SOCIAL
DATA
SCIENCE
LAB.

CARDIFF
UNIVERSITY

PRIFYSGOL
CAERDYⱭⱭ

Ariennir gan
Lywodraeth Cymru
Funded by
Welsh Government

Hate Crime and Bullying in the Age of Social Media

Conference Report

Dr. Matthew Williams &
Dr. Olivia Pearson

Contents

1. Overview

1.1   The All Wales Hate Crime Project (Williams & Tregidga 2013, 2014) highlighted the emerging problem of cyberhate and cyber bulling via social media through interviews with victims. Opportunities for online engagement have increased exponentially over the past two decades. In 1999 only 10 per cent of UK households had access to the Internet. The number had grown to 53 percent in 2005 and to 85 per cent in 2015 (ONS 2015). Estimates put global social media membership at approximately 2.5 billion non-unique users, with Facebook, Google+ and Twitter accounting for over half of these (Sloan et al. 2015, 2015, Williams et al. 2016). Open and widely accessible social media technologies, such as Twitter and Facebook, are increasingly being used by citizens on a global scale to publish online content. The diffusion of information in these networks can manifest itself in a number of ways, ranging from the positive, such as support of social resilience through calls for assistance and advice (Morell et al. 2011), to the negative, through the production and contagion of misinformation and antagonistic and prejudiced commentary (Burnap et al. 2013, 2014, Williams et al. 2013).

1.2   Hate Crime and its commission online is now recognised as a priority by the UK Government. The sending of menacing messages via the Internet is now punishable by up to 2 years imprisonment (Malicious Communications Act 1998 as amended by the Criminal Justice and Courts Bill 2015). The Welsh Government continues to implement 'Tackling Hate Crimes and Incidents: A Framework for Action' and the fast paced evolution of social media is providing significant challenges for partners and agencies. Despite this recognition, practitioners can remain in the dark about the nature, prevalence and resources available to tackle cyberhate and bulling on social media.

This conference aimed to address this knowledge gap via a series of keynote presentations from high-profile leaders in the field and via hands on workshops. This report outlines conference attendee experiences in relation to the current

barriers and potential solutions in the area of cyberhate and cyber bullying and puts forward national recommendations.

2.   Context

2.1   Cyberhate

2.1.1   Despite cyberhate[1] being evident from the beginning of the domestic Internet (initially with the launch of the Stormfront website in 1995[2]), it has only recently become identified as a social problem that requires addressing.   The prominence of the problem is linked to the recognition that online spaces, such as social media platforms now represent new public spaces where key aspects of civil society are played out (Mossberger et al. 2008).   Reflecting this the Crown Prosecution Service has issued guidance to police establishing online networks as 'public spaces' allowing for prosecution to be brought under the Public Order Act as well as the Malicious Communications Act (Crown Prosecution Service 2015).   The former UK Justice Secretary Chris Grayling announced plans in 2014 to increase the maximum sentence for online abusive and hateful content.   In 2015 the sending of menacing messages via the Internet became punishable by up to 2 years imprisonment (Malicious Communications Act 1998 as amended by the Criminal Justice and Courts Bill 2015).

2.2   Manifestation and Prevalence of Cyberhate

2.2.1   Cyberhate has manifested in online communications in various contexts since the Internet became popular amongst the general population in the mid 1990s (Williams 2006). Legal provisions in the UK include the Public Order Act of 1986, the Malicious Communications Act 1988, the Protection from Harassment Act 1997 and the Crime and Disorder Act 1998.   The application of these laws and others that criminalise incitement on the basis of religion and sexual orientation[3] to the online

---

[1] The practice of 'trolling' (the targeting of defamatory and antagonistic messages towards users of social media) has received press attention of late. We avoid using the term in this report as it can encapsulate broader forms of online abuse not restricted to victims with minority or protected characteristics.

*Matthew L. Williams, Social Data Science Lab, School of Social Sciences, Cardiff University, King Edward VII Ave., Cardiff CF10 3WT, UK; WilliamsM7@ cf.ac.uk.
[2]   Stormfront existed in bulletin board format in the early 1990s before being reformed as a website.
[3] Racial and Religious Hatred Act 2008 and the Criminal Justice and Immigration Act 2008.

context is relatively non-contentious as evidenced by several recent high profile social media cyberhate cases (see below).

2.2.2   Despite these provisions, for over a decade much of the hate speech that has manifested online (pre social media) met with little criminal justice response in the UK. Further afield, in countries like the US, cyberhate continues largely unchallenged by law enforcement due to freedom of speech protections. Levin (2002) studied how US right-wing groups promoted their goals on the Web largely unchallenged by law enforcement, concluding that the online medium has been useful to hatemongers because it is economic, far reaching and protected by the First Amendment. Perry and Olsson (2009) found that the Web created a new common space that fostered a 'collective identity' for previously fractured hate groups, strengthening their domestic presence in counties such as the US, Germany and Sweden. They warn a 'global racist subculture' could emerge if cyberhate is left unchallenged. Eichhorn (2001) focuses on how the online environment opens up the possibility for a more immediate and radical recontextualization of hate speech, while also highlighting its affordances for more effective modes of response, such as vigilantism and counter-speech. Leets (2001) in a study of the impacts of hate related web-pages found that respondents perceived the content of these sites as having an indirect but insidious threat, while Oksanen et al. (2014) show how 67 per cent of 15 to 18 year olds in their study had been exposed to hate material on Facebook and YouTube, with 21 per cent becoming victims of such material. This final study evidences how the rise of social media platforms has been accompanied by an exponential increase in cyberhate (see also Williams & Burnap 2015).

