



Internet Regulation and the Online Harms White Paper Stakeholder Workshop¹

Summary by Victoria Nash, Associate Professor & Senior Policy Fellow,
Oxford Internet Institute (OII), University of Oxford

This report summarises key points that emerged from a day-long multi-stakeholder workshop convened to discuss the implications of the 2019 Online Harms White Paper. The range of organisations who attended are listed at the end of this summary, but included human rights NGOs, social media platforms, telecoms and media companies, news media, industry associations, parenting and child rights organisations, academia, think tanks, government departments and independent regulators. The main children’s charities were invited but unable to attend. The day’s main conclusions were:

- There is need for a systematic approach to dealing with problematic content online, but the proposed ‘duty of care’ approach is not the right approach.
- There is broad support for measures that would help tackle the worst illegal content, such as child sexual abuse material and terrorist content but proposals to tackle legal but ‘harmful’ content drew much more criticism.
- Participants wanted to see more accountability across the whole ecosystem of actors involved in decisions about the removal of online content, where that ecosystem also includes government actors, regulators and the wider range of industry players providing third-party services such as content moderation.
- Going forwards, it is vital that policy formulation and delivery be proactively conducted with the involvement of all stakeholders engaged together, rather than dealt with in separate silos.
- In order to serve as a clear, effective and proportionate framework for future policies, key aspects of the proposed framework must be reassessed and clarified, particularly on issues such as scope (which companies are covered and why); practicability of

¹ Convened by the OII, Coalition for a Digital Economy (Coadec), Open Rights Group, Index on Censorship and Global Partners Digital

codes of practice; the likely measures of success and means by which human rights will be protected.

- As part of such a process of clarification, we strongly recommend that government take time to address the trade-offs involved in developing their policy framework, justifying decisions taken on the basis of available evidence and/or value judgements.

Background and aims

The rationale for holding an explicitly multi-stakeholder event with representatives drawn from across all relevant sectors was a determination to confront the various, potentially conflicting interests and values at stake for different actors whilst recognising our common interests in ensuring that any new regulation in this space is clear, effective and proportionate. In order to provide the most useful feedback for government we sought to identify possible points of consensus but also to identify areas where fundamental and persistent disagreements imply that government must acknowledge and justify decisions in the light of unavoidable trade-offs.

The day's agenda was framed around five objectives:

- To identify participants' core concerns relating to the challenges presented by social media and online content and explore what may be reasonable and/or possible for government to address and why;
- To review and scrutinise the government's proposals for tackling unlawful and harmful online content;
- To explore the potential impacts of regulation upon human rights including freedom of expression resulting from the proposals;
- To understand the potential implications of the proposals for business, particularly small and medium-sized businesses;
- To identify where consensus exists in relation to the above and consider possible responses.

Key points of agreement

Given the range of organisations represented in the room, it is unsurprising that it proved easier to find consensus (albeit not unanimity) on issues of broad principle rather than fine detail. Most notably, our attendees did express their support for a more systematic framework to address the challenges presented by social media and online content. There was, however, no confidence that the duty of care approach as currently framed is the right foundation for such a framework. Many participants noted that the concept of duty of care does not translate well from the offline to the online context, and as such it provides little

clarity as to what duties can and should be expected of companies within scope of the OHWP. If it is to be retained in the post-consultation framework, the utility of the 'duty of care' concept will depend on clarification of how its use in the online context differs (for example recognising that platforms provide a space within which content provided by one user may harm another user, rather than themselves directly harming users). One possible approach would be to express an online duty of care as hinging on the adequacy of available platform tools and procedures which users would employ to protect themselves.

In terms of the issues covered by the OHWP, participants were most comfortable with the idea that any new regulatory framework would focus primarily on tackling illegal content, indeed, there was broad support for measures that could enable law enforcement to more effectively tackle the worst forms of illegal content. The inclusion of proposals to tackle legal but harmful content using regulatory measures was much less popular. Many participants suggested that a stronger evidence base was needed to justify intervention in the case of legal but harmful content, indeed, even the term 'harmful content' was questioned by some, on the basis that the effects of legal but risky content such as eating disorder images could not be predicted, and would depend on the context and characteristics of users. A few participants favoured a stepwise approach: any new policies should start with existing categories of illegal content and seek to assess how such content could be more effectively tackled, whilst gradually collecting evidence of harm for other categories of content that might later fall under scope. There was, however, some agreement that if both categories are to be addressed then very different measures would be needed in order to provide appropriate protection for human rights.

Another point which found broad support amongst participants was the need for more accountability (not just transparency) across the whole digital content ecosystem. This emphasis on accountability rather than transparency was explained as a vital step towards incentivising more responsible behaviour by all actors involved in making decisions about online content. Several participants argued that much more data was needed to shed light on how companies deal with take-down requests, appeals and complaints, but that raw figures would not be enough to ensure accountability for these practices. Human rights organisations argued that this principle should apply just as much to governments (for example to enable more effective oversight of notice-and-take-down requests) as businesses. It was also pointed out that with the growth of a new industry sector providing third-party services such as content moderation and GDPR-compliant social media services for children, such accountability requirements should be applied to the wider ecosystem of industry partners responsible for content decisions.

A final point of consensus was that in a policy area as complex and wide-ranging as the OHWP, there is a great need for genuine and sustained multi-stakeholder engagement at every stage of the policy-making process, beginning with policy formulation and continuing through consultation and delivery. Whilst it was recognised that government had reached out to

different sectors throughout the policy process, the silo-ed nature of these discussions was perceived to have prevented robust and honest discussion of the potential trade-offs involved in finding a framework to address online harms.

Priorities for clarification

In addition to finding some points of consensus, our participants also highlighted aspects of the OHWP in need of further clarification.

