Identifying offenders on Twitter: a law enforcement practitioner guide

Abstract

Twitter remains one of the most popular social media network sites in use today and continues to attract criticism over the volume of unsavoury and illegal content circulated by its users. When breaches of legislation occur, appropriate officials are left with the task of identifying and apprehending the physical user of an offending account, which is not always a simple task. This article provides a law enforcement practitioner focused analysis of the Twitter platform and associated services for the purposes of offender identification. Using our bespoke message harvesting tool 'Twitterstream', an analysis of the data available via Twitter's Streaming and REST APIs are presented, along with the message metadata which can be gleaned. The process of identifying those behind offending Twitter accounts is discussed in line with available API content and current Twitter data retention policies in order to support law enforcement investigations surrounding this social media platform.

Keywords: Twitter; Crime; Digital Forensics; Offender Identification; Social Media

1 Introduction

Through the commercialisation of the Internet, modern day communication protocols have been revolutionised. Cyberspace provides a place for the instantaneous transfer of data between individuals both known and unknown to each other. Now, we have the concept of social media, defined by Kaplan and Haenlein (2010) as 'a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content'. These provisions are now common place in society, with estimates showing around 2.2 billion social network users worldwide (Statista, 2016e), where multiple variations exist outside of what can arguably be categorised as bigger players (Facebook, Twitter, for example). Yet social media can be considered a relatively new phenomena, where despite some instances of social media networking sites such as SixDegrees.com existing in the late 1990's, it was not until 2003 and the advent of MySpace, followed soon by Facebook in 2004 did this area generate significant interest (Boyd and Ellison, 2007). Now it is hard to imagine a world without these forms of communication as they have become embedded into almost all aspects of life.

Despite the existence of numerous social media networking sites, this article focuses on Twitter. Twitter 'is an information network made up of 140-character messages called Tweets' (Twitter, n.d.a) and as of 2016, has a recorded 310 million active accounts (Statista, 2016a). The functionality of Twitter is relatively simple, after creating their account users can Tweet content which unless protected in their account settings, remains publically available. Users can also 'follow' other user accounts and view their posted content, or be 'followed', whereby other users can passively observe their actions or interact by way of messages or replies (see Kwak, et al., 2010 for an in-depth account of the functionality of Twitter). Although definitive statistics are difficult to obtain, Twitter is widely reported to witness thousands of Tweets (Twitter's coined term for message) per minute which has led to billions of resident messages. Twitter remains one of the most popular social media platforms in use, often a forum for public expression of opinion on current affairs and matters of interest in the public eye, in real time. To provide an example of Twitter usage volume, Barack Obama's 2012 re-election resulted in over 300,000 tweets per minute (Statista, 2016c) with the post halftime show at Super Bowl 2016 attracting over 150,000 tweets per minute (Statista, 2016d).

Inevitably, with such a large populace engaging with the Twitter platform both positive and negative usage characteristics are frequently witnessed. Twitter has frequently attracted criticism for the volume of abusive and malicious content posted. This problem is best described by Scaife (2013) who states, 'due to the anonymity social media affords, users can potentially express unrestrained and harmful content'. Arguments surrounding freedom of speech are often conveyed in defence of posted online content, yet users do not have free reign to express anything they wish, with multiple jurisdictions having sought to legislate on content which is legally acceptable. This has led to regulatory issues surrounding Twitter and the development of strategies for policing such large volumes of traffic, and, detecting and prosecuting those liable for breaches.

Focusing on Twitter, this article first analyses English law discussing the regulations in place for defining acceptable usage. With a need to identify and prosecute those who breach aforementioned legislation, a discussion of Twitter's application programming interfaces (APIs), both Streaming API and REST API is offered and their use for gathering Twitter message content (see Section 3). We also demonstrate our bespoke 'Twitterstream' application; a method for invoking Twitter's APIs for the collection, analysis and presentation, and preservation of Twitter message content. For the purpose of offender identification, Twitter message metadata fields are highlighted and their relevance discussed in an effort to support law enforcement investigations of this type. Finally regulatory issues involving Twitter are discussed and concluding thoughts offered.

2 The Dark Side of Twitter

In 2015, there were over 5,500 data removal requests submitted by government agencies, police and courts worldwide (Statista, 2016b), with Twitter now frequently identified as a platform which is used for online abuse. Applications to Twitter for account personal information by UK organisations doubled in 2015 making it the biggest requester for Twitter data in the European Union (BBC News, 2015). O'Flinn (2013) highlights this issues stating 'Twitter has unleashed offensive behaviour of a type unseen in real life', leading to significant regulatory issues. Jack Dorsey, CEO of Twitter recently acknowledged the harassment and misuse that takes place via the service (BBC News, 2016a), where incidents include racial abuse (BBC News, 2012a; BBC News, 2016b), death threats (BBC News, 2016c), menacing communications (BBC News, 2012b), contempt of court (BBC News, 2011) and libel (BBC News, 2012c; BBC News, 2013a). Trolling (deliberate behaviour designed to provoke a reaction) is now commonly witnessed on Twitter, and the platform could be perceived as largely unregulated due in part to its size, where breaches of legislation of frequently witnessed.

The following provides an overview of the various types of infringement of English law which can occur on Twitter (see also CPS (n.d.b) and Athena Forensics (n.d.)).

Act: Contempt of Court
Notable Instances: Circulation of images of killer Jon Venables (BBC News, 2013b), breached injunction of Ryan Giggs (BBC News, 2011).

What is it?: The Crown Prosecution Service (n.d.a) guidance defines contempt as "an act or omission calculated to interfere with the administration of justice". In *Dallas v United Kingdom (Application No.38395/12)*, specific attention was drawn to jurors on the potential liability for contempt of court where Twitter and other social media platforms to inform additional persons regarding the details of the case. In fact, Agate and Ledward (2014) indicate that it is now common practice for judges to specifically warn jurors not to engage with case-related social media. Providing a notable example, in *HM Attorney General v Liddle [2013] EWHC 1455 (Admin)*, the defendant Dean Liddle was convicted for contempt of court for posting images of convicted killers Venables and Thompson as adults on Twitter in breach of a global injunction preventing the publication of their new identities. The issue of social media usage in courtroom proceedings is not solely confined to the UK, where concerns in jurisdictions such as the US exist (Lee, 2010; Dysart and Kimbrough, 2013; Hoffmeister, 2015).

Act: Credible Threat
Notable Instances: See consideration of what constitutes a ‘credible threat in *Chambers v DPP [2012] EWHC 2157*

What is it?: Crown Prosecution Service (n.d.b) guidance define credible threats as one of the following categories.

1. Offences Against the Person Act 1861 Section 16 - ‘A person who without lawful excuse makes to another a threat, intending that that other would fear it would be carried out, to kill that other or a third person’.
2. Protection from Harassment Act 1997 Section 4 - ‘A person whose course of conduct causes another to fear, on at least two occasions, that violence will be used against him is guilty of an offence if he knows or ought to know that his course of conduct will cause the other so to fear on each of those occasions’.
3. Malicious Communications Act 1988 Section 1 - prohibits the sending of an electronic communication which conveys a threat.

Threatening communications sent via social media platforms are also regulated in a number jurisdictions including Europe and the US (O'Connor, 2013).

*Act: Grossly offensive, indecent or m*enacing electronic communications breaching Communications Act 2003.
Notable Instances: *Chambers v DPP [2012] EWHC 2157*

What is it?: Section 127 of the Communications Act 2003 regulates the improper use of public electronic communications networks:.