2.3   Cyberhate and Trigger Events

2.3.1   Research has shown that the prevalence and severity of crimes with a prejudicial component are influenced in the short term by singular or clusters of events. Most notably, Philips (1980) evidenced how the occurrence of the most severe of these crimes were influenced by widely media publicized homicides and

Bobo et al. (1994) showed how riots significantly influenced public perceptions of racial minorities. More recently, acts of terrorism have been shown to influence the prevalence of anti-immigrant sentiment and hate crimes and incidents. On a European scale Legewie (2013) established a significant association between anti-immigrant sentiment and the Bali and Madrid terrorist bombings using Eurobarometer data. Similarly, King and Sutton (2014) found an association between terrorist acts and a rise in hate crime incidents in the US. Convincingly, they show that following the 9/11 terrorist attack, law enforcement agencies recorded 481 hate crimes with a specific anti-Islamic motive, with 58 percent of these occurring within two weeks of the attack (4 percent of the at risk period of 12 months). In the UK Hanes and Machin (2014) found significant increases in hate crimes reported to the police in London following 9/11 and 7/7 (28 and 32 per cent increase in the month following respectively). These authors conclude hate crimes cluster in time and tend to increase, sometimes dramatically, in the aftermath of antecedent 'trigger' or galvanizing events, such as terrorist acts. They postulate that hate crimes are communicative acts, often provoked by events that incite retribution in the targeted group, towards the group that share similar characteristics to the perpetrators.

2.3.2   Williams and Burnap (2015) argued that following trigger events, such as terrorists acts, it is often social media users who are first to publish a reaction, and given there are now over 2.5 billion users of social media (Smith 2014) these online communications provide an insight into public opinion on an unprecedented scale at very quick notice (often to the second). Indeed, there is evidence to support this argument in the recent high profile prosecution of social media users who posted negative emotive reactions following various events. For example, in 2012, Liam Stacey was sentenced to 56 days in prison for posting racist comments on Twitter after a footballer's cardiac arrest and Daniel Thomas was arrested after a homophobic message was sent to Olympic diver Tom Daley. In 2014, Isabella Sorley, John Nimmo and Peter Nunn were jailed for abusing feminist campaigner Caroline Criado-Perez and MP Stella Creasy, and Declan McCuish was jailed for a year for tweeting racist comments about two Rangers Football Club players. In relation to

the Woolwich terrorist attack, seven social media users were arrested after posting messages that were suspected of inciting racial or religious hatred (BBC 2013).

2.3.3   In their research Williams & Burnap (2015) developed an automated cyberhate classification tool to identify hate speech originating from individual Twitter users following the Woolwich, UK terrorist attack in 2013 (see also Burnap and Williams 2015). They found that those identifying with right wing political groups were most likely to produce hateful content on Twitter following the attack. Like offline hate, cyberhate was shown to spike and rapidly decline within the first 48 hours of the attack, indicating that cyberhate has a half-life. They conclude social media acts as a force-amplifier for cyberhate as it can open up a potential space for the rapid galvanising and spread of hostile beliefs, via the spread of rumours through online contagion.

2.4   *National and International Responses*

2.4.1   Until very recently, the regulation of cyberhate by the state has been minimal. The slow international response to regulate cyberhate is, in part, due to the now outdated view that offensive online communication is less harmful than offline equivalents. For over a decade, 'virtual' online environments and the experiences had within them were considered separate to 'real' offline spaces. This separation resulted in many, including the public, law enforcement and policy makers, to consider what happened in virtual spaces as somehow not serious, even game-like and without consequence (Joinson 2003, Williams 2006). This form of online 'harm' therefore went unrecognised, considered as a 'petty' *de minimis*[4] phenomenon with no place in the policing crime diet (Hillyard et al. 2004, Wall & Williams 2007, Williams et al. 2013).

2.4.2   In England and Wales, provision comes under a range of legislation including the Malicious Communications Act (1998), the Protection from Harassment Act (1997) and the incitement provisions under of the Public Order Act (1986) that states

---

[4] From: *de minimis non curat lex*: The law does not concern itself with trifles

"A person who uses threatening, abusive or insulting words or behaviour, or displays any written material which is threatening, abusive or insulting is guilty…if (a) he intends thereby to stir up racial hatred, or (b) having regard to all the circumstances racial hatred is likely to he stirred up thereby". Further afield, the EU Framework Decision on Racism and Xenophobia (2008) and the Organisation of Security and Co-operation in Europe Ministerial Council Decision 9/09 on Combating Hate Crimes provide similar protections on stirring up hatred and violence online.

2.4.3   Despite the recent spate of arrests and prosecutions for acts of cyberhate in social media under the various provisions in England and Wales, there remain inadequate criminal justice resources to police the volume of communications that travel through computer networks. This prevalence versus capacity problem results in a loss in confidence from victims of cyberhate and the communities within which they live (Gianassi 2014). These capacity issues are complicated by the pace of change inherent in communications technologies, and by a fragmented and uncooperative industry. Historically, the social media giants (Google, Facebook & Twitter) were reluctant to come together to help law enforcement tackle the growing problem, given the ferociously competitive nature of their industry and their philosophical position on free speech. For the first time representatives from some of these companies came together with politicians and academics at a meeting of the Inter-parliamentary Coalition for Combating Anti-Semitism (ICCA) Task Force on Internet Hate at Stanford University in 2013. At the meeting they established that the location of most hate content is on U.S. servers, that it is extremely difficult to respond to cyberhate due to scale and definition, that the failure of cross-border law enforcement and civil actions preclude any meaningful change in the amount and intensity of cyber hate and that the ever-changing technology makes cross-border law   enforcement and civil actions significantly more difficult (ICCA 2013). Resulting from this meeting the Anti-Defamation League and several major Internet companies established 'Best Practices for Responding to Cyberhate' (ADL 2014) that recommend timely and proportionate responses from social media providers, and for the Internet community to explore avenues for counter-speech as a viable alternative to criminal sanctions.

