Online Harms White Paper Response

Question 1: This government has committed to annual transparency reporting. Beyond the measures set out in this White Paper, should the government do more to build a culture of transparency, trust and accountability across industry and, if so, what?

SWGfL agrees that transparency reporting is helpful. In particular, it is helpful in illuminating the general performance of platforms and the reports they process. SWGfL would suggest care in constructing the reporting indicators, especially numerical indicators. Numerical indicators, alone, may be open to misinterpretation and would require context. For example, a simple rise in reports may both highlight an inherent issue and it may also represent a rise in confidence of users to be making reports. Care is also needed to avoid influencing providers to make amends to their reporting policies or processes that aim to improve ratings rather than serve the interests of users. Again, as an example, a measurement of report closure rates - providers may choose to focus on closing reports more rapidly to increase their performance, however this may be to the detriment of their users and user experience.

SWGfL would like to also take this opportunity to highlight that providers should be required to report on the deployment of the IWF URL/Hash list. SWGfL would encourage all providers to be members of IWF and to deploy their services across their infrastructure. This would highlight those providers who are either not members and/or not protecting their users from illegal online child sexual abuse materials. To support users, particularly schools, in understanding if their filtering or ISP provider protects their connection from websites identified by IWF as containing illegal child sexual abuse material, SWGfL has developed an online utility - http://testfiltering.com/. The utility tests for the deployment of both the IWF and CTIRU URL lists and presents the user with a pass/fail result. The utility is connected with the definitions of 'appropriate filtering and monitoring' published by SWGfL as part of its UK Safer Internet Centre activities to empower users.

In our experience of working with industry, trust is the key to mutually beneficial relationships and effective transparency. It's vitally important that users of the services in scope of the white paper trust the providers and, in order to gain this trust, transparency with users is paramount. Any transparency reports created should be viewable by all and not just available to the government/ regulatory body. This will help to build trust in the wake of revelations about user data leaks and non-compliance with GDPR and the DPA. Involving industry meaningfully in the planning stages will help accountability as this will allow time to establish commitments from industry as to what they can be held accountable for. Ultimately, being seen to be accountable increases trust.

Question 2: Should designated bodies be able to bring 'super complaints' to the regulator in specific and clearly evidenced circumstances?

Yes.

As part of the UK Safer Internet Centre, SWGfL launched the ReportHarmfulContent (also referred to as RHC - https://reportharmfulcontent.com/) reporting hub in late 2018 to support victims facing legal but harmful online content. SWGfL have been providing helpline services since 2011, when the UK Safer Internet Centre launched its Professionals Online Safety Helpline supporting those working with children across the UK. In 2015, this was joined with the launch of the Revenge Porn Helpline on behalf of the Government Equalities Office and to coincide with new legislation. The RP Helpline supports victims who are facing 'Non Consensual Intimate image' (NCII) – having their intimate images shared without consent. The ReportHarmfulContent reporting hub was developed over a 5 year period and following a pilot phase, launched in December 2018.

ReportHarmfulContent reporting hub provides information, support and mediation to all UK users (over 13 years old) with regards to legal but harmful online content. At a basic level, the service provides definitions of what legal but harmful online content is; support for users facing these issues, and direction of how to report these issues to social media and online service providers. If a user has reported their issue and has received a null or an unsatisfactory (from their perspective) response from the social media or online provider, ReportHarmfulContent will assess the case. In assessing the case, ReportHarmfulContent has the opportunity to understand the context in order to determine if the response was unfair. If the response was fair, the user will be provided with advice and an explanation. If the conclusion is that the response was unfair, ReportHarmfulContent might either provide further direction or accept the case and represent the claimant with the social media or online provider. SWGfL, as UK Safer Internet Centre, has developed in-depth understanding of terms and conditions and community standards to enable it to adequately represent claimants. In the 6 months since launching ReportHarmfulContent, of the cases accepted, 87% have resulted in the harmful content being removed.

The RP helpline is the only helpline in the UK supporting adult victims of intimate image abuse by helping to remove intimate images that have been shared without consent online. Similarly, they have an excellent success rate with over 85% of escalated content being removed. Whilst this content is legal in the UK (the sharing of it without consent is illegal) and the RP helpline provides an essential service to victims, helping to prevent further re-victimisation and sharing of this illegal content.

SWGfL would consider all its support services to be unique in the UK, but of particular relevance here is ReportHarmfulContent; providing users and victims of legal but harmful online content with independent support and redress. SWGfL has developed a good understanding of industry policies and the law relating to online criminal behaviour and will escalate content for removal only when we know it breaches community standards and/ or the law.