> (1)A person is guilty of an offence if he—
>
> (a) sends by means of a public electronic communications network a message or other matter that is grossly offensive or of an indecent, obscene or menacing character; or
>
> (b) causes any such message or matter to be so sent.
>
> (2)A person is guilty of an offence if, for the purpose of causing annoyance, inconvenience or needless anxiety to another, he—
>
> (a) sends by means of a public electronic communications network, a message that he knows to be false,
>
> (b) causes such a message to be sent; or
>
> (c) persistently makes use of a public electronic communications network

The case of *Chambers* (referred to as the Twitter joke trial (Akhtar, 2014)) surrounding the post of the Tweet "Crap! Robin Hood airport is closed. You've got a week and a bit to get your shit together otherwise I'm blowing the airport sky high." by Paul Chambers. Chambers was charged under Section 127(1) Communications Act 2003, which was later quashed on appeal. In addition, Newcastle University student Joshua Cryer was charged under Section 127 Communications Act 2003 for sending racist messages for former professional footballer Stan Collymore (BBC News, 2012d). See also offences under the Public Order Act 1986 and the case of Liam Stacey (albeit charged under Section 4A Public order Act 1986), who 'received a 56-day jail term after tweeting 'LOL' ["laugh out loud"] in response to the on-pitch collapse of the footballer Fabrice Muamba and subsequently posting racist and offensive comments when other users criticised him' (Scaife, 2013, pg.5). Further those who utilise Twitter to incite racial hatred via Twitter may be liable under the Public Order Act 1986 (see case of Michael Doyle who was alleged to have incited racial hatred after the Brussels attacks (BBC News, 2016d)).

As a separate instance, those utilising Twitter accounts linked with terrorist organisations or supporting such content may also face liability (Spencer, 2017). Spencer (2017, p503) indicates that in the US, suspected supporters 'of ISIS who have been arrested have been charged is 18 U.S.C. § 2339B (2012) "which punishes "[w]hoever knowingly or attempts or conspires to do so"'.

Defamation (an act which damages the reputation of another) cases stemming from Twitter are also a concern with both professional cricketer Chris Cairns (*Cairns v Modi [2012] EWHC 756 (QB*)) and politician Alistair Mcalpine (*McAlpine v Bercow [2013] EWHC 1342 (QB)*) successfully claiming damages for Tweets made. The impact of Twitter usage can also be witnessed in professional sport where the English Football Association (FA) have reprimanded numerous professional football players including Mario Balotelli and Rio Ferdinand for unacceptable use of the Twitter platform (Coe, 2015; Carpenter and Pendlebury, 2015). Defamation is a legal concept acknowledged beyond the UK (see case involving Courtney Love - *Gordon & Holmes v. Love, No. B256367 (Cal. Ct. App. Feb. 1, 2016)*), however Allen (2014) suggest few cases in this area in relation to Twitter have made it to trial in the United States.

Regardless of the offence committed on Twitter, in order to effectively prosecute, the individual responsible for the act must be identified. To do this, information regarding an offending tweet must be gathered.

3 Gathering Tweets and their Metadata

The gathering of social media content during an applicable criminal investigation is an act often carried out by law enforcement. Brannan (2013) indicates that during the England Riots in 2011, Twitter was actively monitored by law enforcement, and as a public network, without breaching Article 8 of the European Convention on Human Rights as "where someone does an act in public, the observance and recording of that act will ordinarily not give rise to an expectation of privacy" (Gillespie, 2009, pg.555).

As Twitter is a real-time messaging service, the content resident on this platform is volatile, and in cases where a breach is established, information gathering processes must act quickly. Where offending messages have been posted, once identified, it is reasonable to assume that any offender may seek to delete any offending content and/or deactivate their account. Twitter user policies state that deactivation places the account in queue for deletion, where user data relating to the account is only kept for 30 days (a reactivation period), where content may still be available on Twitter up to a few days after deactivation (Twitter, 2016a), which once lapsed deletion takes place in approximately a further week. Content which has been deleted by a Twitter user is generally not available on Twitter servers (Twitter, 2016b). In order to prevent offender data from being lost, ‘preservation requests’ are accepted by Twitter, where data is stored for 90 days (a request to extend this initial period can also be made) and not disclosed until the necessary authority is acquired (Twitter, 2016b). There are two main issues here, time and cooperation. Legal processes are notoriously slow, where delays could result in retained date no longer being available, as noted by Labour MP Lilian Greenwood (HC Deb (2014) 588 col. 14). Further, Twitter is a U.S. based organisation where cooperation with foreign jurisdictions and legal practices is not guaranteed (see sections 4 & 5 for further discussion).

As a result reliance cannot be solely placed upon Twitter by law enforcement alone to acquire and retain all metadata regarding any offending users. As Twitter provides access to both the Streaming and REST APIs, section 3 will provide an analysis of information gleaned from Twitter using these APIs via our Twitterstream platform, which is designed to harvest, store and analyse publicly available Twitter hosted content. We will then evaluate the content of this data in section 4 for the purpose of identifying physical account users.

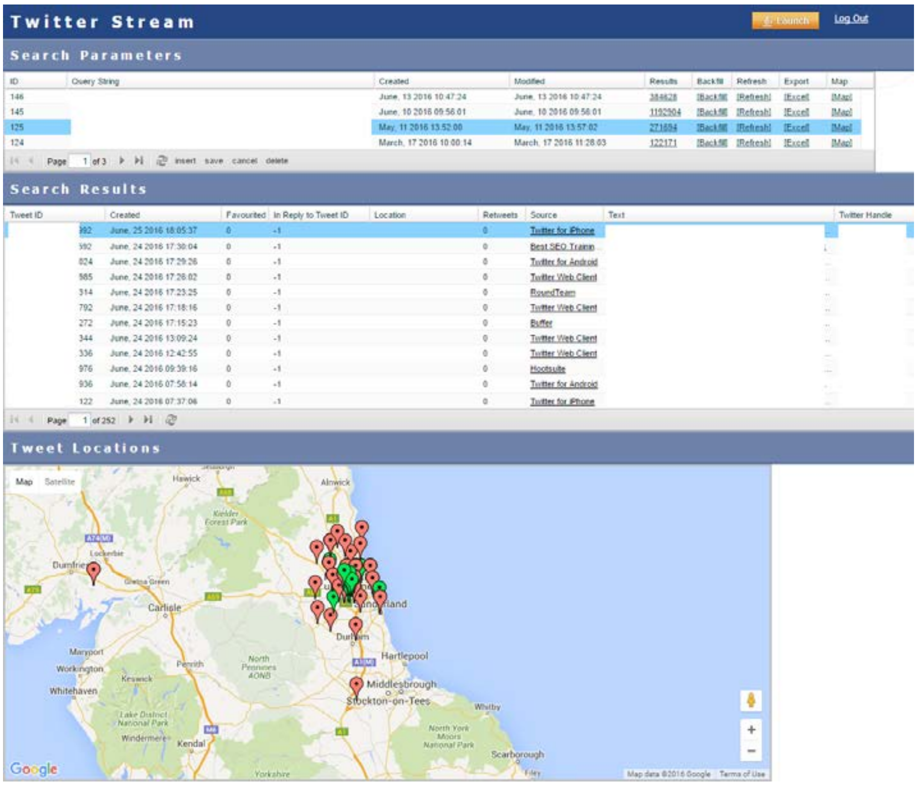
3.1 Twitterstream

Twitter's APIs are frequently utilised in the gathering of message content for research and analysis. Research includes the gathering of message data to establish political consensus surrounding the United Kingdom's exit from the European Union ((Llewellyn and Cram (2016)), malicious URL detection (Venkatesh et al., 2017), stock price prediction (Kordonis et al., 2016) event detection (Li et al., 2012; Alsaedi et al., 2017), cyber-threat detection (Mittal et al., 2016) and the fallout from natural disasters (Gupta et al, 2013; Murthy et al., 2016). The Twitterstream application is designed to capture publically available data from Twitter in both real time and retrospectively via their two APIs (discussed below), storing tweets and their associated metadata in an attached database instance.

