A survey of current social network and online communication provision policies for offender identification

Graeme Horsman
Department of Computing, Engineering and Technology
Faculty of Computer Science

The David Goldman
Informatics Centre
St Peter's Way
Sunderland
SR6 0DD

Email: graeme.horsman@sunderland.ac.uk
Phone: 0191 515 2381

Abstract
Online forms of harassment, stalking and bullying on social network and communication platforms are now arguably wide-spread and subject to regular media coverage. As these provision continue to attract millions of users, generating significant volumes of traffic, regulating abuse and effectively reprimanding those who are involved in it, is a difficult and sometimes impossible task. This article collates information acquired from 22 popular social network and communication platforms in order to identify current regulatory gaps. Terms of service and privacy policies are reviewed to assess existing practices of data retention to evaluate the feasibility of law enforcement officials tracking those whose actions breach the law. For each provision, account sign-up processes are evaluated and policies for retaining Internet Protocol logs and user account information are assessed along with the availability of account preservation orders. Finally, recommendations are offered for improving current approaches to regulating social network crime and online offender tracking.

Keywords: Social Media Networks; Crime; Regulation; Internet; Cyber Crime; Harassment

1 Introduction
Online platforms have now revolutionised modern day communication. However, in light of recent global events, social media has now become a platform for those to voice both positive and negative sentiment, requiring greater regulation by both the social networking sites themselves and the police (Awan, 2016). With a reported 2.3 billion social network users worldwide (Statista, 2016), the regulation of user behaviour on these platforms is a difficult task. In 2015, Vodafone and YouGov surveyed around 5,000 teenagers across 11 countries, where 41% of respondents admitted to feeling depressed or helpless from acts of cyberbullying and a further 18% feeling suicidal (Vodafone, 2015). In addition, a quarter of those surveyed had actively closed their social media accounts due to acts of cyberbullying (Vodafone, 2015). Similarly, the Association of School and College Leaders (2016) reported that 41% of the school leaders surveyed reported an increase in acts of students being cyberbullied. In 2016, anti-bullying charity Ditch the Label (2016) surveyed 8,850 persons aged 12-20, with 6 out of 10 of those reported to have been bullied, indicating that they had experienced this online. Despite the many benefits offered by online communication and social networks, a darker side is also apparent.
Social networks and online forms of communication are frequently identified as problems in the battle against online harassment and abuse. In 2014, ‘a total of 38 out of 45 police forces saw a rise in the number of crime reports that involved Facebook’ (Birchley, 2015) with ‘the Metropolitan Police, receiving 1,207 crime reports which mentioned Facebook, up from 935 in 2013 and 997 in 2012’ (Evans, 2015). Further Evans (2015) reports that over ‘16,000 alleged crimes involving Facebook and Twitter were reported’ across all United Kingdom (UK) police forces for the period of 2014/15. The Twitter platform is regularly subject to scrutiny due to the volume of trolling (an act of posting ‘inflammatory or inappropriate messages or comments online for the purpose of upsetting other users and provoking a response’ (Dictionary.com, 2016)) which occurs (BBC News, 2016a) and the service has been criticised for failing to be proactive in regulating and removing offending content, such as that posted by the extremist cleric Anjem Choudary (BBC News, 2016b). Other examples of social network abuse include the 2011 England riots where such provision were used to organise mass congregations and crime, with Williams et al. (2013) suggesting that at the time, police were ill-equipped to deal with analysing this content. Yet it remains questionable as to whether some five years later, law enforcement are in a better position to tackle these issues.

Reports of sexist and misogynistic comments targeting those on Youtube and Twitch have also received media coverage (BBC News, 2016c), yet such incidents form merely a small part of a far greater issue. Frequently high-profile personalities are targeted, where recent examples include Stephen Fry, Jennifer Lawrence, Matt Lucas and Sinead O’Connor, prompting their exit from such platforms (Cohen, 2014). In addition, attacks upon Sara Payne, the mother of murdered school girl Sarah Payne, and Zelda Williams, the daughter of the now deceased Robin Williams show an altogether more sinister side of the harassment which can be suffered online (Cohen, 2014). These instances form only a small subset of the overall volume of abuse which is experienced by everyday individuals. Acts of online abuse can now be considered relatively common and form part of a greater issue and debate surrounding the need for greater regulation of social networks, a point alluded to in the House of Commons Home Affairs Committee (2016) report into Radicalisation.

Social media companies are consciously failing to combat the use of their sites to promote terrorism and killings. Networks like Facebook, Twitter and YouTube are the vehicle of choice in spreading propaganda and they have become the recruiting platforms for terrorism. They must accept that the hundreds of millions in revenues generated from billions of people using their products needs to be accompanied by a greater sense of responsibility and ownership for the impact that extremist material on their sites is having (House of Commons Home Affairs Committee, 2016, p.34).