SWGfL works in collaboration with a number of international partners. For example, SWGfL supported the Australian eSafety Commissioner's office in establishing their helpline to support victims of NCII in Australia. The ReportHarmfulContent reporting hub has benefitted from support from colleagues in both Australia, but primarily the NGO NetSafe in New Zealand. The Harmful

Digital Communications Act was introduced in 2016 in New Zealand and introduced civil and legal definitions of harmful digital content, together with a regulatory role for NetSafe in assessing cases of Harmful Digital Content. ReportHarmfulContent reporting hub, greatly benefited from the understanding NetSafes experience of in establishing a helpline in New Zealand.

ReportHarmfulContent (as well as the other SWGfL support helplines) is a primary example of a designated body that could bring 'super complaints' to a regulator. Victims would benefit from having strengthened their complaint and bring some further redress for what is likely to be a distressing situation.

In terms of language and for clarity, SWGfL considers that users should '**report'** issues to the social media or online provider. If they are unsatisfied with the response they receive they submit this to ReportHarmfulContent and if ReportHarmfulContent agrees (against criteria), a '**complaint'** is then submitted to the social media or online provider. If the outcome of this complaint is unsatisfactory, ReportHarmfulContent submit a '**super complaint'** to the regulator

Assuming the continuation of this type of complaint and super complaint function and given that it is in support of online providers and the proposed regulator, sufficient funding will be required to sustainably finance this and similar operations.

Question 2a: If your answer to question 2 is 'yes', in what circumstances should this happen?

As outlined above, SWGfL and in particularly ReportHarmfulContent reporting hub, are primary examples of circumstances when this should happen. It is clear that social media and online providers should be providing clear reporting mechanisms for their users but users and victims of legal but harmful content would, without doubt, find routes to mediation and appeal advantageous.

SWGfL Helplines and services benefit from collaborative working with responsible social media and online providers and recognise their efforts. Working to support independent appeal or mediation may be one of the activities by which these responsible providers may be able to demonstrate a 'duty of care'.

Question 3: What, if any, other measures should the government consider for users who wish to raise concerns about specific pieces of harmful content or activity, and/or breaches of the duty of care?

This will very much depend on who becomes the regulator and what powers they will have. SWGfL envisage RHC being the last stop for service users reporting harmful content once the correct industry reporting routes have been exhausted. Throughout the work on the helplines, our practitioners continuously and consistently see failings in the criminal justice system starting with poor police response to victims of crime online. In order to resolve this issue, it is imperative that emergency responders have the capacity/ funding needed to respond adequately. Specific reporting routes for different types of online crime (e.g. as is currently in place for the reporting of CSAM material with the IWF) continue to be required. Appropriate levels of training for all members of the police about the whole array of online crimes and how these might be presented by a victim or on a victim's behalf is essential in order to improve the response victims of internet crimes receive.

As mentioned above, a route for redress will be vital for all internet users and all SWGfL's Helpline Services provide this. Whoever is appointed as the regulator will require a reporting function for arbitration purposes. Report Harmful Content provides this currently and we would strongly recommend that they continue in this capacity, being funded to work in partnership with the appointed body to fulfil this requirement.

SWGfL would also like to highlight that a high degree of issues are reported via schools and established child protection and safeguarding processes. It is vital that all those working with children are suitably competent in their ability to recognise, respond and resolve issues related to online harms. Professional development of staff across UK schools remains consistently the weakest part of a schools wider online safety policy and practice and should be directly addressed.

Question 4: What role should Parliament play in scrutinising the work of the regulator, including the development of codes of practice?

Parliamentary involvement should only be in an oversight capacity and the regulator should remain independent providing reports on a regular basis. Any reporting criteria and codes of practice should be decided in collaboration with industry in order to understand what is reasonably achievable and to establish trust, accountability and transparency. The ICO's code of practice for age appropriate design shows a good example of well thought out and researched principles and a similar approach to this should be considered when developing the statutory duty of care for industry.

Question 5: Are proposals for the online platforms and services in scope of the regulatory framework a suitable basis for an effective and proportionate approach?

The scope of services included in the regulatory framework is vast which does raise questions about how granular any duty of care requirements will be and what sanctions can reasonably be applied in respect of these. It is more likely that a broad set of principles would be agreed upon and, if this is the case, it would be difficult for a regulatory body to enforce removal of harmful but not criminal content. SWGfL would recommend utilising the services highlighted in question 2 and request sufficient funding to continue supporting victims of online harmful content.

More clarity is needed around private communication and what is included in the 'tightly defined categories of illegal content' before being able to decide whether the proposals are a suitable basis for an effective and proportionate approach.