1. *Streaming API*: The Streaming API provides access to globally available Twitter data in real time. In addition to the text of the tweet and any attached imagery and hyperlinks, the API also supplies the full set of metadata associated with the tweet, including information on the account, its holder and location, and system or platform used to send the tweet (see appendix 1 for Twitter’s comprehensive metadata field breakdown).

2. *REST API*: The REST API allows the recovery of tweets for a period of time after their publication. Our tests showed that tweets can only be recovered if they are no more than approximately nine days old (although this timeframe is variable and dependent on the volume of traffic related to a particular handle or hashtag), backed by Twitter’s discussion of the API’s functionality (Twitter, n.d.b). A caveat of the REST API is that message capture is selective where Twitter states it is not a comprehensive recovery of data with focus maintained on the relevance of the message, not a complete recall of data, meaning some historical messages may not be captured, despite being within the nine data time frame (Twitter, n.d.b.).

The Twitterstream interface (shown in Figure 1) supports user interaction with any captured Twitter data whilst allowing the submission of queries and the review and analysis of results. Data which is captured through Twitterstream is stored in an SQL Server database. The interface includes a Google Maps overlay, which displays the location of tweets subject to location data being available.



Twitter Stream

Launch Log Out

Search Parameters

ID	Query String	Created	Modified	Results	Backfill	Refresh	Export	Map
146		June, 13 2016 10:47:24	June, 13 2016 10:47:24	384628	[Backfill]	[Refresh]	[Excel]	[Map]
145		June, 10 2016 09:56:01	June, 10 2016 09:56:01	1192904	[Backfill]	[Refresh]	[Excel]	[Map]
125		May, 11 2016 13:52:00	May, 11 2016 13:57:02	271694	[Backfill]	[Refresh]	[Excel]	[Map]
124		March, 17 2016 10:00:14	March, 17 2016 11:28:03	122171	[Backfill]	[Refresh]	[Excel]	[Map]

Page 1 of 3 insert save cancel delete

Search Results

Tweet ID	Created	Favourited	In Reply to Tweet ID	Location	Retweets	Source	Text	Twitter Handle
992	June, 25 2016 18:05:37	0	-1		0	Twitter for iPhone		
592	June, 24 2016 17:30:04	0	-1		0	Best SEO Trainin		
024	June, 24 2016 17:29:26	0	-1		0	Twitter for Android		
585	June, 24 2016 17:28:02	0	-1		0	Twitter Web Client		
314	June, 24 2016 17:23:25	0	-1		0	RoundTeam		
792	June, 24 2016 17:18:16	0	-1		0	Twitter Web Client		
272	June, 24 2016 17:15:23	0	-1		0	Buffer		
344	June, 24 2016 13:09:24	0	-1		0	Twitter Web Client		
336	June, 24 2016 12:42:55	0	-1		0	Twitter Web Client		
976	June, 24 2016 09:39:16	0	-1		0	Hootsuite		
936	June, 24 2016 07:58:14	0	-1		0	Twitter for Android		
122	June, 24 2016 07:37:06	0	-1		0	Twitter for iPhone		

Page 1 of 252

Tweet Locations

Figure 1: The Twitterstream web interface

3.2 Twitterstream components

Twitterstream has three main structural components with each discussed in turn.

1. The engine of the system is a Java application, which runs on a virtual server and connects to Twitter's Streaming API via the Twitter4J library (Twitter4J, n.d.). An overview of this process is presented in Figure 2.

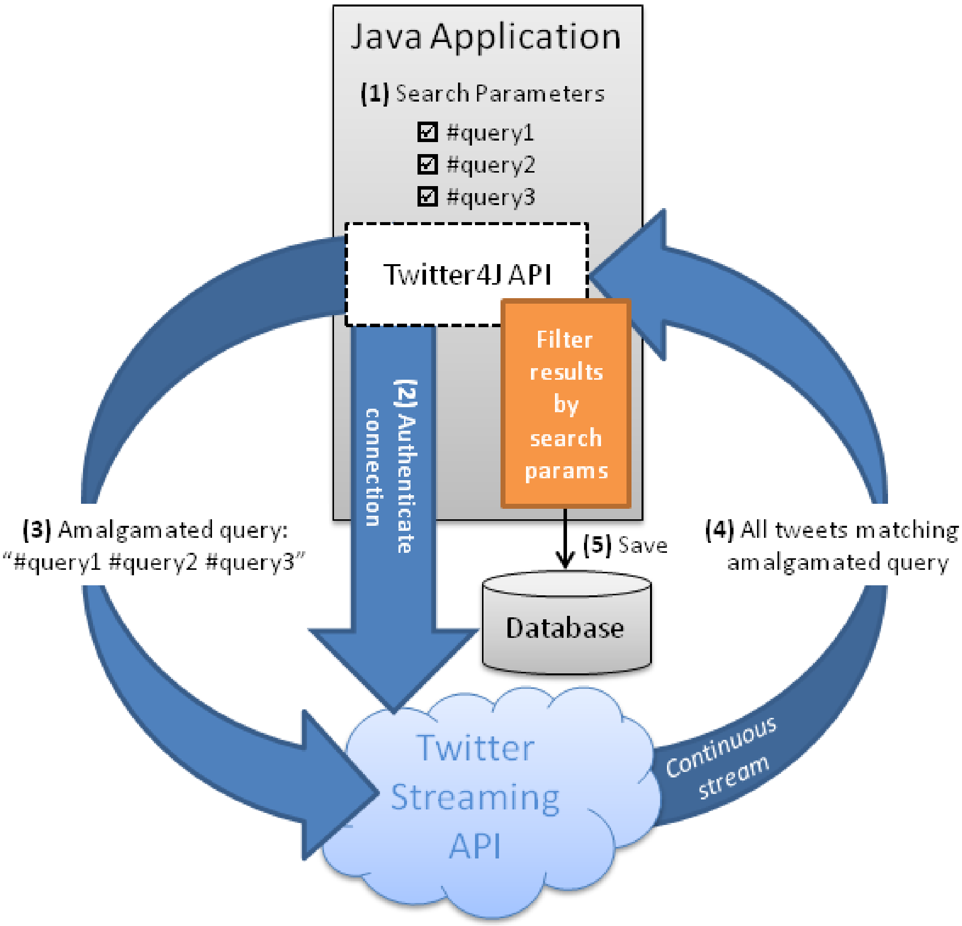

Java Application
(1) Search Parameters
☑ #query1
☑ #query2
☑ #query3
Twitter4J API
Filter results by search params
(2) Authenticate connection
(5) Save
Database
(3) Amalgamated query: "#query1 #query2 #query3"
(4) All tweets matching amalgamated query
Twitter Streaming API
Continuous stream

Figure 2: Java application overview

User defined search parameters are first stored in the database as strings (1). When a Twitter stream is initiated, the user can chose their appropriate Boolean search values (AND, OR etc.) which is automatically embedded into the search string. Search string criteria can be profile handles, hashtags, text content and also metadata relating to the tweet such as geolocation information. The application uses application-user authentication (Twitter, n.d.c) to connect to the Streaming API (2). Credentials linked to a specific Twitter account are passed via an OAuth request (see http://oauth.net/). As Twitter does not allow several connections to its Streaming API with the same credentials, individual search parameters are merged into a single, query string (3), this way, only a single connection is required, improving the efficiency of the search process. A connection to the Streaming API involves making a long-lived HTTP request followed by parsing response data in turn (Twitter, n.d.d). Tweets containing any of the words in the amalgamated query string are periodically returned to the application (4) and filtered to ensure that each individual tweets is correctly linked to its source search parameter (5).

Twitterstream is hosted on a cloud server with scalable resources allowing additional processing power and memory to be configured at runtime if required.

2. The data storage component is provided by a Microsoft SQL Server instance, running on its own dedicated virtual server with scalable resources. This form of local

storage means that once harvested, message data can be stored in a non-volatile state, where further querying and analysis can take place.