Hate crime is also becoming an increasing issue for social network platforms. In the wake of the UK’s vote to leave the European Union (commonly referred to as ‘Brexit’), MP Andy Burnham highlighted a subsequent fivefold increase ‘in race hate comments on social media channels’ (HC Deb, 2016). This is backed by GLA Conservatives’s (2015) survey which reported 68% of the 308 individuals reviewed had encountered hate crime online. The Mayor’s Office for Policing And Crime (MOPAC) (2016) states that currently social media is providing offenders with a ‘veil of anonymity’ which is prohibiting effective regulation of their
conduct, and have recently acquired funding from the Home Office Police Innovation Fund (PIF) to develop an online hate crime hub (MOPAC, 2016).

In any of the aforementioned acts, where forms on online content overstep the mark and fall foul of domestic or international legislation, the identification on an offender is key to the effective regulation of illegal behaviour. Studies have shown online environments can ‘lower behavioral inhibitions’, encouraging disclosures and derogatory actions (Suler, 2004; Lapidot-Lefler and Barak, 2012), yet where an account holder cannot be identified there is a lack of accountability for their conduct. This provides an issue for law enforcement when trying to regulate and apprehend social network offenders, potentially leaving any victims vulnerable for sustained online abuse. This article examines the terms of service, privacy policies and functionality of 22 social network and communication provision in an effort to establish the current feasibility of tracking offenders who post content on these platforms in breach of both policy and law. Account sign-up processes are evaluated along with policies for the retention of data which could be used to identify those in breach. Finally, conclusions and recommendations are drawn.

2 Regulatory problems
Guidelines supporting those subject to cyberbullying and online harassment on social networks exist on various organisation portals designed to support those subject to these acts. Childline (2016) identify Facebook, Twitter, Instagram, Instant messaging (IM), Snapchat, ASKfm and Tumblr, and provide guidance for those subject to abuse on these platforms and how to block and report it. The charity ‘Family Lives’ (2016) provides guidance on dealing with cyberbullying on Facebook, Twitter, Youtube, Whatsapp, Snapchat and Instagram. Other charities offering similar advice and resources include The CyberSmile Foundation (2016) and the NSPCC (2016). Policing social media content is notoriously difficult and arguably, we are yet to see effective forms of regulation in force across many platforms. The Select Committee on Communications's Report (2014) indicated that this is due to the fact that ‘there is no consistent attitude taken by website operators: some require the use of real names (Facebook, although they do not actively confirm users' identities); some allow anonymity but challenge impersonation (Twitter) and others allow absolute anonymity'.

The volume of users combined with large quantities of network traffic continue to pose issues (Kavanaugh et al., 2012). Techniques for regulating online social network content typically fall within one of two categories, proactive or reactive. Proactive measures address content before and as it happens and attempt to prevent its appearance on a given platform in the first instance. Online filters and keyword matching are methods for highlighting posts of a particular type and prevent certain forms of language from being submitted (Bekkers et al., 2013). Yet the speed of linguistic developments mean that these methods can only serve a limited purpose and may quickly become ineffective as new offensive terms or phrases are developed or ways to circumvent their use are discovered (through the use of punctuation, special symbols to break up the plain text meaning of a word). The application of sentiment analysis has also been offered as a way of improving the identification of offending messages (Ceron et al., 2014). Social media platforms have also taken steps to encourage users to be proactive about reporting incidents online as opposed to waiting for a response from the network itself, introducing the notion of self-policing and user-regulation. Facebook have an inbuilt reporting system (Facebook, 2016e) with similar process witnessed on other
platforms such as Twitter (2016d) and Instagram (2016d). Yet despite such methods, it remains arguable that the complete prevention of abuse is unachievable. Regardless of form, where content is posted that reaches its intended target (i.e. a victim’s account) in breach of regulations, a reactive response must be formed in order to reprimand those responsible.

Where message content breaches platform policies or legislation, it may be deemed necessary to identify and prosecute the individual responsible for the post. This is particularly necessary in numerous cases including those of online harassment and threatening behaviour where in the UK, the circumstances of the case satisfy the test defined in the Code for Crown Prosecutors (Crown Prosecution Service, n.d.). The test is twofold where first evidential sufficiency must be achieved (‘a prosecutor must be satisfied that there is sufficient evidence to provide a realistic prospect of conviction’), before consideration must be given as to whether a prosecution is in the public interest. Before this test can be implemented, consideration must be given as to whether there is sufficient evidence available allowing the physical poster of any message content to be identified in the first instance. This can be a difficult process, and one where success is subject to the governance and guidelines of the platform from which the abusive content took place. In many circumstances, there is insufficient data available to identify account holders, hindering effective law enforcement investigation. In absence of the ability to identify an offender, there can often be no legal case to proceed with. On many platforms, regulatory issues exist right from the account creation process, where fake information can often be used to open an account (Barrett, 2016), providing a starting point for analysis in this article.