SWGfL would suggest considering gaming within the scope of online harms. Whilst appstores and in particular the information associated with apps and presented to users (age ratings, data acquisition and sharing, communication, advertising, minimum ages, privacy and terms of service etc), may be covered by the ICO, SWGfL would suggest that these are specifically included within the scope.

SWGfL would recommend that the extent of online platforms and services be as broad as possible to avoid unintended consequences. SWGfL would suggest that it is likely that the services outside of the scope would become more attractive, for example providers may be incentivised to make everything encrypted.

Question 6 In developing a definition for private communications, what criteria should be considered?

Encryption would be a good starting point and the dark web also needs to be considered. When considering the nature of forums/ groups/ pages etc the issue of privacy becomes more nuanced owing to the fact that these would be deemed as private communications but could have upwards of a thousand members/ contributors/ followers. This means that what is being discussed could be quite public. As such, size also needs to be considered.

SWGfL is concerned with the development of DNS over HTTPs. If implemented in the proposed manner, this would potentially circumvent the existing filtering technologies and solutions. The primary purpose of these filtering solutions is to manage the content that children are exposed to both at home and at school. Schools in England and Wales have a statutory obligation to establish 'appropriate levels of filtering'. Equally, the majority of families have access to network level filters that could also be bypassed by the introduction of DNS over HTTPs.

Question 7: Which channels or forums that can be considered private should be in scope of the regulatory framework?

This will depend on the definition of privacy that the government comes up with and is ultimately only an illusion due to what is being proposed. We would recommend that this is defined in collaboration with industry and organisations working in the field of online safety such as the SWGfL alongside the ICO to ensure a thorough understanding of the complexities of privacy. We would also recommend that this particular question is consulted upon with industry to ensure that those services that are considered to be in scope will be able to adhere to the regulatory framework.

SWGfL would suggest that hyper-local channels (eg airdrop and bluetooth) also be included within the scope. These channels are often used by children to share content and in particular illegal content (intimate images or copyrighted content).

Question 7a: What specific requirements might be appropriate to apply to private channels and forums in order to tackle online harms?

There needs to be a quicker, more streamlined process for sharing information in emergency situations. For example, RIPA and MLAT orders can take months to be completed meaning that if content is removed from a platform during this time, law enforcement may never be able to gather the evidence they require in order to make a prosecution. The whole process needs to happen within a couple of weeks in order to ensure data hasn't been removed from servers by the time a request has been submitted.

SWGfL suggests that the understanding amongst front line professionals (in particular social workers, police and health) be improved. This should include the correct referral routes to be followed in an emergency - in particular, the emergency disclosure routes for many of the industry platforms/ games/ apps in scope here. Via the UK Safer Internet Centre Helpline, SWGfL provide this support and training.

The current online climate encourages a choice between privacy and safety with one counteracting the other. This is seen currently with encrypted services where safety is traded off for a user's privacy. When online harm is experienced on these 'private' platforms the course for redress is much harder for the victim. The powers of investigation must be the same across all services regardless of privacy and the trade-off between privacy and safety needs to stop. Internet users should have the right to a private, and a safe online experience.

Question 8: What further steps could be taken to ensure the regulator will act in a targeted and proportionate manner?

It would be advantageous if the regulator would have the trust of industry together with a comprehensive understanding of the complexities surrounding online harms. The use of independent organisations and reporting processes would be an advantage. For example, in the immediate aftermath of the Christchurch attacks, whilst New Zealand Law Enforcement and Government departments worked to identify content to be removed, the operational relationships established by NetSafe (as an independent NGO with a regulatory role) provided capacity and greatly expedited the efficient removal of identified harmful content. This is an important aspect and SWGfL recommend that services, such as ReportHarmfulContent can adopt this role working alongside the regulator.

Government needs to be clear about the parameters of the duty of care to ensure effective regulation. If parameters are not clearly defined this may have a negative impact upon relationships with industry and lead to uncertainty about breaches within the new regulatory framework.

Question 9: What, if any, advice or support could the regulator provide to businesses, particularly start-ups and SMEs, comply with the regulatory framework?

The regulator may be able to provide research and mechanisms to support businesses, especially startups and SMEs. Businesses may benefit from a self-review system to help them understand their obligations together with their readiness as an organisation to operate within the UK. SWGfL launched 360 degree safe (for UK schools) in 2009 and Online Compass (for other organisations) in 2012, to support their self-assessment of their own policies, people, and management in protecting children online. It is evident from the data that organisations find this approach helpful and it clearly demonstrates progress over time.