2.4.4　These best practices include[5]:

Providers:

- Providers should take reports about cyberhate seriously, mindful of the fundamental principles of free expression, human dignity, personal safety and respect for the rule of law.
- Providers that feature user-generated content should offer users a clear explanation of their approach to evaluating and resolving reports of hateful content, highlighting their relevant terms of service.
- Providers should offer user-friendly mechanisms and procedures for reporting hateful content.
- Providers should respond to user reports in a timely manner.
- Providers should enforce whatever sanctions their terms of service contemplate in a consistent and fair manner.

Internet Community:

- The Internet Community should work together to address the harmful consequences of online hatred.
- The Internet Community should identify, implement and/or encourage effective strategies of counter-speech — including direct response; comedy and satire when appropriate; or simply setting the record straight.
- The Internet Community should share knowledge and help develop educational materials and programs that encourage critical thinking in both proactive and reactive online activity.
- The Internet Community should encourage other interested parties to help raise awareness of the problem of cyberhate and the urgent need to address it.
- The Internet Community should welcome new thinking and new initiatives to promote a civil online environment.

---

[5] http://www.adl.org/combating-hate/cyber-safety/best-practices/#.Vo5nMcZ4zHM

2.5   Cyber Bullying

2.5.1   Cyber bullying is a form of bullying which occurs online; through social networking sites, gaming or chat rooms or through mobile phone and tablets. Cyber bullying takes many forms. It can include:

- Harassment or trolling: sending threatening or offensive messages, sharing embarrassing photos and videos or posting upsetting or threatening messages on social networking sites;
- Denigration: fake untrue information to spread rumours;
- Flaming: extreme language to cause a fight;
- Stealing someone's identity or hacking into someone's site;
- Exclusion: intentionally leaving someone out;
- Sending explicit pictures or pressuring others to send sexual images.

2.5.2   In contrast to traditional forms of bullying it can often be much more difficult to escape cyber bulling as it can occur 24 hours a day, 7 days a week and reach victims in their own homes. It can also have a far greater reach, as videos and posts being shared across social networking sites can be seen by large audiences. Cyber bullying may be an extension of traditional offline bullying with bullies being known to the victim, or it can be anonymous and unrelated to known acquaintances offline.

2.6   Prevalence and Patterns

2.6.1   Cyber bullying is fast becoming an area of concern. Many research studies have investigated the prevalence of cyber bullying, particularly among young people. Estimates of the prevalence of cyber bullying victimization and perpetration vary widely between studies. This is because of methodological inconsistencies and definitional differences in the studies that have been conducted.

Figures from leading anti-bullying charities in the UK suggest that most young people will either experience cyber bullying as victims, perpetrators or bystanders at some point. In a recent online survey Bullying UK found that 56% of young people said

they have seen others bullied online (Bullying UK 2014). A systematic review of international research into cyber bullying found that prevalence rates of victimisation ranged from 4-72% (Nixon 2014). By comparison, studies in the UK have typically identified rates of victimization between 15 and 28% (Ditch the Label 2013, 2014, 2015, IpsosMori 2014, Cross et al. 2012).

2.6.2   Anti-bullying charity Ditch the Label's 2013 international Cyber Bullying Report drew on the largest bullying-related dataset of over 10,000 young people. It estimated that 5.43 million young people in the UK have experienced cyber bullying with 1.26 million subjected to extreme cyber bullying on a daily basis. Across the entire survey 7 out of 10 young people had been victims of cyber bullying and 37% had experienced it on a highly frequent basis. Beatbullying's 'Second Virtual Violence Study' found that 20% of young people said that their experience was an extension of offline bullying, while 27% said that the bullying they had experienced had started online (Cross et al. 2012). Therefore, this would indicate that bullying is becoming an increasingly more common phenomenon that starts online.

2.6.3   Similar to more traditional forms of bullying, cyber bullying victimisation and perpetration are not mutually exclusive. A Canadian study (Mishna et al. 2012) found that 25.7% of young people identified themselves as both victims and perpetrators of cyber bullying over the past three months. In another study, when perpetrators were asked why they engaged in cyber bullying, 39% indicated their main motivation was for revenge, while 26% did it for a joke, 16% did it because they were 'angry about stuff' and 15% did it because they were bored (Cross 2012).

2.6.4   Studies have also identified the platforms and channels through which cyber bullying takes place. A Ditch the Label (2014) report showed that Facebook, Twitter and Ask.FM were the most common social networks for cyber bullying. It also found that smartphone apps were a prominent source of cyber bullying. Nearly two thirds

(62%) of people who had been victims of cyber bullying had been sent 'nasty private messages' through this medium.

2.7   Impacts of Cyber Bullying

2.7.1   Cyber bullying can have significant negative effects on people's lives. As with more traditional forms of bullying, it can cause psychological, emotional and physical distress. Research into the impacts of cyber bullying has commonly identified the following factors (although this list is not exhaustive):

- *Depression:* Raskaukas and Stoltz (2007) identified that 93% of victims reported negative effects of cyber bullying, many of which were related to depression, including sadness, hopelessness and powerlessness. Other studies have shown a connection between higher levels of victimization and higher levels of depression.
- *Loss of confidence:* Cyber bullying can have negative effects on victims' self-esteem. Beatbullying found that 19% of respondents said they had reduced self-confidence following cyber bullying incidents (Cross et al. 2012).
- *Fear:* In Finland Sourander et al. (2010) found that victims of cyber bullying feared for their safety more so than with 'traditional' forms of bullying. It was suggested that this was due to the anonymity associated with cyber bullying. Beatbullying's survey identified that 20% of young people were reluctant to go to school as a result of fear of bullies (Cross et al. 2012).
- *Isolation and relationship problems:* Cyber bullying can have a variety of impacts on victims' friendships and relationships. Victims have reported feelings of loneliness, isolation, helplessness and fewer or disrupted peer relationships (Nixon 2014).
- *Self-harming:* In some cases cyber bullying can lead to people self-harming. Beatbullying found that 5% of young people had self-harmed as a result of cyber bullying victimization (Cross et al. 2012). They also suggested that the risk of self-harm is intensified through prolonged and deliberate targeting of victims.