2.1 Implementing effective regulatory measures for identifying account holders

In simple terms, effective regulatory measures on social networks allow for the assessment and evaluation of posted content and the enactment of measures to reprimand those who breach both the social media platform terms of use or legal authority. In order to do this, a social media platform should be able to hold its users to account for the content they post, where essentially they must be identifiable when behavioural breaches occur. To achieve this, social media networks have three current options for account validation, namely direct, indirect and metadata validation:

*Direct validation:* Direct validation requires the user to submit accurately their name and identifying information during the creation of their account, and, being able to attribute any posted content to these details. The problem here relates to the term ‘accurate’ and the fact that as already highlighted, most social media networks do not accurately validate user input. Typically, social network signup procedures require users to input criteria which include name, age and email address. However there is generally no processes in place to accurately authenticate this content. Despite so-called ‘real-name’ policies, seen with platforms such as Google+ (however, their policy was subsequently rescinded in 2014 (Vincent, 2014)) and Facebook (2016e), name data is not authenticated. As a result, account name data cannot be relied upon as a source of offender identifiable information for law enforcement in some cases, which has lead to an abundance of unattributable accounts created solely for purposes such as trolling. The implementation of an effective name validation process would be onerous and require acts similar to those seen in online
monetary transactions or where documentary proof of identity can be captured and verified. When considering that many of the social network provisions seek to attract users via a no-fuss sign-up process to prevent users from disengaging, it is unlikely that these type of verification processes will ever be implemented. Similar concerns surround the use of a valid email address (see Section 3.2 for an analysis of sign-up processes).

**Indirect Validation:** Indirect validation methods may be used to identify an account holder based on their actions and is subject to the behaviour of the offender. For example, those users who have their identity validated indirectly include those who post or send identifiable message content through disclosure of personal information or media. This can include for example, the posting of imagery where an offender is present and can be identified or textual content disclosing identifiable details; these could be potentially accidental disclosures. The problem faced here is that those who create accounts solely for the purpose of abuse are likely to omit the inclusion of personally identifiable content and therefore the use of indirect validation methods is limited.

**Metadata Validation:** Metadata validation surrounds the attributing of an account to an individual from the metadata left behind by their interaction with the service. The effectiveness of this validation form is often dependant on the service provider and their policies regarding the collection and retention of data, and, their willingness to co-operate with law enforcement. Typically relevant metadata could include:

1. Internet Protocol (IP) Address: IP information can (not always, subject to spoofing) be used to identify an internet connection from which offending communications were made, and potentially their location.
2. Device information: This includes information such as the user’s device type and telephone number if they are accessing services via a mobile device.
3. Geolocational data relating to the poster of content.

The above information can be classed as ‘log data’ and is subject to the following limitations.

1. **Log retention:** In regards to metadata account validation, the retention of logged data poses a number of issues. The first issue comes from the assumption that relevant data is stored in the first instance. In cases where no appropriate logs are retained, account identification via metadata is not feasible, creating a significant issue, as account usage metadata is arguably the most likely source of data which is of use to law enforcement. Even when log data is retained, the length of time is critical, where sufficient time must be provided to ensure the necessary legal authority is sought in order to obtain the data. Given that many platforms witness
significant volumes of traffic, a factor impacting the length of time
data is retained is often the cost of having to store the data in the
first place. It is not feasible for many organisations to suffer the
cost of retaining all traffic indefinitely, but in some instances, where
data is only kept for a matter of weeks, it may not be enough time
for an effective response. Therefore log retention times can hinder
investigation attempts.

2. Type of log data collected: The type of data collected is also key
to an effective investigation. Service providers may omit to collect
the necessary metadata to validate an account.

3. Account and data deletion: The removal of posted content can
pose an issue, for example where an individual posts a message
and then intentionally deletes it. If the victim has not made a record
of the offending messages then law enforcement remain reliant on
the service provider to have retained a copy of this content. The
problem here is two-fold, where issues come from those who
delete individual offending content, and/or, decide to delete the
offending account. As with the problem of log retention noted
above, storage space becomes an issue. Given the volume of
traffic, it is unlikely that inactive content (i.e. deleted) will be
retained for any substantial period of time or potentially at all.
Therefore if the user choses to a) delete content and/or b) delete
the offending account. Law enforcement are reliant on the service
providers maintenance of retrievable content. In some cases, this
may not be available.

3 Taking a look at existing platforms
This article has highlighted 22 social network, online communication and blogging websites
in order to analyse their current account signup processes, terms and conditions and privacy
policies in an attempt to assess the feasibility of tracking the physical users of offending
accounts. These provision were chosen based on their popularity and having been
highlighted as potentially problematic platforms by charities like the NSPCC (2016) and
Childline (2016).

The services analysed are:-

Facebook, Twitter, MySpace, Instagram, Snapchat, Youtube, Reddit, ASKfm, Tumblr, Vine,
Bebo, Quora, Flickr, Hi5, Pinterest, Buzzfeed, Badoo, Scribd, Wordpress, Foursquare,
Vimeo and LinkedIn.