At present, the scope of actors targeted by the proposed regulation is too broad. Members of our workshop noted that as currently drawn, the range of actors to be covered could include business-to-business communications platforms, journalism and media companies that host comment sections or mental health charities whose very purpose is to enable discussion of issues such as eating disorders. Even if all are ultimately included, it is not obvious that codes of practice can be produced in a way which would be relevant for all players. Attendees representing the SME sector noted that if the same standards of compliance are to apply to small businesses and start-ups, then it would be helpful to identify areas where compliance costs could be reduced. The example was presented of IWF fees, which are currently necessary to support the operation of this crucial organisation, but which can serve as a disincentive to partnering for small firms. In another area where breadth of scope needs to be reconsidered, carve-outs may be needed to protect freedom of the press for media and journalism providers.

We asked what stakeholders thought of the principle of co-regulation as a means of addressing some of the issues raised in the OHWP. There was broad support in favour of this option if it meant seeing companies sign up to principles or codes. However, as always, the devil is in the detail. Several participants expressed scepticism that workable codes of practice could be developed; if established on an issue-by-issue basis, as exemplified by the forthcoming codes of practice for tackling child sexual abuse and exploitation material and terrorist content, then there was concern that this could result in a fragmented and confusing range of codes. Others worried that it would be challenging to produce codes of practice that would have relevance for all sectors or sizes of organisation. Oversight and approval of the codes was also raised as a concern; several organisations expressed disquiet that the Home Office (rather than an independent regulator) is currently intending to sign off on codes of practice relating to child sexual abuse material and terrorism.

Several participants (and not just those from human rights organisations) expressed disappointment that the OHWP does not place human rights considerations at the heart of the policy framework, given that the recent Age-Appropriate Design Code did endeavour to do this. Within the group there was a willingness to acknowledge that positive rights (users feeling enabled to speak or participate) matter alongside negative rights (formal rights to

speak freely) and that trade-offs could be made between the two. In addition to the focus of the OHWP on protection from harm, one participant also argued that any new regulatory framework should promote rights of due process, for example, ensuring that users who feel their complaints have been ignored or their content removed unfairly, should have clear mechanisms for appeal and adjudication.

Several attendees emphasised the need to consider how the success of any new measures would be assessed. Whilst the group received reassurances from attending civil servants that the focus would be upon systemic failures to remove content rather than individual cases, this still raised the question of how systemic failure would be judged. This is an area where there are explicit trade-offs to be made: for example, specifying short timescales for content removal may be preferred by groups protecting the interests of vulnerable users, but may undermine accuracy in decision-making and pose a threat to freedom of expression by incentivising over-removal of content. Similarly, some types of content will be easier or harder to assess using automated processes due to the vagaries of factors such as context, positionality, or changing uses of language, (we might expect it to be harder to accurately remove hate speech than child sexual abuse material, for example). There is thus another trade-off to be decided between wanting platforms to be completely free of the worst forms of content, and recognising that in some categories (such as hate speech) this could also risk removing legal but offensive speech protected by the right to freedom of expression. The different groups at our workshop individually reach their own decisions about where the balance should be struck on these issues, but ultimately government must also decide. Given the significance of the trade-offs involved, a clear justification will be needed for the position chosen.

Finally, in discussing different types of problematic content, participants wanted to know that government is attentive to the underlying social factors which give rise to users posting and sharing such content online. Issues such as child sexual abuse and exploitation or online disinformation require a systemic 'public health' type approach to eradicate them; they may manifest as a technological problem, but they are fundamentally social, economic (and in the latter case, political) problems. In this light, there was a request that efforts be made to ensure that policies resulting from the OHWP be explicitly joined up with other relevant initiatives to ensure a more consistent and effective approach to tackling the underlying social causes of illegal and problematic user behaviours online.

Making difficult choices

On many issues, it was clear that workshop attendees could not agree, reflecting their different interests, values and communities. But as a group, participants emphasised that government will need to explicitly tackle such points of disagreement and underlying trade-offs between different outcomes. In doing so, they need to provide justifications for these

decisions, reflecting available evidence and/or value judgements. Some of the potential trade-offs identified included:

- Introducing metrics of success that would prioritise speed or completeness of removal over accuracy, as discussed above;
- Choosing codes of practice with very prescriptive requirements which would provide greater regulatory control but could also limit industry's potential to adopt innovative new solutions;
- Expanding requirements for generalised monitoring of individuals' communication in public channels, which seems to unavoidably affect personal privacy;
- Balancing human rights to participation and freedom of expression with rights to protection from harm;
- Imposing broad regulatory responsibilities on companies of all sizes whilst also aspiring to promote innovation and competition.

Conclusion

Many workshop participants praised the UK government's ambition to develop a systematic policy framework that would promote positive uses of new digital technologies. As it stands though, the group did not support the adoption of a 'duty of care' approach, and would like to see the consultation process continue over the summer and into the autumn to ensure that careful consideration can be given to the many nuanced and difficult issues raised so far.

Organisations Represented at the Workshop

Adam Smith Institute
Article19
Ask.fm and Yuoubo
Association for UK Interactive Entertainment
Bird & Bird
British Board of Film Classification
Cabinet Office
Child Rights International Network
Coalition for a Digital Economy
Confederation of British Industry
Demos
Department for Digital, Culture, Media and Sport
Facebook
Global Partners Digital
Google
Guardian
Home Office
Index on Censorship
Liberty
London School of Economics
Microsoft
Mumsnet
Ofcom
Open Rights Group
Parent Zone
Privacy International
Professional Publishers Association
Reporters Without Borders
RightsWatch UK
Society of Editors
Stonewall
TechUK
Twitter
UNICEF UK
University of Oxford
Verizon Media