Netsafe response to the Safer Online Services and Media Platforms Consultation

About Netsafe

1. Netsafe is New Zealand's independent, non-profit online safety charity. Taking a technology-positive approach to the challenges digital technology presents, we work to help people in New Zealand take advantage of the opportunities available through technology by providing practical tools, support and advice for managing online challenges.

2. We are an independent non-profit organisation adjacent to government and law enforcement, supported by the public and private sector and with a focus on online safety. Netsafe provides free support, advice and education seven days a week through a helpline, our website and face to face service delivery across New Zealand

3. Netsafe is also the Approved Agency under the Harmful Digital Communications Act 2015 (HDCA). One of the purposes of the HDCA is to deter, prevent, and mitigate harm caused to individuals by digital communications. Netsafe's functions as the Approved Agency are set out in section 8 of the HDCA. Those functions include:
    a. to receive and assess complaints about harm caused to individuals by digital communications
    b. to investigate complaints
    c. to use advice, negotiation, mediation, and persuasion (as appropriate) to resolve complaints
    d. to establish and maintain relationships with domestic and foreign service providers, online content hosts, and agencies (as appropriate) to achieve the purpose of the Act
    e. to provide education and advice on policies for online safety and conduct on the Internet.

4. Netsafe's comments are focused on our experiences dealing with online harms as both a not for profit agency and as the Approved Agency under the HDCA. Our comments are less focused on aspects of the proposal dealing with offline content (albeit comments may be relevant to offline content as well), and do not necessarily answer all questions posed.

General comments

5. Netsafe welcomes the thinking that the DIA has put into the questions of whether and how our online world should be better regulated. Our experiences in dealing with the many cases of online harm that are reported to us show there is more that can and should be done both to prevent and address such harm. In the last reporting year Netsafe received over 28,000 reports concerning online safety and these numbers continue to rise. Reports made to Netsafe include the following:
    a. online bullying and harassment
    b. intimate image abuse including nonconsensual sharing of images
    c. child sexual abuse
    d. harassment
    e. sextortion and webcam blackmail
    f. hate speech and discrimination
    g. privacy breaches
    h. scams

6. We know with increasing digitalisation, whether in the context of the drive to interact with public or government services online and the use of devices and internet based teaching in schools, that it is becoming increasingly difficult simply to opt out of online life. For children growing up today the distinction between online and offline life is almost meaningless. So it is not simply a case of user beware. We also think it is difficult to draw analogies to regulation in the offline world. Communication and consumption of content via the internet and social media has its own challenges which do not exist in the offline world. The systems designed for regulating the online world need to be customised to the online environment.

7. The reports to Netsafe likely represent only a fraction of the online harms experienced by New Zealanders. It is clear that the system as it currently operates does its best to tackle online harm, but is imperfect. In particular, the current system is designed to be reactive and focusses on seeking to remedy harm only once it has occurred. Current legislation does not oblige platforms, nor indeed any other actors, to take proactive measures to reduce or prevent online harm. While many platforms have sought to take unilateral action or to self-regulate, a combination of voluntary and mandatory approaches may be necessary to better protect people in New Zealand. We note that many, if not most, platforms are owned by profit driven companies, and while there has been investment in trust and safety in recent years, trust and safety teams also bear the brunt of cost cutting measures in leaner times. We have recently witnessed a direct correlation between the significant downsizing of the trust and safety functions in many social media organisations and an increase in harmful online content. It is clear therefore that we cannot rely only on self-regulation. The case for proactive action to reduce or prevent harm before it occurs as an additional lever in the online safety eco-systems is in our view justified. However, we are not yet convinced of all of the proactive measures suggested in this consultation paper and think further analysis ought to be undertaken and detail provided before being able to comment in detail.

