., Osseiran, S., and Skleparis, D. (2016) *Mapping Refugee Media Journeys Smartphones and Social Media Networks*, Milton Keynes: Open University.

Jongman, B., Wagemaker, J., Romero, B.R., de Perez, E.C. (2015*) 'Early Flood Detection for Rapid Humanitarian Response: Harnessing Near Real-Time Satellite and Twitter Signals'*, *ISPRS International Journal of Geo-Information, 4: 4*, pp 2246-2266.

Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014) 'The Parable of Google Flu: Traps in Big Data Analysis', Sciecne, 343: 6176 pp. 1203-1205

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I. and Boyd, D. (2011) 'The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions', *International Journal of Communication* (5) Feature: 1375-1405.

Nagar, R., Yuan, Q., Freifeld, C.C., Santillana, M., Nojima, A., Chunara, R. and Brownstein, J.S (2014) 'A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives', *Journal of Medical Internet Research*, 16:10, e236.

NatCen (2014) *Research Using Social Media: Users' View*, London: Natcen.

ONS (2011) *2011 Census – Population and Household Estimates for England and Wales*. Newport: ONS.

Ripjar (2015) *Nepal Earthquake: A news and social media case study*, Cheltenham: Ripjar. Available at: < https://ripjar.com/2015/06/nepal-earthquake/>

Sloan, L. and Morgan, J. (2015) 'Who Tweets with Their Location? Understanding the Relationship Between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter', *PLOS ONE*.

Sloan, L., Morgan, J., Burnap, P. and Williams, M. L. (2015) 'Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data', *PLOS ONE.*

Sloan, L., Morgan, J., Williams, M. L., Housley, W., Edwards, A., Burnap, P., Rana, O. (2013), 'Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter', *Sociological Research Online* 18:7.

Smith, D. (2012), *How Many People Use the Top Social Media?*. Digital Market Ramblings. Available at: <http://expandedramblings.com/index.php/resource-how-many-people-use-the-top-social-media/>

Weber, J. A. (2012) *Frameworks for Inclusion of the Private Sector in Disaster Preparedness and Response*, London: DFID.

Williams, M. L. (2015) 'The Ethics of Using Social Media Data in Social Research', presented at the *Workshop of Ethics in Online Research*, Social Research Association, UCL.