Uniting Church in Australia
SYNOD OF VICTORIA AND TASMANIA

Centre for Theology and Ministry
29 College Crescent
Parkville Victoria 3052
Telephone: (03) 9340 8807

Director, Online Content and eSafety Section
Department of Communications and the Arts
GPO Box 2154
Canberra ACT 2601
E-mail: onlinesafety@communications.gov.au

Submission of the Synod of Victoria and Tasmania, Uniting Church in Australia to the Online Safety Charter – consultation paper 5 April 2019

The Synod of Victoria and Tasmania, Uniting Church in Australia, welcomes the opportunity to provide a submission in response to the Online Safety Charter consultation paper. The Uniting Church in Australia has a strong commitment to protect children from child sexual abuse. The 2017 *Uniting Church National Child Safe Policy Framework* states:

> *The Uniting Church in Australia (The Church) believes that all people, including children, are made in the image of God. As a Christian community we believe that God reaches out to us in love and acceptance, and that our relationships with each other should express love, care and respect (Safe Place Position Statement developed by the UCA Commission on Women and Men in 1997). Central to living out the gospel is to love God and to love others. As a community of faith, we are committed to providing safe environments for all people including children, so that they may live life in all its fullness. We also acknowledge the rights of children as detailed in the Convention on the Rights of the Child (United Nations, 1990, Article 49) that States shall protect children from physical or mental harm and neglect, including sexual abuse and exploitation.*

It is in this context that we make this submission on the proposed Online Safety Charter.

1. What are examples of technology-facilitated solutions to enhance online safety, and how effective have these solutions been in addressing harms and mitigating risks?

In 2009, Microsoft partnered with Dartmouth College to develop PhotoDNA, a technology that aids in finding and removing known images of child exploitation. Today, PhotoDNA is used by organizations around the world and has assisted in the detection, disruption, and reporting of millions of child exploitation images.[1] Qualified organizations can use the PhotoDNA Cloud Service for free to help detect and report the distribution of child exploitation images. Microsoft began sharing the first release of PhotoDNA as an on-premise technology in 2009.[2]

PhotoDNA is an image matching technology which creates a unique 'signature' for a digital image, like a fingerprint. This can be compared to signatures of other images to find copies of the image.

The National Centre for Missing and Exploited Children creates PhotoDNA signatures of the worst known images of child sexual abuse online – images which capture the act of rape via physical penetration of an identified prepubescent child – and shares those signatures (never the images themselves) with online service providers like Microsoft and Facebook to help disrupt the redistribution of those images online.

[1] https://www.microsoft.com/en-us/photodna
[2] https://www.microsoft.com/en-us/PhotoDNA/CloudService

This technique is known in the technology industry as "hashing", but PhotoDNA's 'robust hashing' differs from other common hashing technologies because the signature is based on the essence of the image and not the file. Therefore, if an image has been resized, recolored, saved in a different file format or otherwise similarly altered, PhotoDNA can still reliably identify copies of the same image when other hashing technologies that require every file characteristic to be precisely the same could not.

In March 2012, Microsoft made PhotoDNA available to law enforcement worldwide at no charge to support child sex abuse investigations and help law enforcement more quickly identify and rescue victims and bring their abusers to justice. Law enforcement can now get PhotoDNA source code through direct licensing or in select tools they already use, including NetClean Analyze or the Child Exploitation Tracking System (CETS).[3]

In 2003, Microsoft designed a new software known as "CETS" which supports criminal investigators to efficiently organise and share media they come across during investigations. It allows units from various countries to effectively classify track and identify links between indecent material, enabling them to identify owners and uncover international child sexual abuse syndicates. By 2011 the tool was in use in seven countries, including Indonesia and Canada[4], by over 400 investigators worldwide. Microsoft offers the program to interested law enforcement agencies free of charge and donates all training and server software required to deploy the application at no cost.[5]

Australian law enforcement agencies have been making use of CETS.[6]

7. How should content moderators be trained? What minimum standards should apply?
The 2018 documentary by Hans Block and Moritz Riesewieck, *The Cleaners,* about how social media corporations outsource content moderating to places like the Philippines provided disturbing testimony from those who carry out the content moderation.[7] Those interviewed from the Philippines stated they are required to assess 25,000 images a day and have only eight seconds to allow the image to stay or to delete it. Despite speaking of horrific images of child rape and sexual exploitation and acts of terrorism, none of those interviewed referred to any procedure to report such material to law enforcement for investigation. It appeared that that the evidence is simply wiped. The documentary interviewed content managers disturbed by the material they needed to view and without the adequate psychological support. The documentary reported on one case where a content manager took their own life after their requests to move to another area of work was denied.

Thus training for content managers relevant to Australia should include:

- Should include an understanding of the Australian context, culture and community standards, including in relation to First Peoples' culture;
- Awareness of when content needs to be removed, but preserved, for referral within the organisation for further review and, where the material is illegal and not previously detected, should be referred to the appropriate law enforcement agency; and
- Awareness of when content is having a negative impact on their well-being and when to be able to seek psychological support.