8. We are grateful for the work to date and are optimistic that meaningful change is in the offing. However, we consider further public engagement backed up by more detailed

qualitative and quantitative research would put these proposals on a firmer foundation. In particular, while Netsafe has its own experience of online harms as a result of the work we do, we think in terms of public engagement the consultation could and should have included more in the way of empirical evidence on harms being experienced by New Zealanders to back up the proposals.

9. We also consider greater analysis and discussion is warranted in respect of rights under the New Zealand Bill of Rights Act and in particular freedom of expression. The proposals would ultimately have the effect of removing or restraining legal content. Without addressing this issue more explicitly, and addressing how the right to freedom of expression can be balanced with rights not to be harmed online, the proposals may not achieve the necessary public acceptance.

10. We are also concerned about the length of the consultation period for such an important and potentially far reaching change, and question whether there has been sufficient publicity and public engagement around the proposals. We also note that the consultation paper has left many questions open ended or unanswered. It is our assumption that the current proposals are a step in the process of reform, do not represent the only chance to engage, and that there will be more meaningful consultation in future as proposals evolve and crystallise.

Interaction with the HDCA

11. Netsafe's main and overarching comment on the consultation paper is its failure sufficiently to mention, analyse and engage with the Harmful Digital Communications Act 2015. New Zealand already has a relatively well developed system for dealing with harmful online content in the form of the HDCA. While the HDCA does not regulate the platforms per se and is reactive in its design, it is an important piece of legislation which has many cross overs with the proposals in the consultation paper. It is surprising, therefore, that the existence of the HDCA is barely acknowledged and that the consultation paper does not discuss what the HDCA already does, why it may be insufficient when dealing with platforms, how the HDCA might be improved and how it intersects with the current proposals. Further work and analysis of the HDCA in light of the proposals is clearly required.

12. We also note that significant work was undertaken by the Law Commission in its paper *Harmful Digital Communications: The adequacy of the current sanctions and remedies* (Ministerial Briefing Paper, August 2012) as a precursor to HDCA. Much of what is set out in the Law Commission's paper remains relevant in the context of the current proposals (for example the discussion on freedom of expression). It would be useful to engage with that pre-existing work.

13. In our view the HDCA is a key pillar in the New Zealand's online safety regime to which elements of the current proposals might be added. The HDCA is by nature and design reactive. It contains both civil and criminal processes for reacting to harm

caused to individuals by harmful digital communications. The consultation paper (albeit not very clearly) appears to proceed on the assumption that the HDCA will continue to exist and operate and makes no mention of any particular reform to that legislation or consider relevant jurisprudence under that Act. We would endorse the continuation of the HDCA as a pillar in New Zealand's online safety regime and would endorse an additional pillar which would place proactive obligations on platforms (which the HDCA does not currently do).

14. However, we think further in depth analysis is required on the interaction between the HDCA and the current proposals and consider the following sorts of questions need to be answered and with an opportunity for further consultation:
    a. Will any of the Approved Agency's functions change and if so what functions?
    b. Who should an individual complain to? Will the SOSMP proposals require a complaint to the platform first? When can or should an individual complain to the proposed regulator instead of the Approved Agency and vice versa?
    c. Can the District Court still order a platform to remove content under the HDCA (i.e. exercise existing take down powers)? How would this interact with proposals under SOSMP given the different context, definitions and thresholds?
    d. The Approved Agency is obliged under s8(1)(d) HDCA to establish relationships with online content hosts and platforms and indeed has done so, including signing up to and participating in schemes to escalate matters related to harmful content. What will or should happen to these relationships?
    e. The Approved Agency is required to provide education about online safety under s8(1)(e) HDCA. How will this obligation interact with similar obligations the new regulator might have in this space?