Question 10: Should an online harms regulator be: (i) a new public body, or (ii) an existing public body?

SWGfL has no preference.

Question 10a: If your answer to question 10 is (ii), which body or bodies should it be?

Question 11: A new or existing regulator is intended to be cost neutral: on what basis should any funding contributions from industry be determined?

SWGfL has no preference but is clear that industry funding of a regulator should not impact any existing industry funding in this area.

SWGfL would like to highlight that the introduction of a digital tax may present opportunities of funding for a regulator and other initiatives. SWGfL, as a UK Safer Internet Centre partner operates thanks to a 50% funding contribution from the EU, and industry contributions either financial or in-kind are therefore vital to our work. The concern we have as UKSIC is to ensure that industry recognise and are able to meet this need, whilst also meeting any requirements to support the regulator. It is important that consideration of any requirements on industry to fund the regulator should also be given to ensuring the continuation funding for SWGfL and other UK Safer Internet Centre partners receive to continue their vital role.

Question 12: Should the regulator be empowered to i) disrupt business activities, or ii) undertake ISP blocking, or iii) implement a regime for senior management liability? What, if any, further powers should be available to the regulator?

This will need to be decided on a case by case basis as one model will not fit all scenarios. The regulator should have the power to remove illegal content. However, if they are to also have the same power to remove harmful but not illegal content, this will need to be defined clearly in the statutory duty of care industry are to abide by. Until this duty of care is defined, it is difficult to say which sanction(s) would be the most relevant.

Disrupting business activities and applying financial sanctions would make the most sense in terms of sanctions applied. However, it is imperative that whoever is posting the content deemed harmful and/ or illegal, is also held to account accordingly. SWGfL would highlight that whilst industry, by its very nature, facilitates the sharing of harmful content, it cannot be held solely responsible for the behaviours of the users who are doing the sharing.

SWGfL would highlight that senior management liability and 'duty of care' are terms and liabilities that have been in force for Local Authorities in England and Wales for the last 15 years following the introduction of the Children's Act 2004. Here, Local Authorities and nominated senior officers have accountability and a 'duty of care' for the safety of all children within their jurisdiction.

Question 13: Should the regulator have the power to require a company based outside the UK and EEA to appoint a nominated representative in the UK or EEA in certain circumstances?

SWGfL might assume that the introduction of a duty of care liability for senior management would require the appointment or nomination of a representative in the UK/EEA. SWGfL would exercise care in this regard in that not to discourage businesses from locating in the UK.

Question 14: In addition to judicial review should there be a statutory mechanism for companies to appeal against a decision of the regulator, as exists in relation to Ofcom under sections 192-196 of the Communications Act 2003?

We would recommend using the judicial review system already in place as it is an appeals process in itself and if a case gets to this point it may well be that legislation regarding the issue is out dated and in need of changing.

Question 14a: If your answer to question 14 is 'yes', in what circumstances should companies be able to use this statutory mechanism?

N/A

Question 14b: If your answer to question 14 is 'yes', should the appeal be decided on the basis of the principles that would be applied on an application for judicial review or on the merits of the case?

If this was implemented it would have to be decided on a case by case basis.

Question 15: What are the greatest opportunities and barriers for (i) innovation and (ii) adoption of safety technologies by UK organisations, and what role should government play in addressing these?

The biggest opportunities come from AI being used as a tool for good and the gamification of safety features to appeal to younger users.

Often a start-up will be developed by someone with little comprehension of the potential safeguarding risks and it is only in hindsight that safety and policy is considered. By joining up these 2 teams from the start there is a better chance for safety by design to be implemented in the initial development phase.

Other key barriers include encryption of services making it harder to retrieve information; regulation itself could prevent new business opportunities in the UK;

Uncertainty surrounding Brexit is the cause of challenge for many organisations including the UK Safer Internet Centre (of which SWGfL are a part of). UK Safer Internet Centre partners are in receipt of European funding to maintain many vital national services and their continued contribution is at great risk.

SWGfL would recommend that all social media and online providers utilise the technologies available to identify and prevent illegal and harmful content (images and video) to be uploaded, specifically the use of the IWF Image Hash list.

Question 16: What, if any, are the most significant areas in which organisations need practical guidance to build products that are safe by design?

Age Verification has the potential to be a game changer in this space and if it can be made to work effectively for the adult industry it would make sense to deploy this across all services in scope to allow them to prevent underage users from accessing their platforms. SWGfL and the UK Safer Internet Centre partners are well placed to provide practical guidance to organisations about safety by design with regard to young people.