- *Suicide:* Research suggests that *both* victims and perpetrators are more likely to have had suicidal thoughts or attempted suicide than peers not involved in cyber bullying. Perpetrators of cyber bullying are 1.5 times more likely to have attempted suicide than non-involved peers, while for victims this is twice as likely. Beatbullying also reported that 3% of young people have attempted suicide as a result of cyber bullying (Cross et al. 2012).

2.8   Reporting and Responding to Cyber Bullying

2.8.1   With these figures in mind, it is clear that providing young people support to deal with incidences of cyber bullying and encouraging reporting of these is vital. However research suggests that there is still much to be done in respect of reporting and responding to cyber bullying.

2.8.2   Ditch the Label found that 52% of young people have never reported abuse that has occurred through smartphone apps. When they have reported incidences 26% felt like a report was not taken seriously. One reason why young people may be averse to reporting cyber bullying online, therefore, is that they feel it will not be acted upon. This was indicated in a recent study by Bullying UK. They found that when reporting bullying to a social network site, only 8.8% of respondents said that the network took any action. Parents have also been found to be reluctant to report incidences of cyber bullying. Only 30% of parents have reported bullying online and in these cases only 50% of social networks responded (Bullying UK 2014).

2.9   National and International Responses

2.9.1   There is currently little in the way of a formal national or international response to dealing with the problem of cyber bullying. Anti-bullying websites, national /international charities and research studies tend to offer some advice and recommendations for preventing and responding to cyber bullying. There has also been some recent published advice from government departments in England and

Wales. There are currently no existing legislative frameworks specifically designed to tackle cyber bullying although a number of laws in the UK do apply (see below).

International Advice and Prevention

2.9.2   National and international recommended responses to cyber bullying suggest prevention in the first instance, through educating students, school staff and parents about Internet safety. It is suggested that schools build awareness of these issues into curriculum education, particularly with the growing involvement of young people with digital technologies.

2.9.3   Advice from both North American and UK organisations[6] (Hindula 2014) recommend that young people should use privacy settings built into social networking sites to protect themselves from the risk of cyber bullying. In the event of bullying online, young people should report the incident to an adult. They should keep any evidence of cyber bullying that they can and not respond or retaliate to cyber bullies. Advice is also issued to parents. StopCyberBullying tell adults to be aware of what young people are doing online, to establish rules about technology use and for parents to talk to their children about Internet use. If cyber bullying does occur, adults are encouraged to report this to Internet Service Providers (ISPs), to schools and to social media sites if appropriate. If the cyber bullying includes instances that are considered criminal (e.g. threats of violence, receiving explicit photos, being stalked or receiving racial or homophobic hate messages), adults are also recommended to take legal action.

National Legislation, Policy and Guidance (England and Wales)

2.9.4   In 2014 the Department for Education published two guidance documents (DoE 2014, 2015) for schools relating to the safeguarding of children and policies and processes for responding to bullying (including cyber bullying). These included legal

---

[6] ChildLine (www.childline.org.uk); Bullying UK (www.bullying.co.uk); Stop Cyber Bullying (http://www.stopcyberbullying.org)

duties that must be complied with, the government's approach to bullying and the powers that schools and colleges have to deal with bullying. The Welsh Government has issued their own guidance (Welsh Government 2011)[7] on bullying in schools, which includes advice and practical solutions on preventing and responding to bullying.

2.9.5   Schools across England and Wales are required under The School Standards and Framework Act 1998 and the Education (Independent Schools Standards) Regulations 2003 to have anti-bullying policies and processes which should include provisions for dealing with cyber bullying. Other legislative frameworks including the Education Act 2011, the Disability Discrimination Act 2005, and the Equality 2010 Act also place duties on schools and other public bodies to protect students from bullying, discrimination and harassment. The Education Act 2011 also allows teachers to search electronic devices for evidence of bullying or threatening behavior between students.

2.9.6   Cyber bullying is not a specific criminal offence in the UK. However, a range of bullying activities such as harassment and threatening behavior, sending offensive emails or other messages that are indecent or obscene and making anonymous or abusive telephone calls are considered criminal under various legislation: the Protection from Harassment Act 1997; the Crime and Disorder Act 1998; the Telecommunications Act 1984; the Harassment Act 1997; the Malicious Communications Act 1988 and; the Communications Act 2003.

2.9.7   Recently the UK Council for Child Internet Safety published the report Child Safety Online: A Practical Guide for Providers of Social Media and Interactive Services (2015). The guidance includes:

Managing Content on Your Service

- Decide what content is acceptable on your service, and how you'll make this clear to users.

- Be clear on minimum age limits, and discourage those who are too young.

- Consider different default protections for accounts that are opened by under 18s.

- Plan and regularly update how you'll manage inappropriate or illegal content posted on your site.

- Consider using available age verification and identity authentication solutions.

- Plan now for dealing with illegal content.

- For under-13s, consider a walled garden environment and pre-moderating content before users see it. Also become familiar with the UK rules to advertising to children.

Parental Controls

- Consider parental controls that are designed for your service.

- Be aware how different parental controls might interact with your website or app.

Dealing with Abuse/Misuse

- Explain to users the type of behaviour you do and don't allow on your service.

- Make it easy for users to report problem content to you.

- Create a triage system to deal with content reports.

- Work with experts to give users additional information and local support.

- For under-13s, talk in their language, and pre- and post-moderate their content.

Dealing with Child Sexual Abuse Content and Illegal Contact

- Give your users a standardised function for them to report child sexual abuse content and illegal sexual contact.

- Have a specialist team, who are themselves supported, to review these reports.

- Consider technology such as PhotoDNA and working with relevant bodies such as the Internet Watch Foundation (IWF) to help remove child sexual abuse content.

- Escalate reports of child sexual abuse content and illegal sexual contact to the appropriate channel for investigation.

- Tell users how they can report child sexual abuse content or illegal sexual contact directly to the relevant authorities and/or where to obtain further advice.