3.1 Age of use
As with most online services, the terms and conditions associated to each platform define
the age a person must be before they can legally sign up for the service. From the 22
services, 16 define this age as 13 years and above. Badoo, Hi5 and Buzzfeed identify this
age as 18. LinkedIn specify age requirements specific to country of origin (13 years old for
all countries except People’s Republic of China, Netherlands, United States, Canada,
Germany, Spain, Australia and South Korea). Finally Youtube and Vine state that users must be old enough to form a binding legal contract. Despite defining an age, similar to the issues with validating name information on account sign-up, there are no age validation procedures when a user creates their account (although many services (see Myspace (2016d) for example) will delete your account if they suspect you are not of an appropriate age) and therefore this information cannot be relied upon to support tying a physical individual to an account. This lack of validation is also potentially placing some population demographics at risk, particularly children. Nominet’s (2014) survey into child social media usage states that by ‘age 10, over half (59%) of children have used a social network’ in breach of platform terms and conditions. As a result, ‘21% of those surveyed stated they had posted negative comments by the average age of 11 and 43% had messaged strangers by an average age of 12’ (Nominet, 2014). Such concerns are exacerbated when it is considered that any form of inappropriate content targeted at a child may not be attributable to a physical offender, through a lack of traceability on these platforms.

3.2 Sign-up process

The account sign-up process for the 22 services highlighted in this article provides the first source of potentially useful information for tracking down an individual attributed to an offending account. When analysing the signup process it was found that all 22 services requested at least the name of the individual (full or first and last names), with no means of validating the accuracy of inputted content. All 22 services offered signup via an email address, with Twitter and Facebook offering a dual sign-up process of either email or mobile telephone number (verified via SMS message). Of the 22 services, only LinkedIn prevented users from using the service (successfully logging on) without first validating the email address used to signup for an account via sending a registration confirmation email to that account. As a result, the other 21 services could be accessed by supplying fake name details and a fake email address, providing that it was entered in the format of a legitimate email account provider. For tests within this article, a fake randomly generated email prefix string was added to ‘@gmail.com’ and successfully used to sign in. Although this may seem trivial, in reality it causes greater difficulty for law enforcement attempting to track individuals. Not only does it encourage the creation of trolling accounts as it can take seconds to generate an account on these platforms without the need to first create a valid email address, but it also lessens the chance of the user inputting identifiable content in the signup process. For example, where a valid, accessible email account is needed in order to activate an account on sign-up for any of the above social network services, in some cases an individual may opt to use their personal email address without contemplating the potential to be identified later, which may have been set up for legitimate purposes with accurate details (see Figure 1 for an example of how the use of an email service provider may also provide traceable information). At which point, a service provider may store a record the email address used at signup and attribute it to the offending account and therefore law enforcement may be able to contact to the email service provider for additional content.

Figure 1. An overview of the signup process

Yet the problem faced by social networks and the lack of validation of account sign-up details also exists with email service providers. This article also analysed the signup process at both Outlook and Google Mail in order to acquire an email address, arguably two of the most popular email service providers. Outlook requested that on sign-up, users provide a
secondary backup email address and verification mobile number in addition to name, location and age data. The benefit here is that where a user must provide a mobile number to verify, providing a record of the number is kept, network service provider records may be queried in order to potentially identify the actual user of the account (subject to spoofing and pay-as-you-go numbers). Outlook requires this data before allowing the account to be created, however it accepted a number of ‘55555555555’ and a randomly generated email address, with the suffix ‘@gmail.com’. The problem this presents is that even though there is a requirement for what appears to be extra validation, it also cannot be relied upon. Similar processes were witnessed on the Google Mail platform.

3.3 Terms, conditions and privacy
Given the issues present on sign-up for the 22 services, offending tracking may be reliant on what was previously coined as metadata account validation in Section 2.1. To assess the feasibility of law enforcement achieving this, Table 1 presents a breakdown of the key inclusions in the terms of service and privacy policies of all 22 platforms, with a focus on IP log retention and account content and deletion procedures.

<table>
<thead>
<tr>
<th>Service</th>
<th>Is IP info stored?</th>
<th>How long is it retained?</th>
<th>Key policy points of interest</th>
</tr>
</thead>
</table>
| Facebook | Yes | Does not say. | **IP related information:** “When you delete IP content, it is deleted in recycle bin on a computer. However, you understand that removed copies for a reasonable period of time (but will not be available to others).”  
**Information around data deletion:** “Information associated with account is deleted, unless we no longer need the data to provide products and services.” Accounts can be deactivated, essentially placing it in a suspended state. After the account is canceled, the following are key points of interest in deletion information (Facebook, 2016c):
- Deletion is delayed for a few days after it's requested. If you canceled.
- Access can't be regained once the account is deleted.
- It may take up to 90 days to delete data stored in backup systems on Facebook during this time.
- Data related to an account may not be deleted. For example, if we remove copies of some material (ex: log records) may remain disassociated from personal identifiers.