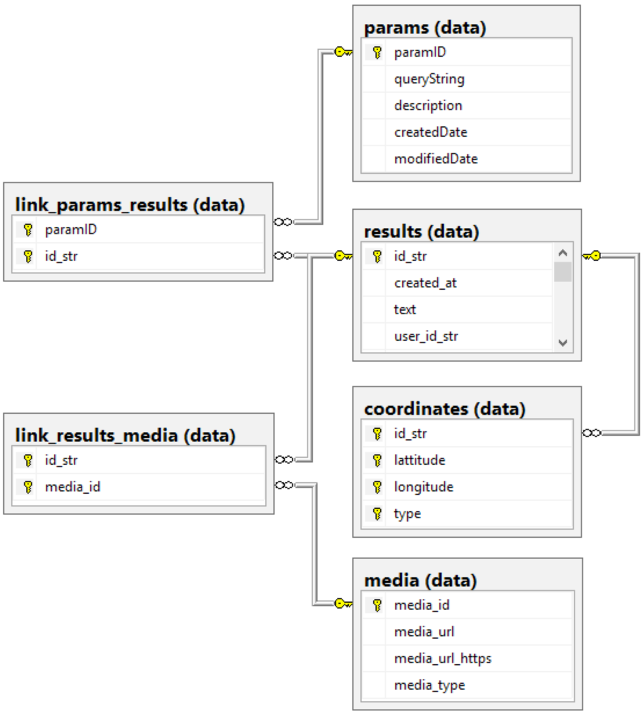


Figure 3: Database structure

To provide an overview of Figure 3, the link_params_results table exists because a single result (i.e. tweet) could be linked to more than one search parameter. This is the most efficient way of storing tweets, as each tweet only exists once in the database. Similarly, a single result can have multiple coordinates. Again, it's more efficient to store them in a separate table.

3. The third component is a web interface which is written in Adobe ColdFusion. The web interface allows users to create new queries, view and sort result sets, export result sets to Microsoft Excel, and view tweets with geolocation data on a Google Map (see Figure 4). The web interface also allows users to backfill result sets retrospectively, using the monkehTweet ColdFusion wrapper (RIAForge, n.d.) which connects to Twitter's REST API.

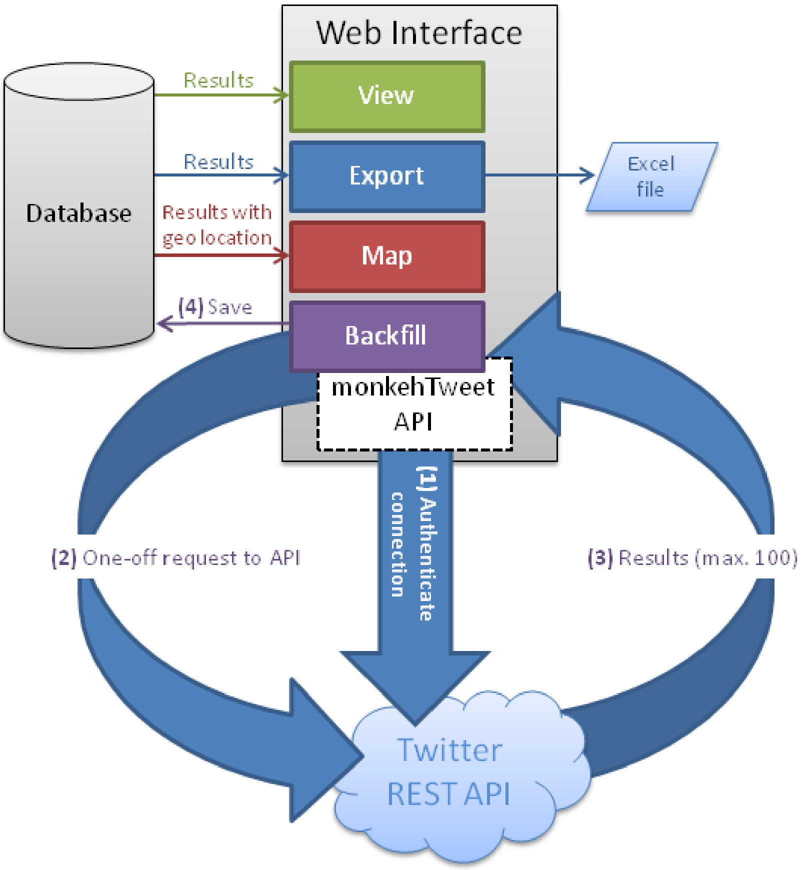


Figure 4: Web Interface and REST API

In relation to the backfill process using REST API, a connection is first established with Twitter via an OAuth request, the same as with the Java application (1). Next, a one off HTTP request is sent to the REST API (essentially containing the search query ('q'), count ('100') and result_type ('recent'). (See https://dev.twitter.com/rest/reference/get/search/tweets) (2). A max of 100 results (matching tweets) at one time are returned to the ColdFusion application (3). Once received, the results are saved in the Twitterstream database for later analysis.

3.2.1 Limitations of the API Usage

Twitter's APIs can only be used to extract information accounts deemed 'active'. The problem this presents is that there remains ambiguity surrounding the actions that are required to deem an account active. Twitter accounts which are first registered are not active, where guidance (Twitter, 2016c) indicates activity to be based on actions such as the volume of tweets, re-tweeted messages mentions (profile handle included in other messages). Therefore offenders who create new accounts solely for the purpose of offending may not have the accounts deemed active quick enough to have their content included in the API data.

A further limitation stems from those who opt to protect their messages. Protected accounts cannot have their data passively collected using Twitterstream and Twitter's APIs. However, law enforcement may still be able to access a user's protected messages by obtaining the relevant legal documentation and following Twitter's formal procedures for requesting a disclosure of a user's account information (Horsman, 2017; Twitter, 2016b).

3.2.2. Twitterstream Summary

The Twitterstream framework provides a dynamic method for collecting, storing and analysing Twitter message content via Twitter's public APIs. Given that Twitterstream allows message metadata collection both by targeted (i.e. from a particular account handle) and large scale (such as a hashtag collection) methods, significant volumes of information may be collected, which is in need of subsequent storage and analysis. Twitterstream provides a method for the flexible and effective allocation of resources for collecting message content, and robust database storage for housing Twitter data for subsequent querying and analysis. Twitterstream's bespoke interface also provides the user with SQL-based filtering options and a visual geographical mapping facility. Although the Twitter's APIs are available for public use, Twitterstream is currently not available to DF practitioners as its resources and infrastructure are managed internally. The authors are currently developing the portability of Twitterstream to allow practitioners to implement and utilise the framework in Twitter investigations. Expanding Twitterstream's availability to others will also allow additional testing and validation of the framework to take place.

3.3. Message Metadata

The Twitterstream platform allows the value of data gleaned from Twitter's APIs for the purposes of law enforcement and offender detection to be analysed. Such data is publically available and may support an investigation without the need to engage in legal processes with the Twitter organisation over data retention and disclosure. In turn, it may in some instances provide law enforcement with the necessary information to target specific retained information from Twitter via a submitted request. This is particularly due to the type and availability of metadata associated with each message gathered through the APIs.

When acquiring Tweets from Twitter stream, almost 60 distinct metadata fields are associated with each message (for a full list, see Appendix 1). However, not all of these fields are of potential value to law enforcement during an investigation. To distil this content, potentially evidential metadata fields are now highlighted having been verified using a test Twitter account and placed into three categories, metadata surrounding message content, location and profile information.

3.3.1 Message Contents Fields

The following are metadata fields gathered via Twitterstream from Twitter's APIs surrounding a potential offending message's content.

Created_at: A timestamp field denoting the created time of the message.

Text: A string field containing the contents of the message.

Favorite_count: An integer field indicating approximately how many times this Tweet has been 'liked' by Twitter users.

Id_str: This field is the unique identifier of a message. This information can be used to potentially tie a message to a user where only a retweet of an offending message has been found (see appendix 1 for an explanation of the linked fields *in_reply_to_status_id_str* and *in_reply_to_user_id_str*).