Privacy and Controls

- Only collect the personal data you actually need for your service.

- Tell users what information you collect, why and how long you'll keep it.

- Give users reasonable choices about how to use their personal information and specific types of data, such as geolocation data.

- Offer privacy settings options, including privacy-by-default, to give control to your users.

- Involve parents/guardians if you collect personal data from under-18s.

- For under-13s, have stricter privacy measures to help them understand the implications of sharing information.

Education and Awareness

- Educate users about safety as part of the experience on your platform.

- Work with parents, educators, users and their communities to raise awareness about online child safety.

- Work with experts to help develop your messages and to reach different communities.

- For under-13s, tailor the language and approach so they will take an interest.

3.   Conference Findings

The conference was opened by Lesley Griffiths AM, the Minister for Communities and Tackling Poverty. Keynotes were delivered on topics, including: Online Safety for Children and Challenges, Claire Lilley, Head of Child Safety Online, NSPCC; Cyberhate: International and UK Contexts, Paul Giannasi, Ministry of Justice; Hate Crime Challenges for Policing across Wales, Rt Hon Alun Michael, Police and Crime Commissioner for South Wales; and Challenges and Responses for Social Media, Nick Pickles, Head of Policy Twitter UK.   Eight workshops led by experts elicited perceptions on the barriers and solutions to tackling cyberhate and cyber bullying on social media from conference participants.

3.1   Cyberhate

3.1.1   Four conference workshops focussed upon identifying the barriers and solutions to tackling cyberhate on social media networks: Terrorist Incidents and the Propagation of Cyberhate (Dr Matt Williams, Cardiff University); Cyber Hate: Criminal Justice and the Law – Prosecuting Online Hate (Paul Giannasi, Ministry of Justice and Mike Whine, Community Security Trust); Don't Feed the Trolls: Responses and Challenges of On-line Behaviour (Dr Amy Binns, University of Central Lancashire); and Islamaphobia and the Far Right: Social Media Patterns and Challenges (Fiyaz Mughal, Tell MAMA). The key themes identified from across these workshops are presented below.

3.1.2   Barriers

- Evidence and data on cyberhate is poor. Anecdotal evidence points to an emerging problem, but there are no official sources to corroborate this. Without solid evidence of the scale and nature of the problem we cannot develop practical and policy solutions.

- Little is currently known about the impacts of cyberhate and how these compare to offline hate crimes.

- Cyberhate has been shown to emerge following 'trigger' events such as terrorist attacks, but the technology to identify hate speech on social media networks is not widely adopted by authorities charged with protecting the public.

- Victims and offenders need better information sources on the law pertaining to cyberhate in England and Wales.

- Counter-hate-speech is being used by online community members to stem the spread of cyberhate, but in some cases the forms of counter-speech can create a cycle of hate, worsening the problem.

3.1.3   Solutions and Examples of Best Practice

- The Crown Prosecution Service now recognise the internet as a 'public space' allowing the use of the public order offences under the Public Order Act 1986.

- True vision is a web facility that is maintained by the Association of Chief Police Officers. It was launched in 2011 and provides an online platform for the reporting of hate crimes and provides information for victims and advocates. It contains official strategies and policies that guide police and partners about how to respond to incidences of hate, what happens when you report a hate crime, personal safety tip and organisations, which can offer support. The site also offers up to date hate crime data and reports. In its first year of operation the site was visited more than 6,000 times (Gianassi 2012). In 2011/2012 1,900 reports of hate crime where made to true vision by the public, many of which had not been reported directly to the police. This figure has risen continuously with 2,957 incidences being reported in 2012/2013 and 3,461 in 2013-2014 (Home Office 2014). In 2013 a true vision mobile phone app was launched to support the website. It is free to download and was downloaded more then 3,000 time between the periods of 2013-2014 (Home Office 2014).

- A Welsh language version of True Vision is available. In 2014/15 91 reports were made via the True Vision website that were referred to one of the four Welsh forces, compared to 66 reports in 2013/14.

- Academic research examining the production and spread of cyberhate, led by the Social Data Science Lab at Cardiff University, has produced an automated cyberhate detection tool for Twitter communications. This tool can be integrated into computer systems within organisations that have a responsibility for protecting the public (e.g. police, pubic sector and third sector organisations). This tool allows for the monitoring of cyberhate on social media, in particular following 'trigger' event. The ability to identify and monitor hateful content online allows action to be taken to prevent the spread of harmful or antagonistic content. The Social Data Science Lab is currently working with the Community Security Trust in London to integrate the tool into their systems to monitor the production and spread of anti-Semitic content on Twitter.

- The Social Data Science Lab is also conducting research into the effective use of counter-hate-speech by online community members. Initial results indicate that while it can be useful in stemming the spread of cyberhate, it can also worsen situations if not used appropriately.

3.2   Cyber Bullying

3.2.1   Four conference workshops focussed upon identifying the barriers and solutions to tackling cyber bullying on social media networks: Online Safety and Practice in the UK (Andrew Williams and David Wright, South West Grid for Learning); Online Behaviour: How Young People use Social Media (Rachel Benson, Youth Cymru); Cyber Safety for Parents and Children (Andrew Williams and David Wright, South West Grid for Learning); and Revenge Porn and Sexting: What you

need to know (Shereen Williams, Gwent-East Community Cohesion Co-ordinator). The key themes identified from across these workshops are presented below.

3.2.2   Barriers

- A skills barrier exists in relation to social media technology amongst some teaching staff in Wales.

- Teachers are not being allowed the time to be trained in how to use these technologies and how to best to prevent cyber-bullying.

- Workshop participants highlighted that some parents were more hard-to-reach than others in relation to communicating e-safety messages.

- There is a lack of consistency between schools in relation to the provision of e-safety training for teachers and instruction for students.