Additional retained device information: “Attributes such as the operating system, device settings, file and software names and types, battery and signal strength, device locations, including specific geographic locations, such as GPS signals. Connection information such as the name of your mobile device, language and timezone, mobile phone number and IP address” (Facebook, 2016g). Users can self-disclosure account information and download their data.

| Twitter | Yes | Does not say. | **IP related information:** “Given Twitter's real-time nature, some information may be stored for a very brief period of time” (Twitter, 2016b). The privacy policy may state that data may be available for up to 18 months.  
**Information around data deletion:** Users can deactivate their account.
reinstate the account before permanent deletion. “After deactivation within a few minutes, however some content may be viewable on twitter (Twitter, 2016a). If the account is not reinstated after 30 days, Twitter will permanently delete the account, which can take up to 1 week (Twitter, 2016a). Some content may be viewable via search engines beyond Twitter’s control.

Twitter’s ‘default is almost always to make the information you provide public for as long as you do not delete it’ (Twitter, 2016c). ‘Content (Tweets) is generally not available’ (Twitter, 2016b).

<table>
<thead>
<tr>
<th>Platform</th>
<th>Access</th>
<th>Retention Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instagram</td>
<td>Yes</td>
<td>Does not say.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>IP related information:</strong> Given the volume of real-time content on Instagram, user-generated content can only be stored for a short period of time (Instagram, 2016a).</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Information around data deletion:</strong> Users have both the option to recycle, delete or disable their account option (Instagram, 2016b). There is no indication on the Instagram user profile or sign up details, analytics, log and device identifier content are retained.</td>
</tr>
<tr>
<td>Myspace</td>
<td>Yes</td>
<td>Does not say.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Information around data deletion:</strong> As per the Myspace Privacy Policy, if you delete your Account or delete information or Profile Content, copies of some Profile(s) may remain viewable in circumstances where, for example, another User’s Profile, shared information with a Third-Party Link, another Member copied, stored or shared your information or has a copy of the Myspace Services. To the extent permitted by applicable law or our internal security policies. Please be aware that, due to mobile technologies outside of our control, such as caching and new devices, information may not be instantly inaccessible to others, and there may be information and content from elsewhere on the internet and from search engines.</td>
</tr>
<tr>
<td>Foursquare</td>
<td>Yes</td>
<td>Does not say.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Information around data deletion:</strong> As per Foursquare’s privacy policy, we maintain Foursquare, such deletion may not be immediate, and information or posts may remain on backup media for up to nine months. For the foregoing, we will retain information as required by applicable law or our internal security policies. Please be aware that, due to mobile technologies outside of our control, such as caching and new devices, information may not be instantly inaccessible to others, and there may be information and content from elsewhere on the internet and from search engines.</td>
</tr>
<tr>
<td>Snapchat</td>
<td>Yes</td>
<td>Does not say.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Information around data deletion:</strong> As per Snapchat’s privacy policy, we automatically delete the content of your Snaps (the photo and video) and messages (text and audio) from our servers after we detect that a Snap has been “expired” (Snapchat, 2016a). In addition, the policy also acknowledges that Snapchats from the local device - “It’s also possible, as with any digital information, to access messages forensically or find them in a device’s temporary storage.”</td>
</tr>
</tbody>
</table>

The period of time information exists before deletion depends on the type of content:

- **Snapchat:**
  - Snaps: Automatically deleted after they’ve been viewed and typically cannot be retrieved from Snapchat’s servers by anyone. Snapchat’s servers are designed to automatically delete unopened Snapchats after they expire. (Snapchat, 2016a).
  - Chat messages: Automatically deleted after sender and recipient close the Chat screen – unless either party presses and holds the Chat button to prevent deletion.
  - My Story: Automatically deleted 24 hours after each Snap is sent.

---
<table>
<thead>
<tr>
<th>Site</th>
<th>Availability</th>
<th>Account Deletion Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live Stories</td>
<td>Live Story or Local Story Snaps may be (Snapchat, 2016c).</td>
<td>Information around account deletion: As per Snapchat's privacy policy, if you delete your account, you will have up to 30 days to restore your account before our servers. During this period of time, your account will not be available to others. (Snapchat, 2016a).</td>
</tr>
<tr>
<td>Youtube</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>Reddit</td>
<td>Yes</td>
<td>Potentially 100 days (Reddit, 2016a)</td>
</tr>
<tr>
<td>ASKfm</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>Tumblr</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>Vine</td>
<td>Yes</td>
<td>Potentially up to 18 months</td>
</tr>
<tr>
<td>Service</td>
<td>Dr.</td>
<td>Does not say</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>--------------</td>
</tr>
<tr>
<td>Badoo</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>Scribd</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>Wordpress</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>Buzzfeed</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>Quora</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>Flickr</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pinterest</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>Hi5</td>
<td>Yes</td>
<td>Indefinitely</td>
</tr>
<tr>
<td>Vimeo</td>
<td>Yes</td>
<td>Does not say</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Yes</td>
<td>Potentially 7 days after account deletion, at which point logs are depersonalized (LinkedIn 2016a). However, LinkedIn’s law enforcement data retention policy states that there is a 24 month limit on IP information from law enforcement data.</td>
</tr>
</tbody>
</table>
4 Discussion of regulations

As can be seen within Table 1, the stance taken by social media platforms in regards to data deletion and retained information varies. To analyse this content, discussions are broken down into two areas, data deletion and IP information.