3.3.2 Location Information

The following fields gathered via Twitterstream provide metadata surrounding a potential location of an offender.

User_geo_enabled: A boolean field, when true, indicates that the user has enabled the possibility of geotagging their Tweets.

Geo_lattitude and *geo_longitude.* A decimal field denoting location information where location services have been activated on the message sender's phone.

user_location. A user defined string for the profile location. This may not be accurate and is considered as secondary location information.

3.3.3 User Identification Information

The following fields gathered via Twitterstream provide metadata surrounding the potential identification of an offender posting a message.

User_id_str: A string representation of the unique identifier for this User.

User_created_at: A timestamp field denoting the UTC datetime that the user account was created on Twitter.

User_description: The user-defined string describing their account.

User_name: A string field showing the name of the user, as they've defined it. Not necessarily a person's name. Typically capped at 20 characters.

User_screen_name: A string field showing the screen name, handle, or alias that this user identifies themselves with, usually preceded with '@'. Screen names are unique but subject to change.

3.3.4 Caveat: Metadata Manipulation

The metadata available via the Twitter APIs cannot in all circumstances be authenticated, with particular attention drawn to the location information. Geolocation data is only available on messages where the user has enabled it both on their mobile device and within the Twitter application itself. Kumar et al., (2013) state that only approximately 1% of all tweets on Twitter are maintain geolocation information. By default, location data is not initialised, therefore arguably in many circumstances location data will not be available. Even where location data is initiated by the user, the following three outcomes are possible (having been verified through testing using a test Twitter account with messages collected by Twitterstream).

1. The user may choose to initialise the option to use their exact location. Here, the *Geo_lattitude* and *geo_longitude* fields should provide accurate co-ordinates of the poster's position (subject to point 2).

2. It is possible to spoof the geolocation data from the handset. Tests showed that utilising the application ‘Fake GPS Location’ (Google Play, 2016), coordinate data can be effectively spoofed. In this case, the *Geo_lattitude* and *geo_longitude* fields supplied to Twitterstream via the APIs cannot be relied upon. There is no way of verifying (from the data gathered from the Twitter APIs) whether spoofing has taken place, with an analysis of the device used to send the message needed. As this would only be available once a suspect has been apprehended, the purpose of validating geo-location data is defeated, and therefore location data must be treated with caution.

3. Finally, users have the ability to search for a location or venue within the Twitter interface, meaning that users can be based in one location but tag the message with a different one (for example, post from Manchester and tag the tweet as being in Liverpool). Tests showed that data gathered from the Twitter API showed coordinate data reflected the location chosen by the user, not the actual location. As with point 2 above, in general, the *Geo_lattitude* and *geo_longitude* fields cannot be relied upon in all instances.

3.3.5 Data Gathering on Twitter: Examples

To provide context to the capabilities of Twitterstream and the volume and type of data provided through the Twitter Stream and REST API’s the following examples are presented.

3.3.6 Streaming API: Sunderland vs Everton Premier League football game

Using the Streaming API, the following is a breakdown of extracted data gathered from Twitterstream when semantically focused on the Sunderland vs Everton Premier League football game taking place on the 11th May 2016 (see Table 1).

The following search parameters were defined:

Table 1. Tweets captured from Sunderland vs Everton Premier League football game	
Search Terms (generated from local dialect and terms associated to the game)	safc; nufc; ftm; mackem; geordie; safcefc; efc; smb; sunvefc; sunefc; everton; mags; haway; #sol
Stream started	11th May 2016 12:50:27
Stream finished	12th May 2016 16:10:21

	No. of Tweets	Total %	Other Info
Volume of Tweets	271,694	100.00%	N/A

Tweets with exact location	387	0.14%	I.e. geo_lattitude + geo_longitude
Tweets with bounding box	8666	3.19%	I.e. either a rectangular bounding box (4 sets of coordinates) or a point of interest (1 set of coordinates)
Tweets from Tyne and Wear	1392	0.51%	42 exact location; 23 points of interest; 1327 bounding box
Tweets from Sunderland	438	0.16%	28 exact location; 22 points of interest; 388 bounding box

3.3.7 Streaming API: European Championships Opening Ceremony

Using the Streaming API, the following is a breakdown of extracted data gathered from Twitterstream when semantically focused on the start of the European Championships 2016 on 10th June 2016 (see Table 2).

Table 2. Tweets captured from European Championships Opening Ceremony	
Search Terms (generated from local dialect and terms associated to the game)	#euro2016 #euro16 #frarom #frarou
Stream started	10th June 2016 @ 09:58
Stream finished	11th June 2016 @ 04:06

	No. of Tweets	**Total %**	**Other Info**
Volume of Tweets	1,192,904	100.00%	N/A
Tweets with exact location	387	0.22%	I.e. geo_lattitude + geo_longitude
Tweets with bounding box	8666	2.64%	I.e. either a rectangular bounding box (4 sets of coordinates) or a point of interest (1 set of coordinates)
Tweets from Paris	1392	0.09%	282 exact location; 0 points of interest; 816 bounding box

3.3.8 Streaming & REST API: Orlando Pulse Shooting

Using the Streaming and REST APIs, the following is a breakdown of extracted data gathered from Twitterstream when semantically focused on the Orlando Pulse Shooting June 12th 2016 (see Table 3).

Table 3. Tweets captured from Orlando Pulse Shooting	
Search Terms (generated from local dialect and terms associated to the game)	#LoveWins #pulseshooting #praysForOrlando #Loveislove #PrayForOrlando #orlando #pulse #pulsenightclub
Stream started	12/06/2016 @ 14:11:20
Stream finished	13/06/2016 @ 09:51:33

	No. of Tweets	**Total %**	**Other Info**
Volume of Tweets	384,628	100.00%	N/A
Tweets with exact location	0	0.00%	I.e. geo_lattitude + geo_longitude
Tweets with bounding box	8845	2.30%	I.e. either a rectangular bounding box (4 sets of coordinates) or a point of interest (1 set of coordinates)
Tweets from Orlando	226	0.06%	0 exact location; 21 points of interest; 205 bounding box

3.3.9 API Considerations

Al-garadi et al., (2016) utilised Twitter's APIs for the gathering of publically available message data to develop cyberbullying detection methods and acknowledge the limitations of these functions. It must be noted that Twitter does not endorse mass surveillance via its API (Twitter, 2016f) and nor does it facilitate it due to incomplete data sets provided. Depending on implementation, the APIs do not provide a complete extraction of data. For example, if an API user wanted an extraction of all messages containing '#security', the APIs only provide a subset of available data. At the time of writing, quantification on the size of the subset in terms of what volume of these messages would be collected is not possible, but has been noted by Al-garadi et al., (2016) and others (Cheng and Wicks, 2014; Gonzalez-Bailon et al., 2014) to potentially create sample bias due to incomplete downloads. The ability for mass collection is limited when using the freely available APIs. However, a user can pay for complete access to 100% of Twitter messages via the APIs or can be acquired via 3rd party data resellers (Kumar et al., 2013).

However, the REST API can be effectively utilised for targeted extraction of messages and their associated metadata, allowing the last 3200 messages from a Twitter account to be extracted searching for a specific Twitter handle (Twitter, 2017). This would allow the reporting of a specific tweet to law enforcement and then providing the user account associated with the message has not deleted the content, associated message metadata can be extracted. During tests carried out using a research Twitter account, the REST API was able to target and recover posted messages by utilising the account's Twitter handle.

Where a sample of tweets are sought (for example, from a specific hashtag) Janetzko (2017) indicates that the REST API focuses on relevance as criteria for message recovery, rather than completeness therefore not all messages are indexed and potentially retrievable. As a result, the REST API suffers from the same issues as the Streaming API where complete data is not available without cost and historical messages may not be recoverable (Janetzko, 2017). However, it is assumed that in instances of malicious tweets, appropriately quick reporting would make the use of the Twitterstream application a viable option. In addition, restrictions in the amount of messages which can be downloaded from the REST API is also an issue limiting historical data collection (Sheela, 2016).