[3] Microsoft PhotoDNA Fact Sheet
[4] https://www.cio.com/article/2445807/indonesia--microsoft-team-to-battle-child-predators.html
[5] Jeffrey Avina, *Public-private partnerships in the fight against crime: An emerging frontier in corporate social responsibility,* Journal of Financial Crime **18(3)** (2011), pp 289-290.
[6] https://www.acic.gov.au/our-services/child-protection/child-exploitation-tracking-system
[7] http://www.gebrueder-beetz.de/en/productions/the-cleaners

10. How should records of removed content be kept to ensure that evidence is available if needed by authorities?

We are concerned some of the technology firms destroy evidence of serious human rights abuses and other criminal activity without reporting it and do not hold records of such material for any adequate length of time. Given that a Mutual Legal Assistance request from a law enforcement agency to a foreign court can take a year or more, there is a need for evidence of such activities to be held by technology firms for at least that length of time. Facebook's policy in this regard seems reasonable:[8]

> *Information we receive about you (including financial transaction data related to purchases made with Facebook) can be accessed and preserved for an extended period when it is the subject of a legal request or obligation, governmental investigation or investigations of possible violations of our terms or policies, or otherwise to prevent harm. We also retain information from accounts disabled for term breaches for at least a year to prevent repeat abuse or other term breaches.*

However, despite being part of the same multinational corporation, Whatsapp does not make clear how long they will preserve data for.[9]

By further contrast, Google makes it less clear they will preserve and report actionable child sexual abuse material in all cases:[10]

> *Do not upload or share content that exploits or abuses children. This includes all child sexual abuse imagery (even cartoon images) and all content that presents children in a sexual manner. We will remove such content and take appropriate action, which may include disabling accounts and reporting to the National Center for Missing & Exploited Children (NCMEC) and law enforcement.*

We were unable to identify any part of Google's policies that state how long records of removed content that involve criminal activity would be retained for.

We believe the Online Safety Charter should make clear that where content detected is of human rights abuses or other criminal activity that could be subject of a law enforcement agency investigation, the record of the material should be maintained for at least a year.

Further, the Online Safety Charter should ask that technology firms host data about Australian users in Australia whenever it is feasible to do so. Allowing technology firms to select where they decide user data is located can frustrate and hinder investigations into human rights abuses and criminal activity to the detriment of users and others who have been victims of such abuse.

For example, Brian Lee Davis in the US confessed to owning hundreds of digital photos and videos that showed young children being raped. In July 2017 he was sentenced to a decade in a state prison. Law enforcement sought to pursue the entire child exploitation network he had been part of. State investigators were unable to access emails that could have helped them identify victimized children and track down the offenders Mr Davis admitted to contacting. Although Google tipped off law enforcement about the child exploitation files that had crossed its network, the corporation refused to give them access to his gmail account, despite the fact that police had a search warrant.[11] Google's argument was reported to be that the data is "out of jurisdiction." Some of the data in that Gmail account was stored on Google servers outside the United States so the corporation refused to co-operate.

---

[8] https://www.facebook.com/about/privacy
[9] https://www.whatsapp.com/legal/#privacy-policy-law-and-protection
[10] https://www.google.com/+/policy/content.html
[11] https://money.cnn.com/2017/10/19/technology/criminal-investigations-microsoft-ireland-invs/index.html

In the case of the murder of Lucy McHugh, aged 13, Facebook refused to allow UK police access to the account of Stephen Nicholson who has been charged with her rape and murder.[12] Mr Nicholson exchanged messages with Ms McHugh on Messenger. The access was denied despite pleas from Lucy's mother that Facebook co-operate with the police.[13] Lucy from Southampton, was found in woodland at Southampton Sports Centre after being stabbed to death on 25 July 2018. Stephen Nicholson, a family friend who was staying in Lucy's home until shortly before her death, was questioned on suspicion of murder and sexual activity with a child but twice refused to give detectives his Facebook password.[14] Mr Nicholson was sentenced to 14 months imprisonment over his refusal to hand over the password to his account. Failing to cooperate with police is an offence under the *Regulation of Investigatory Powers Act*. Facebook argued that it will only grant UK police access to the Facebook accounts if compelled to do so by a US court.[15] Facebook is insisting on a Mutual Legal Assistance Treaty request, which is likely to take months to complete.[16] We have not been able to find any reports if Facebook has granted police access to Ms McHugh's account. If not, then this raises serious concerns that the corporation treats user data as its own. Given the clear desire of the parent of the murdered child for Facebook to co-operate with police, Facebook should have granted police access to Ms McHugh's account if they have not done so. The case does highlight the need for laws that allow law enforcement agencies to work quickly in the interests of justice and avoid overly long and complicated legal processes. Facebook has received strong public criticism for its accessing user data to target advertising at them, while claiming to protect user privacy in the case of a murdered child.[17]

14. What are the potential pitfalls and risks with content removal? How can these risks be mitigated?

With content removal there is always risks that inappropriate and illegal content is not removed, which if often a factor of the level of resourcing assigned to the content removal. There is also the risk that content that should be referred to law enforcement, because it provides evidence that is useful in a criminal investigation is deleted and destroyed, rather than preserved and referred.