4.1 The problem with deleting content

Content on these platforms is volatile and often vulnerable to the original poster’s intentions to delete or keep any posted messages. As a result, police guidelines indicate that an individual should attempt to make record of any offending content as soon as it is identified as it may be the only chance to capture a record of it (ACPO, 2012; Hampshire Police, n.d.; Westyorkshire Police, n.d.). The problem remains that even where content is identified or recorded by a victim, it could be removed at any point from the service in question, potentially before relevant metadata can be preserved for identification purposes. There is no consistent approach to deleted data in the 22 platforms analysed, and although guidance is provided around deleting an account, there is little guidance as to how deleted account content is handled or archived by service infrastructure. In relation to deleted account content, where an offender chooses to delete their actual account time is crucial to ensuring potential evidential data is retained. The problem here is that law enforcement must obtain the relevant legal authority to comply with the terms and conditions of the social network services before data can be disclosed and this process can be slow. This is compounded by the fact that there is no actual defined consistent period of retention from which law enforcement authorities can make an informed judgement as to the feasibility of requesting data from these services. This ambiguity within policies can cost both time and money, leading to unsatisfactory case outcomes, and provides an area in need of improvement and clarity.

From the 22 platforms examined in this article, the following platforms provide guidance targeted at law enforcement: Facebook (2016d), Pinterest (2016b), Twitter (2016b), Snapchat (2016b), Myspace (2016b), Tumblr (2016b), Instagram (2016a), ASKfm (2016b), Badoo (2016a), Youtube (Google, 2016), Reddit (2016b), LinkedIn (2016b), Flickr (Yahoo, 2016b), Foursquare (2016b) and Wordpress (2016). These guidelines aim to explain the processes involved in submitting a request for information. One of the key processes for
supporting law enforcement when interacting with social network platforms is the preservation order, allowing potentially evidential account data to be stored (not disclosed), pending relevant legal authority. Here, requests can be made to preserve data in connection with an offence, providing the correct legal authority is obtained for future disclosure, where generally a 90 period of preservation is set (Brunty and Helenek, 2014). For example, Facebook will preserve account data for 90 once relevant legal requests have been made (Facebook, 2016d). A 90 day preservation period is also obtainable from Twitter (2016b), Tumblr (2016b), ASKfm (2016b), Reddit (2016b), Instagram (2016a), Foursquare (2016b), Flickr (Yahoo, 2016b), LinkedIn (2016b), Snapchat (2016b) (one additional 90-day period of extension can be requested in addition, making a total of 180). Myspace (2016c) preserve account data for 180 days. Wordpress (2016b) preserve data for 45 days and Pinterest’s (2016b) law enforcement guidance does not define a period of preservation.

Often, the preservation order must specify the boundaries of the request in terms of time and amount of information from a specific account (Sammons, 2015). A blanket request for information may not be suitable, placing emphasis on suitable recognisance around an event to ensure that evidential information is contained within the requested period. It must also be noted that many social networks inform the account holder subject to a preservation order once it has been submitted (Sammons, 2015).

Many current investigations involve Social Networking Sites. It is imperative that early consideration is made around securing Social Networking Profiles that fall within the investigation. The best evidence is available from the service provider however they are often located outside of the UK and may or may not secure the content on the appropriate request via the force CSP/ISP SPOC. As such the investigator should always secure a copy of what is seen by them as this may be the only opportunity to secure this evidence before it changes. ACPO!!!!- link in

4.2 IP logs
As highlighted previously, IP log information can be a key source of information for identifying offenders. However, none of the 22 platforms clearly define their log retention period. For example, Reddit states that ‘will delete any IP addresses collected after 100 days’ (Reddit, 2016a) but does not indicate what IP information it retains and whether it stores it for at least 100 days before deletion, with Hi5 stating this information is kept indefinitely (Hi5, 2016a) and Bebo does not comment on the issue. From Table 1, it can be seen that a common stance is to indicate that IP logs are collected but to omit to determine a time frame. As with ambiguity surrounding account and content deletion policies, the lack of a determined and consistent stance on retention can be detrimental to law enforcement’s ability to mount an effective investigation and in some circumstances may also deter an attempt in belief that the relevant information does not exist.