4 Putting it all Together: Evaluating API Data for Offender Identification

Where illicit acts are noted, the global reach of Twitter now means that the regulation of users and their actions is not straightforward, often involving multiple jurisdictions and large geographical areas. Even when breaches of law can be established, identifying an offender in order to instigate prosecution remains difficult and in some cases impossible. The fundamental problem posed by Twitter in regards to the aforementioned offences in section 2 is the ability to attribute an offending message to a physically identifiable user. There is no definitive method for identifying a physical individual who sent a message on Twitter, only a series of data which when collated may lead to this outcome. Ledward and Agate (2014) best summarise the issue as follows:

> "While pressure from UK Government grows on Facebook, Twitter and other social media providers to enforce their terms of use in relation to unacceptable/criminal content on social media, and to respond more quickly and more effectively to reports of abuse, the tension with users' art.10 and art.8 rights continues, and obtaining identity information or content in evidential form remains problematic and time-consuming in this jurisdiction".

As discussed in section 3, the metadata fields which can be gathered using Twitterstream from Twitter's APIs often cannot be used alone to attribute a message to an individual.

Although message content, visible profile information or fields such as *user_profile_image_url*, may disclose personally identifiable information in some instances, this is not always the case. Many individuals relinquish the opportunity to apply personally identifiable information to their Twitter profile, particularly where they intend to posting offending content (Birbeck, 2013). In turn, locational data (providing it is accurate) may place someone in a general area or in a best case scenario, a place of residence, but further investigation of devices is required in order to identify an individual sender.

Where API metadata alone cannot be used to directly identify an offender, the next step is to use this information to attribute a message to a specific account. However, support from Twitter may be limited, where Colin Crowell, Twitter's head of public policy indicated that most Twitter data retained is already available through the APIs (Joint Committee on the Draft Communications Data Bill, 2013).

Where additional account information is needed from Twitter, reliance must be placed upon Twitter and their maintenance of user logs for account holders. Using API metadata such as the *Id_str* and *User_Id_str* it may be feasible for Twitter to identify an active offending account. At which point, account information stored by Twitter may support the identification of the physical account holder, but again, this is not guaranteed in all circumstances. As of the time of writing, those who seek a Twitter account have to supply 3 criteria, a full name, valid email address and password. Yet, Twitter states that it 'doesn't require real name use, email verification, or identity authentication' (Twitter, 2016b) and as a result the account details may not be accurate or attributable to a physical user.

A lack of validation means that any name can be entered leading to varying levels of reliability, depending on whether an offender has submitted correct information and often dependant on where the user intends to use their account legitimately of whether it is for the purpose of abuse. The need for a valid email address does provide some support in the offender identification process. Services such as Gmail and Microsoft Outlook may implement various authenticity checks which can involve the use of SMS messages to validate the account. What this means is that when a Twitter user creates an account they maybe doing so with an email address which has been validated with personal details linking themselves to the account. At which point, the email associated with a Twitter account can be queried with the provider (for example, Google mail) records for any identifiable information. However, this would be an onerous process and not feasible to implement in all cases. In addition, the requirement for email account verification varies where in some instances, it is not required during the email account creation process. Ultimately, the current account sign up processes mean it is possible for Twitter users to create and send content without leaving behind an identifiable trail of evidence.

The final point to consider in the Twitter offender tracking process is the acquisition of Internet Protocol (IP) information related to a message to an offending message. It is key to note, sender IP information is not supplied via Twitter's APIs. Twitter's policy guidance states that IP information is only maintained for a 'short period of time' (Twitter, 2016b) leaving ambiguity as to whether the necessary authority can be sought in time in order to utilise this information during an investigation. In addition, IP address information is also subject to spoofing and does not automatically guarantee offender identification even when present.

5 Concluding Thoughts

Given the continued popularity of Twitter, it is likely that law enforcement will continue to witness acts of illegitimate use. Although publically available data may in some instances support the identification of an offender during an investigation, it is likely that support from Twitter's is needed in order to disclose account specific and IP information, despite the limitations on this data in place. When distilled, the problem of detecting offending users maintains numerous issues. Time is clearly an inhibiting factor where Twitter's policies note (as discussed in section 3) that limitations of data retention are apparent. When coupled with time consuming legal processes (although 'emergency disclosure requests' can be made if the request regards the danger of death or serious physical injury to a person (Twitter, 2016b)), even where data capable of identifying an offender exists, a lack of timely procedures to access this content may thwart law enforcement. At present, international co-operation between police and organisations like Twitter are limited. This problem is exacerbated by a potential lack of cooperation by organisations like Twitter. For the period July-December 2012, Twitter complied with information requests by UK police in only 4% of cases (Twitter, 2016d). Figures for July - December 2015 show this figure now stands at 76% (Twitter, 2016e). Although there is a clear improvement in response rates, it still leaves what could be considered by law enforcement, an unsatisfactory gap. Twitter maintains a strong stance on upholding what it considers freedom of speech, noting the following statement.

> "Twitter continues to defend and respect the voices of our users including their right to engage in free expression anonymously or pseudonymously. For example, in June 2015, Twitter received a U.S. non-government legal request seeking to unmask several anonymous users. Twitter pushed back on the request based on First Amendment grounds, specifically the right to speak anonymously" (Twitter, 2016e).

Twitter also maintains the right to refuse information requests which fail to identify a specific account or message, of if the request is overly broad in nature (Twitter, 2016e). Further, when requests are made, the account holder is notified allowing them to mount a challenge or in turn, Twitter may challenge the request themselves (Twitter, 2016e). It is difficult to propose or envisage improvements to the process of identifying offending Twitter users as Twitter itself omits to retain the type and volume of information needed to regularly identify these individuals. The presented Twitterstream platform constitutes an effective mechanism for gathering Twitter data and storing it statically for later analysis and use. However, the identification of individual twitter account owners will likely continue to depend on the cooperation of Twitter due to limitations in publically available data via the APIs. As a result, law enforcement may have to settle for the fact that in some cases, the necessary information needed to pursue prosecution is simply just not available, or time constraints may inhibit access to it.

5.1 Future Work

Future work includes the needs to effectively quantify the limitations in data retrieved from using both of the APIs through Twitterstream. This would allow the precision and recall of the APIs to be established and therefore help law enforcement to make an effective decision as to when to utilise it during an investigation. In addition, coupling the platform with existing techniques for behavioural and sentiment analysis would support the use of Twitterstream in investigations of crime occurring during specific events, congregates via specific hashtag usage.

References:

Agate, Jennifer & Ledward, Joycelyn (2014) 'Contempt and social media: update' Entertainment Law, 25(2), 52-54

Akhtar, Zia (2014) 'Malicious communications, media platforms and legal sanctions' *Computer and Telecommunications Law Review.* 20(6), 179-187 at 180

Al-garadi, M.A., Varathan, K.D. and Ravana, S.D., 2016. Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network. Computers in Human Behavior, 63, pp.433-443.

Allen, A.A., 2014. Twibel Retweeted: Twitter Libel and the Single Publication Rule. J. High Tech. L., 15, p.63.

Alsaedi, N., Burnap, P. and Rana, O., 2017. Can We Predict a Riot? Disruptive Event Detection Using Twitter. ACM Transactions on Internet Technology (TOIT), 17(2), p.18.