15. Netsafe has a further specific concern in that during the consultation period Government made a decision to move the Approved Agency contract and funding arrangements from the Ministry of Justice and Ministry of Education to the Department of Internal Affairs. Neither Netsafe, nor apparently the MoJ or the MoE were consulted about this. Moving the contractual relationship to the DIA raises questions such as whether there are greater plans afoot to consolidate the online safety regulatory ecosystem within the DIA, whether there are plans to change the role of the Approved Agency or to amend the HDCA, and whether there might be any potential conflicts of interest in DIA's role in the wider online safety ecosystem. This change and its potential effects was not signaled in the consultation paper and significant questions therefore arise as to whether this is a genuine open minded consultation or whether certain decisions have been made behind closed doors and on which the New Zealand public and affected parties have no input. Whatever the situation it impacts DIA's proposed regulatory model and ideas in the discussion document.

Definitions of "harmful" and "unsafe"

16. We note that there is a pre-existing definition for "harm" in the HDCA. "Harm" is defined as "serious emotional distress" (see s4 HDCA). On the assumption that the HDCA would continue to operate, further work is required to ensure that any new

definitions arising from the SOSMP proposals do not contradict or otherwise undermine the HDCA and growing jurisprudence about this definition.

17. While we agree that key concepts ought to be defined for the purposes of any legislation we do not consider that the definitions provided for in the consultation paper for “harmful” and “unsafe” are sufficiently precise or useful at this stage of the process (even though they appear to cover the right concepts in broad terms). The definitions as drafted for the purposes of the consultation paper are vague and at times confusing. For example, definitions are provided in the bullets on paragraph 10 on page 18 of the consultation document but then appear to be added to in paragraph 12 which appears to consider group or societal harms in the form of discrimination or interference in democratic processes as also being harms with which the proposals are concerned. It also talks about “troubling but not illegal content” (p49). Are these concepts included in the definitions? It is important also to be clear exactly when and how these definitions might be relevant. The paper talks about harmful “conduct” and also harmful “content” – but it’s not always clear whether this is in respect of the platforms, posters or content creators. It is vitally important that those who will have to apply to definitions, or will be subject to them, have as much clarity as possible. Presumably how any legislation will define these terms will ultimately be for the legislative drafters.

18. We are also concerned about what appears to be an entirely subjective definition. Without any objective element it may be difficult if not impossible to apply in practice.

19. Whatever definitions are settled on, we consider there ought to be consideration of a seriousness threshold either built into the definitions or applied depending on the particular obligation. A seriousness threshold was included in the definition of “harm” under the HDCA as a justification for the exercise of powers which would interference with the rights contained in the New Zealand Bill of Rights Act 1990, in particular the right to freedom of expression. We consider further analysis of the definitions, and indeed the proposals as a whole, in light of the NZBORA is required.

20. Finally, whatever definitions are chosen, given not all harms are equal and not all harms justify the same intervention it is difficult to provide further comment until these definitions can be seen and applied in context.

Definition of platform

21. On the definition of platform while we understand the objective of not imposing unnecessary regulatory burdens on smaller operators the paper does not set out clearly how or why the particular threshold figures have been arrived at and why they are suitable or relevant in the New Zealand context. How will the numbers be established? Will they be publicly available? Will platforms know definitively if they are within scope of regulation or not? It would be helpful to understand this better.

22. We also note the consultation paper acknowledges the proposals will not be able to prevent every bad actor. However, having thresholds could result in harmful content or

those who seek to produce or post it to move to smaller platforms. As a way of mitigating this, platforms that present a particular risk would also be brought within scope. It will be important, therefore, to establish clear criteria that would bring a smaller platform that fell below the numbers threshold into the regime. What are the relevant risk factors? Who would assess them? Would such a decision be subject to review or appeal for example?

23. For those platforms that do not meet the numbers threshold and do not objectively present a risk such as to bring them within scope of the regulation is it safe to assume they will have no obligations other than those that apply at present under the proposals? To the extent they end up hosting unsafe or harmful content as defined is this the sort of scenario that the consultation paper accepts is beyond the scope of the proposals?