4.2.1 Request method and extracted data formats
Request methods and extracted data format Establishing levels of access and engagement with social media platforms by law enforcement is difficult. The procedures associated with making a request for data vary depending on the social media platform, with limited information available. Facebook operate a ‘Law Enforcement Online Request System’ accessible and authenticated through the use of a valid law enforcement email address (also
used for their associated product Instagram (2017)) (Facebook, 2017). Instagram highlight that those who do not submit via the law enforcement portal (opting for mail or email), then longer response rates may be witnessed (Instagram (2016a)). This process should be compared with services such as Myspace who simply offer contact methods of fax, mail or email, directed towards an appropriate legal department (Myspace (2016c)). In addition, there is limited information available within available policy information on the platforms analysed to indicate expected request response times. Establishing the format which law enforcement will receive any retained data is also difficult to ascertain due to limited disclosure of information in policy information. Twitter (2016b) indicate that content will be provided in electronic format which can be opened using generic word processing software. Further, Google’s ‘Takeout’ function allows standard users to download an archive of their data containing information relating to Google’s various products, where the format of data varies depending on content (for example, contacts are provided in vCard form, Google Drive documents provided in Microsoft Office associated formats). Although not stated, a similar standard for extracted data may be adopted for law enforcement, but this would need to be established on a case by case basis.

4.3 Location and cooperation

Location has an impact on communication and cooperation with the provision in question. Of the 22 platforms analysed, 20 are based in the United States (US) and are governed by US law, with ASKfm is based in Dublin and Badoo in UK. Therefore those operating outside of these locations must seek cross-jurisdictional compliance with these organisations which may not in all circumstances be straightforward, and costly in terms of time and money. This can sometimes be achieved through Mutual Legal Assistance (MLA), which ‘is a method of cooperation between states for obtaining assistance in the investigation or prosecution of criminal offences. MLA is generally used for obtaining material that cannot be obtained on a police cooperation basis, particularly enquiries that require coercive means’ (Gov.uk, 2016). However, cooperation with social network provision is not guaranteed, a problem which has become a somewhat controversial matter of discussion in recent years, centering around arguments of freedom of speech, privacy and alleged government spying regimes. For example, The Select Committee on Communications (2014) indicated that recently, French authorities had to endure a lengthy court battle with Twitter for the disclosure of account details for those posting of anti-Semitic tweets. Twitter’s (2016e) Transparency Report surrounding received information requests, indicates that compliance can vary. The United Kingdom (UK), Japan and US form the 3 current biggest requesters of data from Twitter. From statistics reported from January to June 2016, information was supplied by Twitter in 82% of requests from the US, 61% from Japan and 76% from the UK. When taken in context, 453 requests for information from the US were unsuccessful, 282 from Japan and 151 from the UK. In comparison, Facebook’s (2016h) Government Request Report shows the US, India and UK to be the biggest requesters. From statistics reported from July to December 2015, information was supplied by Facebook in 81% of requests from the US, 51% from India and 82% from UK. When taken in context, 3654 requests for information from the US were unsuccessful, 2724 from India and 754 from the UK. In these circumstances, potentially evidential information may have been withheld, potentially prohibiting an investigation.

With a focus on developments within the United Kingdom, the recent enactment of the Investigatory Powers Act 2016 and its potential impact on social media investigations must be considered. Often termed the ‘Snoopers Charter’, the Investigatory Powers Act 2016 (IPA16) received royal assent on the 29th November 2016. Despite receiving criticism for implementing various powers considered by some as being for the purposes of mass surveillance through bulk data retention, the IPA16 came into force on December 30th 2016 and its many powers include the facilitation of the preservation of Internet Connection Records (ICRs) for 12 months by telecommunications service. ICRs are defined under section 62(7) IPA16 as records of the visits made to online services and websites by a user, used to examine where a user has been online. This is elaborated by the Home Office's (2016a, p17) Communications Data Draft Code of Practice stating ‘an ICR will only identify the service that a customer has been using. It is not intended to show what a customer has been doing on that service’. Yet, the bulk maintenance of ICRs by Internet Service Providers may only provide limited assistance in social media investigations (by potentially determining if a user has accessed a particular platform) due to the fundamental limitations imposed by an ICR, particularly in relation to the user of mobile device social media applications. This was reported by Adrian Kennard (2015), Managing Director Internet Service Provider Andrews & Arnold Ltd to the Joint Committee on the Draft Investigatory Powers Bill who stated that even if the mobile provider was even able to tell that she had used twitter at all (which is not as easy as it sounds), it would show that the phone had been connected to twitter 24 h a day, and probably Facebook as well. This is because the very nature of messaging and social media applications is that they stay connected so that they can quickly alert you to messages, calls, or amusing cat videos, without any delay'. Despite this being an issue with retained service provider information, the IPA16s deliberately wide definition of ‘telecommunications service’ still may provide some assistance to law enforcement. The Home Office’s (2016b, p6) Interception of Communications Draft Code of Practice indicates that ‘Internet based services such as web-based email, messaging applications and cloud-based services are, therefore, covered by this definition’, where Smith (2016) indicates that social media services would be incorporated. In addition, the powers of the IPA16 extend to ‘an operator outside the UK who provides a telecommunications service to people in the UK, or controls a telecommunication system in the UK’ (Smith, 2016). Providers are under an obligation to take reasonable practicable steps to give effect to any served warrant (see section 43 IPA16), regardless of their geographical location (Stringer, 2017). Therefore powers under the IPA16 are potentially enforceable via injunction against foreign social media providers (noted previously in Section “Location and cooperation” in relation to geographical location of many providers) subject to potential conflict with any foreign jurisdictional conflict (Smith, 2016). In such instances, social media providers may be required not only to maintain information regarding accesses to their platforms, but in turn, disclose this information subject to the correct legal processes being followed.