Athena Forensics (n.d.) 'Think before you Tweet - Twitter and the Law' Available at: http://www.athenaforensics.co.uk/think-before-you-tweet-twitter-and-the-law#.V1bJmfkrKUl (Accessed 1 June 2016)

BBC News (2011) 'Injunctions doubt as footballer Ryan Giggs named by MP' Available at: http://www.bbc.co.uk/news/uk-13516941 (Accessed 1 June 2016)

BBC News (2012a) 'Man sentenced over racist tweets about Newcastle United' Available at: http://www.bbc.co.uk/news/uk-england-tyne-17183384 (Accessed 1 June 2016)

BBC News (2012b) 'Robin Hood Airport tweet bomb joke man wins case' Available at: http://www.bbc.co.uk/news/uk-england-19009344 (Accessed 1 June 2016)

BBC News (2012c) 'Ex-cricketer Chris Cairns wins £90,000 libel damages' Available at: http://www.bbc.co.uk/news/uk-17512027 (Accessed 1 June 2016)

BBC News (2012d) 'Stan Collymore Twitter race abuser Joshua Cryer sentenced' Available at: http://www.bbc.co.uk/news/uk-england-tyne-17462619 (Accessed 1 June 2016)

BBC News (2013a) ‘Alan Davies pays Lord McAlpine damages over tweet’ Available at: http://www.bbc.co.uk/news/uk-24654289 (Accessed 1 June 2016)

BBC News (2013b) ‘Twitter users: A guide to the law’ Available at: http://www.bbc.co.uk/news/magazine-20782257 (Accessed 1 June 2016)

BBC News (2015) ‘Twitter 'snooping' requests double in UK’ Available at: http://www.bbc.co.uk/news/technology-33882688 (Accessed 1 June 2016)

BBC News (2016a) ‘Jack Dorsey: Twitter will take time to fix’ Available at: http://www.bbc.co.uk/news/technology-36376037 (Accessed 26 May 2016)

BBC News (2016b) ‘Azealia Banks' Twitter account suspended after Zayn Malik abuse’ Available at: http://www.bbc.co.uk/newsbeat/article/36276783/azealia-banks-twitter-account-suspended-after-zayn-malik-abuse (Accessed 1 June 2016)

BBC News (2016c) ‘James May slams Twitter 'threats' to Sue Perkins‘ Available at: http://www.bbc.co.uk/news/entertainment-arts-32331218 (Accessed 1 June 2016)

BBC News (2016d) ‘Man arrested after 'mealy mouthed' Brussels tweet’ Available at: ‘http://www.bbc.co.uk/news/uk-england-london-35888748 (Accessed 1 June 2016)

Birbeck, Sarah (2013) ‘Can the use of social media be regulated?’ Computer and Telecommunications Law Review, 19(3), 83-85

Boyd, Danah, m. & Ellison, N.B., 2007. Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, *13*(1), pp.210-230.

Brannan, Jade (2013) ‘Crime and social networking sites’ Juridical Review, 1, 41-51

Carpenter, Kevin & Pendlebury, Adam (2015) ‘Tweeting the game into disrepute: regulation of social media by governing bodies - lessons from English football’ International Sports Law Review, 1, 3-10

Cheng, T., & Wicks, T. (2014). Event detection using Twitter: a spatio-temporal approach. PLoS One, 9(6), e97807.

Coe, Peter, (2015) ‘Footballers and social media "faux pas": the Football Association's "cash cow"?’ Entertainment Law Review, 26(3), 75-78

Crown Prosecution Service (n.d.a) ‘Contempt of Court and Reporting Restrictions’ Available at: http://www.cps.gov.uk/legal/a_to_c/contempt_of_court/ (Accessed 1 June 2016)

Crown Prosecution Service (n.d.b) ‘Guidelines on prosecuting cases involving communications sent via social media’ Available at: http://www.cps.gov.uk/legal/a_to_c/communications_sent_via_social_media/ (Accessed 1 June 2016)

Dysart, Katie L.; Kimbrough, Camalla M. (2013) Justice - Social Media's Impact on the U.S. Jury System, Trial Evidence, Vol. 21, Issue 2 (Summer 2013), pp. 12-14

Gillespie, Alisdair A., "Regulation of Internet Surveillance" (2009) 4 E.H.R.L.R. 552, 555.

Google Play (2016) ‘Fake GPS Location’ Available at: https://play.google.com/store/apps/details?id=com.lexa.fakegps&hl=en_GB (Accessed 1 June 2016)

Gonz alez-Bailon, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. Social Networks, 38, 16e27.

Gupta, A., Lamba, H., Kumaraguru, P. and Joshi, A., 2013, May. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In Proceedings of the 22nd international conference on World Wide Web (pp. 729-736). ACM.

Hoffmeister, T. (2015). Preventing juror misconduct in digital world. Chicago-Kent Law Review 90(3), 981-1000.

Horsman, G., 2017. A survey of current social network and online communication provision policies to support law enforcement identify offenders. Digital Investigation, 21, pp.65-75.

Janetzko, D., 2017. The Role of APIs in Data Sampling from Social Media. The SAGE Handbook of Social Media Research Methods, p.146.

Joint Committee on the Draft Communications Data Bill, (2013) ‘Draft Communications Data Bil’ Available at: https://www.parliament.uk/documents/joint-committees/communications-data/Oral-Evidence-Volume.pdf (Accessed 1 June 2016)

Kaplan, A.M. and Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), pp.59-68.

Kordonis, J., Symeonidis, S. and Arampatzis, A., 2016, November. Stock Price Forecasting via Sentiment Analysis on Twitter. In Proceedings of the 20th Pan-Hellenic Conference on Informatics (p. 36). ACM.

Kwak, H., Lee, C., Park, H. and Moon, S., 2010, April. What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World Wide Web (pp. 591-600). ACM.

Lee, L. (2010). Silencing the twittering juror: The need to modernize pattern cautionary jury instructions to reflect the realities of the electronic age. DePaul Law Review 60(1), 181-222.

Ledward, Jocelyn & Agate, Jennifer (2014) 'Cyberbullying: legislative reform and an international perspective' Entertainment Law Review, 25(7), 236, at 238

Li, R., Lei, K.H., Khadiwala, R. and Chang, K.C.C., 2012, April. Tedas: A twitter-based event detection and analysis system. In Data engineering (icde), 2012 ieee 28th international conference on (pp. 1273-1276). IEEE.

Llewellyn, C. and Cram, L., 2016, March. Brexit? Analyzing Opinion on the UK-EU Referendum within Twitter. In ICWSM (pp. 760-761).

Mittal, S., Das, P.K., Mulwad, V., Joshi, A. and Finin, T., 2016, August. CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities. In Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on (pp. 860-867). IEEE.

Murthy, D., Gross, A. and McGarry, M., 2016. Visual Social Media and Big Data. Interpreting Instagram Images Posted on Twitter. Digital Culture & Society, 2(2), pp.113-134.

O'Connor, C. (2013). Cutting cyberstalking's gordian knot: simple and unified statutory approach. Seton Hall Law Review 43(3), 1007-1040.

O'Flinn, James (2013) 'Pushing back the trolls' S.J. 157(32), 11.

RIAForge (n.d.) '(monkeh) Tweet Twitter API' Available at: http://monkehtweet.riaforge.org/ (Accessed 7 June 2016)

Scaife, Laura (2013) 'The DPP and social media: a new approach coming out of the Woods?' Communications Law, 5

Sheela, L.J., 2016. A Review of Sentiment Analysis in Twitter Data Using Hadoop. International Journal of Database Theory and Application, 9(1), pp.77-86.

Spencer, T. (2017). Twitter in the age of terrorism: Can retweet constitute true threat. First Amendment Law Review 15(3), 497-521.