3.2.3   Solutions and Examples of Best Practice

- The Welsh Government e-Safety Zone was highlighted as a good example of best practice from government. Conference participants felt that further promotion campaigns would improve the resource's reach and use. The Welsh E-safety zone is a partnership between the Welsh Government and the South West Grid for Learning, which is part of the UK Safer Internet Centre. The e-safety Zone aim's to promote safe and responsible use of the Internet by all. The project was based on a needs assessment survey undertaken by the South West Grid for Learning and Plymouth University (Phippen 2014). The survey sought to define the landscape of Wales and the findings and recommendations of the report provided the foundation that underpins the work of the e-safety partnership (Phippen 2014). The e-safety zone provides a variety of resources, links, advice and support for children and young people, parents/carers and education professionals via its online platform the Hwb e-saftey zone. Resources included a

self-review e-safety online tool adapted specifically for the use of Welsh schools and digital literacy resources that guides teachers on how to embed digital literacy skills in their learners through all Key Stages.

- The e-safety partnership also offers a number of training sessions. They are currently offering 2 days of e-Safety training and support for every local authority in Wales. The events are specifically aimed at schools and school staff and local authorities safeguarding teams.

- The O[2] Partnership with NSPCC (www.o2.co.uk/nspcc) was identified as an example of good practice. The partnership's aims is to give parents the information and support they need to keep children safe online. The partnership includes; 1) A free online safety helpline for parents to call for technical advice (such as includes setting up parental controls, adjusting privacy setting and understanding social networking sites; 2) Online safety workshops for parents and carers into schools and workplaces, and; 3) Trained staff so all O2 stores can help adults with their online safety concerns. The partnership was formed in 2015 in response to new research, which found that thousands of children were not receiving necessary advice and support about the Internet at a time when they need it most. A survey conducted by YouGov highlighted a 'digital delay' where parents were leaving it to late to have vital conversations with their children about how to stay safe online. The partnership is designed to help parents and guardians see the Internet as their children do and understand its real dangers.

- The Welsh Network of Healthy Schools Scheme was highlighted as an example of best practice that might be used to promote e-safety messages and training. The Network (WNHSS) was launched in 1999 and was designed to encourage the creation of healthy schools within a national framework. The WNHSS is funded by the Welsh Government and Public Health Wales and provides guidance to health and education services in setting up healthy school schemes in the 22 local authorities across Wales. The aim of the healthy school schemes is to promote and protect physical, emotional and social health and well being of

children in Wales.   The WNHSS encourages schools to work across seven different topics. These include: Food and Fitness; Mental and Emotional Health and Wellbeing; Personal Development and Relationships; Substance use and misuse; Environment; Safety, and; Hygiene.   Every Healthy Schools Scheme in Wales is working towards achieving the same set of aims. However, a key component of the scheme is the autonomy it provides schools in setting their own priorities and agendas. As the scheme encourages pupils to be involved in the planning and implementation of healthy school action the methods used to achieve the schemes aims might differ between schools. A National Quality Award of Excellence exists to award schools that have achieved the highest standards in all Healthy Schools topic areas. Schools are required to provide evidence that they have achieved all indicators for each of the seven Healthy Schools topic areas and demonstrate a whole school approach in meeting these aims.

- The Eyst 'Think Project[8]' was identified as a good example of reducing racism on social media.   It is a 3-year Big Lottery funded project designed to challenge racism and far right extremism in vulnerable young people in Wales. The project aims to provide disengaged young people in Wales with greater knowledge and understanding about race, religion, immigration asylum and extremism so that they can challenge and deconstruct racist views.   A key objective is to listen to young people and enable them to express their real and perceived concerns, increase their confidence in their own sense of belonging and build resilience to extremist ideologies.   The Think Project follows on from the pilot project funded by the Welsh Government Community Cohesion Fund.   In 2011 the pilot project evaluation found that young people believed that they had increased their knowledge of different races, religions and cultures (iworks 2013). Before taking part in the Think Project the young people largely held negative views. Most of the young people believed that their communities did not like people from racial and religious groups. After taking part in the Think Project, the majority of the

---

[8] http://eyst.org.uk/think-project/

young people expressed increased levels of understanding and empathy for different groups. Young people felt more confident in expressing themselves particularly in the presence of people from Black Minority Ethnic backgrounds. In regards to the long term impact of the project, the report noted that attitude change was difficult to measure and that 'attitudes and beliefs are influenced by a wide range of social factors and one single project will not fix all social problems.' However they acknowledged that increasing awareness and education is an important practical step in developing greater cohesion and resilience and the project provides a useful model in the process of formal and informal education.

- The Crown Prosecution Service Hate Crime Schools Project[9] resources were being used by some workshop participants, but some noted their use was far from uniform. The CPS has developed free resource packs for teachers so that they can help students explore issues about hate crime and bullying. There are currently a variety of different packs available. These included a set of resources to help teachers explore: disability hate; racist and religious hate crime, and homophobic and transphobic bullying and hate crime. The resource pack contains a DVD, which provides scenarios based on real life incidents that young people have experienced. It also includes lesson plans, which offer suggestions for classroom activities, designed to develop young people's knowledge and understanding of hate crime and the impact of this type of bullying on victims. The resources are free to download (or order) and have been developed in collaboration with education, criminal justice and voluntary sectors. These included the National Union of Teachers, The Ministry of Justice, Stonewall, Anthony Walker Foundation and Gender Intelligence. A number of young people from Sheena Amnos Youth Trust, University of Central Lancaster, Dame Elizabeth Cadbury Technology College, LBT North West and a school in the North West, have also helped create and act out scenarios based on real life experiences for the resources.