There is also the risk that content that should not be removed is. For example, a ban on nudity on a platform may end up removing fine art paintings or images of breastfeeding.

The risks can be mitigated by the quality of algorithms used and the number and level of training of content managers. Technology firms also need clear policies and procedures about identification of material that should be removed and that which should also be removed but preserved and forwarded to the relevant law enforcement agency.

---

[12] https://mirrorherald.com/stephen-nicholson-murder-lucy-mchugh/

[13] Shehab Khan, 'Lucy McHugh: Murdered schoolgirl's mother urges Facebook to give suspects' password to police', *The Independent,* 4 September 2018, https://www.independent.co.uk/news/uk/crime/lucy-mchugh-mother-facebook-password-urges-police-stacey-white-stephen-nicholson-a8521761.html

[14] Shehab Khan, 'Lucy McHugh: Murdered schoolgirl's mother urges Facebook to give suspects' password to police', *The Independent,* 4 September 2018, https://www.independent.co.uk/news/uk/crime/lucy-mchugh-mother-facebook-password-urges-police-stacey-white-stephen-nicholson-a8521761.html

[15] Alex Hern, 'Why won't Facebook give access to Lucy McHugh murder suspect's account', *The Guardian,* 6 September 2018, https://www.theguardian.com/uk-news/2018/sep/05/why-wont-facebook-provide-access-lucy-mchugh-suspect-account

[16] Alex Hern, 'Why won't Facebook give access to Lucy McHugh murder suspect's account', *The Guardian,* 6 September 2018, https://www.theguardian.com/uk-news/2018/sep/05/why-wont-facebook-provide-access-lucy-mchugh-suspect-account

[17] Mick Hume, 'Facebook's two-faced bosses show disregard for murdered Lucy McHugh's case as they protect his suspected killer', *The Sun,* 6 September 2018, https://www.thesun.co.uk/news/7188577/mick-hume-facebook-lucy-mchugh/

15. What should minimum standards for behaviour be? Should they be higher for products and services directed at children, or that have a substantial number of child users?
Products and services directed at children, or that have a substantial number of child users should have higher standards and effort should be made to ensure children can understand the terms of use, community standards and how to make complaints.

18. Are there positive examples of improving user experience currently in use?
It needs to be acknowledged that measures that improve the experience for some users may impact on the what other users believe is acceptable. For example, the online Scopely game *The Walking Dead*: *Road to Survival* recently implemented that common swear words that users used were replaced by #s. It is clear the game has a large number of child users. The measure is appropriate given there is no age control on the use of the game, but some adult users of the game reacted negatively to the measure. The proceeded to pursue measures to defeat the screening of the swear words, testing which words would be replaced by #s and testing which languages the screening applied to.

The game moderators have also been more active in placing temporary and permanent bans on players for inappropriate and abusive behaviour, such as extreme racism. This has not been welcomed by all users, but appears to be an appropriate response in the majority of cases.

20. What timeframe is reasonable to respond to complaints and reports?
Reasonable timeframes for response depends on the platform and the severity of issue. As has been publicly pointed out the removal of the footage of the terrorist attack in Christchurch required immediate response. Similarly, complaints involving behaviours that are likely to cause serious harm, such as grooming for the purposes of child sexual abuse or sextortion, require very rapid responses, while a complaint about mild swearing might be something that receives a much lower priority of response and might be dealt with within 48 hours.

21. Should reporting and complaint response timeframes vary depending on the complaint (e.g child or adult), the type of content or other factors?
As noted above, complaints about severe and serious criminal conduct, especially those which can result in the harm to a user, especially a child, require very rapid response. Technology firms need appropriate systems in place to triage complaints and detected abuses of their products and services.

33. What elements should be reported on and how can consistency of reporting be achieved?
To ensure consistency, the content of reports by technology firms should be legislated. Such reports should include:

- Policies to deal with complaints and inappropriate material and how the policies are reviewed;
- How many complaints have been received;
- How many complaints have been resolved;
- The average speed for complaint resolution;
- The volume of content removed;
- The number of users banned from the platform or service;
- The number of users suspended from the platform or service;
- The volume of content referred to law enforcement agencies for further investigation;
- The number of requests for assistance by law enforcement agencies (a sign of how safe a platform might be for its users);
- The number of law enforcement requests fully complied with;
- The number of law enforcement requests partially complied with; and

- How users can get help and access safety centres on their platform.

34. How often should reporting take place? The UK requires country-specific reporting. To what extent should a similar arrangement be developed in Australia?
The reporting should be annually and should require country specific reporting.

Dr Mark Zirnsak
Senior Social Justice Advocate
Synod of Victoria and Tasmania
Uniting Church in Australia
Phone: (03) 9340 8807