Statista (2016a) 'Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2016 (in millions)' Available at: http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ (Accessed 26 May 2016)

Statista (2016b) 'Number of data removal requests issued to Twitter from January 2012 to December 2015' Available at: http://www.statista.com/statistics/315147/total-number-of-requests-for-data-removal-twitter/ (Accessed 26 May 2016)

Statista (2016c) ‘Most popular global events on Twitter as measured in tweets per minute in 2012’ Available at: http://www.statista.com/statistics/249222/worldwide-events-with-the-most-tweets-per-minute/ (Accessed 26 May 2016)

Statista (2016d) ‘Most popular Super Bowl 50 moments on Twitter as measured in tweets per minute on February 7, 2016’ Available at: http://www.statista.com/statistics/252221/super-bowl-moments-with-the-most-tweets-per-minute/ (Accessed 26 May 2016)

Statista (2016e) ‘Number of social network users worldwide from 2010 to 2019 (in billions)’ Available at: http://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ (Accessed 25 June 2016)

Twitter (n.d.a) ‘Getting started with Twitter’ Available at: https://support.twitter.com/articles/215585 (Accessed 26 May 2016)

Twitter (n.d.b) ‘The Search API’ Available at:
https://dev.twitter.com/rest/public/search (Accessed 26 May 2016)

Twitter (n.d.c) ‘Obtaining access tokens’ Available at:
https://dev.twitter.com/oauth/overview (Accessed 26 May 2016)

Twitter (n.d.d) ‘Connecting to a streaming endpoint Overview’ Available at:
https://dev.twitter.com/streaming/overview/connecting (Accessed 26 May 2016)

Twitter (2016a) ‘Deactivating your account’ Available at: https://support.twitter.com/articles/15358 (Accessed 26 May 2016)

Twitter (2016b) ‘Guidelines for law enforcement’
Available at: https://support.twitter.com/articles/41949#3 (Accessed 26 May 2016)

Twitter (2016c) ‘Why am I missing from search?’ Available at: https://support.twitter.com/articles/66018 (Accessed 26 May 2016)

Twitter (2016d) ‘Information Requests’ Available at: https://transparency.twitter.com/information-requests/2012/jul-dec (Accessed 26 May 2016)

Twitter (2016e) ‘Information Requests’ Available at: https://transparency.twitter.com/information-requests/2015/jul-dec (Accessed 26 May 2016)

Twitter (2016f) ‘Developer Policies to Protect People’s Voices on Twitter’ Available at: https://blog.twitter.com/2016/developer-policies-to-protect-people-s-voices-on-twitter (Accessed 13th April 2017)

Twitter (2017) ‘GET direct_messages’ Available at: https://dev.twitter.com/rest/reference/get/direct_messages (Accessed 13th April 2017)

Twitter4J (n.d.) ‘Introduction’ Available at: http://twitter4j.org/en/index.html (Accessed 7 June 2016)

Venkatesh, R., Rout, J.K. and Jena, S.K., 2017. Malicious Account Detection Based on Short URLs in Twitter. In Proceedings of the International Conference on Signal, Networks, Computing, and Systems (pp. 243-251). Springer India.

Appendix A

Results Table: The results table contains data pertaining to tweets (see: https://dev.twitter.com/overview/api/tweets)

Field Name	**Data Type**	**Description**
id_str	String	The string representation of the unique identifier for this Tweet.
created_at	Timestamp	UTC time when this Tweet was created.
text	String	The actual text of the status update.
favorite_count	Integer	Indicates approximately how many times this Tweet has been “liked” by Twitter users.
in_reply_to_screen_name	String	If the represented Tweet is a reply, this field will contain the screen name of the original Tweet’s author.
in_reply_to_status_id_str	String	If the represented Tweet is a reply, this field will contain the string representation of the original Tweet’s ID.
in_reply_to_user_id_str	String	If the represented Tweet is a reply, this field will contain the string representation of the original Tweet’s author ID. This will not necessarily always be the user directly mentioned in the Tweet.
is_possibly_sensitive	Boolean	This field only surfaces when a tweet contains a link. The meaning of the field doesn’t pertain to the tweet content itself, but instead it is an indicator that the URL contained in the tweet may contain content or media identified as sensitive content

lang	String	When present, indicates a BCP 47 language identifier corresponding to the machine-detected language of the Tweet text.
geo_lattitude	Decimal	The latitude of the status update, if the user has geotagged their Tweet.
geo_longitude	Decimal	The longitude of the status update, if the user has geotagged their Tweet.
metadata_iso_language_code	String	
metadata_result_type	String	
place	String	When present, indicates that the tweet is associated (but not necessarily originating from) a place
place_country	String	Name of the country containing this place.
place_country_code	String	Shortened country code representing the country containing this place.
place_full_name	String	Full human-readable representation of the place’s name.
place_id	String	ID representing this place.
place_name	String	Short human-readable representation of the place’s name.
place_type	String	The type of location represented by this place (e.g. 'city').
retweet_count	Integer	Number of times this Tweet has been retweeted.
source	String	Utility used to post the Tweet, as an HTML-formatted string. Tweets from the Twitter website have a source value of 'web'.
truncated	Boolean	Indicates whether the value of the text parameter was truncated, for example, as a result of a retweet exceeding the 140 character Tweet length. Truncated text will end in ellipsis, like this ... Since Twitter now rejects long Tweets vs truncating them, the large majority of Tweets will have this set to 'false'.
user_id_str	String	The string representation of the unique identifier for this User.

user_created_at	Timestamp	The UTC datetime that the user account was created on Twitter.
user_default_profile	Boolean	When true, indicates that the user has not altered the theme or background of their user profile.
user_description	String	The user-defined string describing their account.
user_favourites_count	Integer	The number of tweets this user has favorited in the account's lifetime.
user_followers_count	Integer	The number of followers this account currently has.
user_friends_count	Integer	The number of users this account is following.
user_geo_enabled	Boolean	When true, indicates that the user has enabled the possibility of geotagging their Tweets.
user_lang	String	The BCP 47 code for the user's self-declared user interface language. May or may not have anything to do with the content of their Tweets.
user_listed_count	Integer	The number of public lists that this user is a member of.
user_location	String	The user-defined location for this account's profile. Not necessarily a location nor parseable. This field will occasionally be fuzzily interpreted by the Search service.
user_name	String	The name of the user, as they've defined it. Not necessarily a person's name. Typically capped at 20 characters.
user_profile_background_color	String	The hexadecimal color chosen by the user for their background.
user_profile_background_image_url	String	A HTTP-based URL pointing to the background image the user has uploaded for their profile.
user_profile_background_image_url_https	String	A HTTPS-based URL pointing to the background image the user has uploaded for their profile.
user_profile_background_tile	Boolean	When true, indicates that the user's profile_background_image_url should be tiled when displayed.

user_profile_banner_url	String	The HTTPS-based URL pointing to the standard web representation of the user’s uploaded profile banner.
user_profile_image_url	String	A HTTP-based URL pointing to the user’s avatar image.
user_profile_image_url_https	String	A HTTPS-based URL pointing to the user’s avatar image.
user_profile_link_color	String	The hexadecimal color the user has chosen to display links within their Twitter UI.
user_profile_sidebar_border_color	String	The hexadecimal color the user has chosen to display sidebar borders with in their Twitter UI.
user_profile_sidebar_fill_color	String	The hexadecimal color the user has chosen to display sidebar backgrounds with in their Twitter UI.
user_profile_text_color	String	The hexadecimal color the user has chosen to display text with in their Twitter UI.
user_profile_use_background_image	Boolean	When true, indicates the user wants their uploaded background image to be used.
user_protected	Boolean	When true, indicates that this user has chosen to protect their Tweets.
user_screen_name	String	The screen name, handle, or alias that this user identifies themselves with. Screen names are unique but subject to change.
user_statuses_count	Integer	The number of tweets (including retweets) issued by the user.
user_time_zone	String	A string describing the Time Zone this user declares themselves within.
user_url	String	A URL provided by the user in association with their profile.
user_utc_offset	String	The offset from GMT/UTC in seconds.
user_verified	Boolean	When true, indicates that the user has a verified account.
id_str	String	The string representation of the unique identified for these coordinates.
latitude	Decimal	The latitude point of the bounding box.
longitude	Decimal	The longitude point of the bounding box.