---

[9] http://www.cps.gov.uk/northwest/working_with_you/hate_crime_schools_project/

- The All Wales School Liaison Core Programme (schoolbeat.org) was highlighted as an example of best practice by some conference participants. The All Wales School Liaison Core Programme (AWSCLP) is a crime prevention programme designed to educate young people about the dangers associated with a number of current issues such as substance use and misuse, bullying, anti-social behaviour, strangers, Internet safety, weapons, mobile phone usage and so on. The Welsh Government and the Welsh police forces jointly fund the programme. The main aims of the AWSCLP are to 1) work towards achieving a reduction in crime and disorder in the young of our communities, through the medium of education; and 2) promote the principles of positive citizenship in schools and their wider communities. In recognition of the role that schools play in educating young people about anti-social behavior, substance misuse and personal safety, the AWSCLP is focused on formal lesson delivery by 85 educationally trained police officers. Lessons are delivered during school hours in partnership with PSE teachers. The formal approach to the programme is designed to ensure that all children across Wales receive the same up to date information. In 2008-2010 an evaluation of 10 schools was undertaken. In relation to Internet safety the evaluation found that teachers and programme police officers reported that pupils acted upon the information shared in AWSLP lessons. They stated an example, that pupils were reporting suspicious behaviour on the Internet. The report also suggested that some pupils had asked for additional lessons on this topic.

- In March 2015 the Cyber Crime Week of Action Campaign was launched by the National Crime Agency working alongside national and international law enforcement and the education sector to respond to the threat of cyber-crime. Dyfed Powys Police has a specialist Digital Cyber and Cyber Unit which, alongside Get Safe Online is helping people and business stay safe online. As part of Cyber Crime Week of Action, they focused on providing guidance to people on how to stay safe online and reduce the risk of becoming victims of cyber crime. A number of events were organised, including a series of school assemblies

addressing issues with 'sexting', training for police officers and online events and a pop up shop in the town centre of Aberystwyth. A key focus of the week was on key cyber-crime issues including online child sexual exploitation, cyber bullying, revenge porn and online fraud through social media.

3.3 Initial Recommendations and Next Steps

3.3.1 A growing body of evidence indicates that incidents of cyberhate and cyber bullying are increasing in prevalence and are having a measurable impact upon the health and wellbeing of victims. Despite this recognition many practitioners remain unaware about the nature, prevalence and resources available to tackle cyberhate and bulling on social media. The conference workshops drew upon the experiences of the 100+ participants from across industry, public and third sectors to set out the current barriers and to identify potential solutions to draw together national recommendations. These recommendations are outlined below.

3.3.2 Cyberhate

- Better data on the scale, nature and impacts of cyberhate are needed to best target practical and policy initiatives. **To estimate the picture in Wales dedicated questions should be included in the National Survey for Wales** (The Crime Survey for England and Wales does not currently include questions on hate crime committed via the Internet. Furthermore, extrapolating to Wales from the CSEW is currently not possible given the low number of respondents (n=2600) and victims of hate).

- Cyberhate should become a standing item on Hate Crime Criminal Justice Boards in order to monitor and develop evidence sources and to advise and play a leading role on preventative campaigns.

- A national campaign on the legal protections available to victims of cyberhate should be launched in hate crime awareness week.

- A national guide on 'effectively using counter-hate-speech' should be developed for users of social media.

- The cyberhate automated detection tool for social media developed by the Social Data Science Lab should be trailed in an operational setting to gauge its usefulness in monitoring and responding to hate speech online.

3.3.3 Cyber Bullying

- Teaching and support staff in schools need to be released from duties for training on preventing cyber-bullying. Training should be sought from criminal justice agencies (potentially delivered thorough National Cyber Crime Awareness Week) and third sector organisations with expertise in online safety (such as the South West Grid for Learning).

- The introduction of an e-safety kite-mark system linked to Estyn inspection should be considered. Such a system would highlight schools that have implemented a series of protections for students and staff and would introduce a degree of consistency between schools.

- Partnerships should be established between organisations with responsibility for child welfare and mobile phone and social media companies, mirroring the O^2 NSPCC arrangement.

- A secure Cyber-Bullying Information Sharing Point should be established to facilitate inter-agency collaboration and best-practice exchange.

4.   References

BBC (2013) *Woolwich murder sparks anti-Muslim backlash*, Available at: http://www.bbc.co.uk/news/uk-22664835

Bullying UK (2014) National Bullying Survey 2014, retrieved from: http://www.familylives.org.uk/about/press/bullying-continues-to-impact-on-wellbeing-of-pupils-parents-and-professionals/

Burnap, P, Williams, M. L. & Sloan, L. (2014) 'Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack', *Social Network Analysis and Mining*, 4:206.

Burnap, P. & Williams, M. L. (2015) 'Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making', *Policy & Internet.*

Burnap, P., Williams, M. L. Rana, O., Edwards, A., Avis, N., Morgan, J., Housley, W. and Sloan, L.. (2013) Detecting tension in online communities with computational Twitter analysis, *Technological Forecasting & Social Change*.

Cross, E., Piggin, R., Douglas, T. and Vonkaenel-Flatt, J. (2012). *VIRTUAL VIOLENCE II: Progress and Challenges in the Fight against Cyberbullying*, London: Beatbullying/Nominate Trust.

Department of Education. (2014). Preventing and tackling bullying: advice for headteachers, staff and governing bodies. Retrieved from, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/374850/Cyberbullying_Advice_for_Headteachers_and_School_Staff_121114.pdf

Department of Education. (2015). Keeping children safe in education: statutory guidance for schools and colleges. Retrieved from https://www.gov.uk/government/publications/keeping-children-safe-in-education--2

Ditch the Label (2014). *The Annual Bullying Survey. Retrieved from http://www.ditchthelabel.org/annual-bullying-survey-2014/*

Ditch the Label (2014). *The Wireless report 2014.* Retrieved from http://www.ditchthelabel.org/the-wireless-report-2014/

Ditch the Label (2015). *The Annual Bullying Survey. Retrieved from http://www.ditchthelabel.org/the-annual-bullying-survey-2015-is-here/*

Ditch the Label. (2013). *The Annual Cyberbullying Survey. Retrieved from http://www.ditchthelabel.org/the-cyber-bullying-survey-2013/*

Eichhorn, K. (2001) 'Re-in/citing linguistic injuries: speech acts, cyberhate, and the spatial and temporal character of networked environments', *Computers and Composition*, 18:293-304.