It is apparent from our work across all SWGfL helplines that Violence against Women and girls (VAWG) accounts for a disproportionately large amount of the cases we deal with and we propose that guidance is issued in this area specifically. SWGfL helplines would be keen to collaborate with other stakeholders to produce this.

SWGfL would recommend support for organisations developing products that accept or host digital images and video to be required in order to prevent illegal images from being uploaded by using hashing technologies. For example, SWGfL would like to see social media and online providers utilise the IWF hash list to prevent known illegal CSAM images from being uploaded.

SWGfL, as lead partner in the UK Safer Internet Centre, publishes definitions to support schools in England and Wales in the responsibility to provide ‘appropriate filtering systems’. The definitions also support filtering providers in developing their solutions. The area of extremism and radicalisation is complex and more support from Government, academics and experts to be provided to providers, particularly with regards to definitions, search terms and illegal image and video identification.

Question 17: Should the government be doing more to help people manage their own and their children's online safety and, if so, what?

Online safety is everyone's responsibility and there is always room for more to be done to protect users. SWGfL would like to see support for awareness raising activities, for example for Safer Internet Day.

SWGfL would like to see the creation of a labelling schema to support users, particularly parents' understanding of apps and services. The use of labelling schemas is widespread, for example nutritional labelling on food packaging, laundry labelling and eco labelling etc. A labelling schema for apps or services, particularly available through appstores, would be a great advantage to articulate primary aspects of the particular app, for example, minimum age, advertising, acquisition and use of data, communication as well as nature of content. This would be a significant development to help people (especially parents) manage their own and their children's online safety.

Question 18: What, if any, role should the regulator have in relation to education and awareness activity?

Education and awareness should be a cornerstone of any strategy, as 'prevention is better than cure'. Many online safety resources have been developed over the last two decades, most in response to particular issues and solely focus on harms, often extreme harms. SWGfL would question these and whilst children may find them entertaining or indeed shocking, the majority cannot see the relevance to them. The analogy used is that we don't learn to drive a car by merely watching films of car crashes. This said, it is important to highlight the potential harms, but suggest that education and awareness should focus on the skills and competencies required to navigate and benefit from technology and online services. It is important that Governments invest suitable sustainable expectations and funding to ensure schools and other education providers safeguard and educate children to ensure they benefit from technology, free from harm. A whole new approach is required

In terms of looking to a role that regulator may have in relation to formal education - education is clearly a devolved mater and therefore the regulator is only able to work to support the Department for Education, Welsh and Scottish Governments and Northern Ireland Executive. The variety of education systems (including private sector), including the legislative environments across the four nations is broad and sensitivity is evidently required in order to make any engagement or support universally applicable.

Looking at the challenges facing the formal UK education sectors, they are clear from the annual assessment report that SWGfL publishes annually regarding UK schools' online safety policy and practice. The latest report (2018), authored by Prof Andy Phippen, draws on self-assessment data from the 14,500 UK schools using 360 degree safe and highlights that "*areas of concern, primarily around training and the development of knowledge in the wider community:*

- *50% have carried out no governor training around online safety issues with only a slight improvement on 2017*
- *43% have no staff training to date around online safety, although this has improved on 47% in 2017. Staff training remains consistently one of the weakest aspects*
- *The majority (54%) of schools are not evaluating the impact of their online safety efforts.*
- *Whilst there has been a 2% improvement, it remains that 30% of schools have insufficient data protection provision.*

The issue with training is something that continues to cause concern and we will continue to raise as this is the other part of the foundation of effective online safety practice. Without effective knowledge by staff, and those who scrutinize the staff, we cannot hope to have effective practice. We know from our work with young people that one of the things they call for is knowledgeable and understanding staff. If over 40% of schools have no staff training programme in place, not only are the failing in their statutory duties, but it is unlikely they would be able to effectively support young people in their care when addressing online safety incidents. Schools need effective training to deliver online safety and ensure young people and the wider school community engage with the online world in a resilient and risk mitigating manner."

The regulator may have a role in helping organisations, such as schools, in highlighting effective online safety organisations, emulating the role of the Australian eSafety Commissioner. SWGfL, on behalf of UK Safer Internet Centre, in conversation with the Australian eSafety Commissioners Office is exploring the adaptation of their Trusted provider scheme.

Additionally, the regulator should help existing services to address complaints (as previously outlined) that may arise regarding incidences in schools.   There may also be a role for the regulator in supporting non formal or extra curricula education programmes, for example emulating the CyberDiscovery programme that aims to provide cybersecurity skills to children through extra curricula programmes.

Turning to awareness, there might be a significant role for the regulator, lending support to initiatives such as Safer Internet Day and other parallel opportunities.