5 Concluding thoughts

As shown above, there are two main areas which currently pose an issue in the regulation of social network crime, account validation and retention policies. First, account validation remains a clear issue and is unlikely to be addressed soon. Validated accounts would not
only allow effective identification but also act as a deterrent to individuals as the perceived anonymity offered by the Internet is partially combatted. Yet as already highlighted, the processes needed to achieve this are unlikely to be favourably adopted by social network services. Second, data retention also remains an issue, and one which is likely going to get worse, largely due to the significant volumes of traffic witnessed across these platforms. At present, the task of retaining data is onerous and in absence of legal regulations forcing retention, the task of doing so is simply voluntary. Retaining all data indefinitely is not feasible and not necessarily needed, and to some extend, retention is already occurring. However, to improve current processes, clarity, transparency and consistency of retention processes is needed, an issue shown in Table 1. It is proposed that the global adoption of a single standard for retaining data in terms of length of time and type is required for all social network services, yet the feasibility of achieving this without legal regulation is debatable.

Ideally, IP logs should be maintained for long enough for this information to support investigatory processes, yet arbitrarily defining a period of time is not effective. Instead, collaboration between platforms must take place with law enforcement organisations to ensure that the data is retained for long enough to cover the period from first reporting of a crime by a victim up until legal processes are suitably activated to secure the disclosure of information. Preservation orders (as discussed above) provide some support but it is necessary to scrutinise further whether the commonly define period of 90 days is long enough to meet the required standard for disclosure of information from the service in question. Transparency in data retention times may also serve as a deterrent to those utilising the platforms. Where clear guidance is given regarding how long user actions are maintained, there may be a greater chance of potential offenders not breaching regulations in fear of detection. Achieving satisfactory retention of data must be balanced against the right for privacy and freedom of speak, a debate which is likely to continue for some time.

References


Association of School and College Leaders (2016) ‘School leaders voice concerns over children’s mental health care’ Available at: http://www.ascl.org.uk/news-and-
views/news_news-detail.school-leaders-voice-concerns-over-children-s-mental-health-care.html


Bebo (2016b) Terms of Service’ Available at: https://bebo.com/legal/terms (Accessed 10 September 2016)

Brunty, Kevin and Helenek, Katherine (2014) ‘Social Media Investigation for Law Enforcement’ Routledge,

Buzzfeed (2016a) ‘Privacy Policy’ Available at: https://www.buzzfeed.com/about/privacy (Accessed 26 September 2016)


Crown Prosecution Service (n.d.) ‘Guidelines on prosecuting cases involving communications sent via social media’ Available at:
(Accessed 26 September 2016)


Facebook (2016a) ‘Terms of Service’ Available at: https://www.facebook.com/terms (Accessed 10 September 2016)


Facebook (2016c) ‘What's the difference between deactivating and deleting my account?’ Available at: https://www.facebook.com/help/125338004213029 (Accessed 10 September 2016)


Facebook (2016f) ‘What names are allowed on Facebook?’ Available at: https://www.facebook.com/help/112146705538576?helpref=faq_content (Accessed 23 September 2016)


HC Deb 29 June 2016, Volume 612, cols 326-339


Snapchat (2016c) ‘When does Snapchat delete Snaps and Chats?’ Available at: https://support.snapchat.com/en-GB/article/when-are-snaps-chats-deleted (Accessed 13 September 2016)


Twitter (2016a) ‘Deactivating your account’ Avaiable at: https://support.twitter.com/articles/15358 (Accessed 10 September 2016)

Twitter (2016b) ‘Guidelines for law enforcement’ Available at: https://support.twitter.com/articles/41949 (Accessed 13 September 2016)


Twitter (2016d) ‘Reporting abusive behavior’ Available at: https://support.twitter.com/articles/20169998 (Accessed 13 September 2016)

Vine (2016a) ‘Vine Privacy Policy’ Available at: https://vine.co/privacy (Accessed 13 September 2016)

Vimeo (2016a) ‘Terms of Service’ Available at: https://vimeo.com/terms (Accessed 13 September 2016)


Westyorkshire Police (n.d.) ‘Offensive messages and posts on social media’ Available at: https://www.westyorkshire.police.uk/contact-us/social-media-sites/offensive-messages-and-posts-social-media (Accessed 10 September 2016)


Wordpress (2016a) ‘Terms of Service’ Available at: https://en.wordpress.com/tos/ (Accessed 26 September 2016)