Gianassi, P. (2012) The United Kingdom approach to Hate crime, Unpublished conference paper, Riga Graduate School of Law, (available at http://www.rgsl.edu.lv/uploads/files/RGSL_Konference_Cela_uz_ieklaujosu_sabi edribu_241012_UK_approach.pdf)

Gianassi, P. (2014) *Hate on the Internet: Progress on the UK Government's Hate Crime Action Plan*, presented at the All Wales Hate Crime Criminal Justice Board, British Transport Police, Cardiff, 23rd July, 2014

Greenawalt, K. (1989) *Speech Crime & The Uses of Language*, Oxford University Press.

Hanes, E. and Machin, S. (2014) 'Hate Crime in the Wake of Terror Attacks: Evidence from 7/7 and 9/11', *Journal of Contemporary Criminal Justice*, 30:247-267.

Hillyard, P, Tombs, S, & Gordon, D. (2004) *Beyond Criminology: Taking Harm Seriously*, Pluto Press.

Hinduja, S. & Patchin, J. W. (2014). Cyberbullying Identification, Prevention, and Response. Cyberbullying Research Center (www.cyberbullying.us)

Home Office (2014), Challenge it, report it, stop it: Delivering the Government's hate crime action plan, London:HMSO

IpsosMORI. (2014). *Health and Wellbeing of 15 year olds in England: Findings from the What About YOUth? Survey 2014.* The Health and Social Care Information Center. Retrieved from http://www.hscic.gov.uk/catalogue/PUB19244

iworks (2013), The Think Project Interim Evaluation, Devon: iworks ltd

Joinson, A. N. (2003) *Understanding the Psychology of Internet Behaviour*, London: Palgrave Macmillan.
Savage, M. and Burrows, R. (2007) 'The Coming Crisis of Empirical Sociology', *Sociology*, 41/5: 885-899

King, R. D. and Sutton, G. M. (2014) 'High Times for Hate Crimes: Explaining the Temporal Clustering of Hate Motivated Offending', *Criminology*, 51:871-894.

Leets, L. (2001) 'Responses to Internet Hate Sites: Is Speech Too Free in Cyberspace?', *Communication Law and Policy*, 6:287-317.

Legewie, J. (2013) 'Terrorist events and attitudes toward immigrants: A natural experiment', *American Journal of Sociology,* 118:1199–245.

Levin, B. (2002) 'Cyberhate: A Legal and Historical Analysis of Extremists' Use of Computer Networks in America', *American Behavioral Scientist*, 45:958-988.

Mishna, F., Khoury-Kassabri , M., Gadalla, T., and Daciuk. J. (2012). Risk factors for involvement in cyber bullying: Victims, bullies and bully–victim. *Children and Youth Services Review*, 34, (1) 63–70

Morell, G., Scott, S., McNeish, D. and Webster, S. (2011) *The August Riots in England*, NatCen.

Mossberger, K., Tolbert, C. J., and McNeal, R. S. (2008) *Digital Citizenship: The Internet, Society and Participation*, MIT Press.

Nixon, C. L. (2014). Current perspectives: the impact of cyberbullying on adolescent health. *Adolescent Health, Medicine and Therapeutics*, *5*, 143–158.

Office for National Statistics (2015) *Internet Access – Households and Individuals 2015, Statistical Bulletin*, ONS.

Oksanen, A., Hawdon, J., Holkeri, E., Nasi, M. and Rasanen, P. (2014) 'Exposure to Online Hate among Young Social Media Users', in M. Nicole Warehime (ed.) *Soul of Society: A Focus on the Lives of Children & Youth,* 253-273. Emerald.

Perry, B. and Olsson, P. (2009) 'Cyberhate: The Globalisation of Hate', *Information & Communications Technology Law*, 18:185-199.

Phippen, A (2014) The online safety landscape of wales, University of Plymouth (assessed http://swgfl.org.uk/products-services/esafety/wales-project/Content/WG-E-Safety-Landscape-Report-English)

Raskauskas J, Stoltz AD. (2007) Involvement in traditional and electronic bullying among adolescents. *Dev Psychol*. 43(3):564–575

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

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

Sourander A, Brunstein Komek A, Ikonen M, et al. (2010) Psychosocial risk factors associated with cyberbullying among adolescents: a population- based study. *Arch Gen Psychiatry*. 67(7):720–728

Wall, D. S. & Williams, M. (2007) 'Policing diversity in the digital age: maintaining order in virtual communities', *Criminology and Criminal Justice*, 7:391–415.

Welsh Government. (2011). Respecting others anti-bullying guidance. Retrieved from http://gov.wales/topics/educationandskills/publications/circulars/antibullying/?lang=en

Williams, M. L, Burnap, P. and Sloan, L. (2016), 'Crime Sensing with Big Data: The Affordances and Limitations of using Open Source Communications to Estimate Crime Patterns', *British Journal of Criminology*, Online Advance Access.

Williams, M. (2006) *Virtually Criminal: Crime, Deviance and Regulation Online*. Routledge.

Williams, M. L. and Tregidga, J. (2013) *All Wales hate crime research project: final report.* Cardiff: Race Equality First. Available at: http://www.cardiff.ac.uk/socsi/research/publications/Time%20for%20Justice-All%20Wales%20Hate%20Crime%20Project.pdf

Williams, M. L. and Tregidga, J. (2014) 'Hate crime victimisation in Wales: psychological and physical impacts across seven hate crime victim-types', *British Journal of Criminology*, 54:946-967.

Williams, M. L., Edwards, A., Housley, W., Burnap, P., Rana, O., Avis, N., Morgan, J., and Sloan, L. (2013) 'Policing cyber-neighbourhoods: Tension monitoring and social media networks', *Policing and Society* 23:461-481.