24. As a potential answer to the question posed above, we think further consideration should be given to proactive versus reactive obligations on platforms. For example, greater proactive obligations on larger platforms may be justified from a resources and regulatory burden perspective e.g. obligations in respect of risk analysis, transparency reporting, threat detection etc. But to the extent persistent harmful or unsafe content is hosted on smaller platforms there could be a case for imposing certain less burdensome reactive obligations on them (e.g. the requirement to have a complaints process) as a response to the concern of creating loopholes or creating an unlevel playing field.

25. We also note that messaging platforms such as Telegram, Signal or WhatsApp appear not be in scope, presumably on the basis that they are “private” messaging systems. We think the public versus private dichotomy may be harder to justify and apply in practice. For example, “private” Telegram groups can have up to 200,000 members. While these groups are opt in, so too are subscription services, private groups on e.g. Facebook or paid for websites and content. Indeed in many instances simply going online and consuming content may be a choice (although as set out in our general remarks we acknowledge that for many this “choice” may not really apply in practice). We think the proposals need to be clearer about the justifications for bringing platforms within scope or not, and if platforms are out of scope what alternative remedies there may be.

26. Finally, whatever thresholds are chosen, if regulated platforms within scope are able to be held to account in New Zealand further thought is needed on how exactly enforcement against such platforms (should it be necessary) can take place. For example will regulated platforms be required to have a New Zealand based representative who is legally responsible?

Codes of Practice

27. Codes of practice are a model of regulation that Netsafe endorses in principle. In the absence of any intervention by Government or Parliament the voluntary Aotearoa New Zealand Code of Practice for Online Safety was drafted to encourage proactive action

on the part of platforms. While the Code is still in its infancy, we think industry led codes are likely to be more sustainable in the long term, and allow for flexibility as the industry and technology evolve. We note this proposition is consistent with international practice and agree it makes sense to ensure a measure of consistency with international practice. But this should not lead to the lowest common denominator. We agree that ensuring the New Zealand context is reflected in codes is vitally important.

28. However, it is the actual content of codes that matter. The content of a voluntary code is likely to differ quite significantly to the content of a code that is backed up by regulation or enforcement, even if the drafting of that code is industry led. It is for that reason that harder edged obligations or outcomes set out in legislation may be warranted.

29. We also agree that there should be meaningful participation of civil society and other affected groups in the drafting and governance or oversight of the codes. It should not be left up to the discretion of the regulator who to involve and when. These obligations should be set down in legislation.

30. Further detail is required on how such codes are to be endorsed - only by the regulator? Presumably Courts will be able to scrutinise code provisions or decisions, and actions or omissions of the regulator that fail to comply with administrative law obligations or are inconsistent with the NZBORA? Further detail on this is necessary.

31. We think the consultation paper has not been as open or transparent as it ought to be about the effect of enforceable codes of practice on freedom of expression. If the intent is that platforms will moderate or remove more lawful content than they do at present in response to enforceable provisions of a code of practice then this undoubtedly will have an effect on freedom of expression of those who have either created or posted the content. While removal or moderation of otherwise lawful content which has the effect of protecting children and other vulnerable groups from online harms may be a worthy objective, we think there needs to be a much more open and transparent debate. This indirect route to content moderation and removal needs further discussion, and greater assurances for content producers and authors is needed.

The proposed regulator and the wider online safety ecosystem

32. If there is to be an independent regulator and public buy in or acceptance of a regulator which has censorship and take down powers (amongst others) affecting freedom of speech, then such a regulator must be truly independent of the Executive. We would be concerned with there being a role for government in “monitoring” the regulator (see p 58/59) or producing policy statements to influence the regulator or in ensuring “Codes deliver Government consumer safety expectations”. That does not suggest independence. We think Parliament is the appropriate body to set expectations by way of legislation. If the regulator does not comply with what is set out in legislation then the regulator’s actions or omissions should be reviewable by the

Courts. We think more thought is required on how a regulator can fulfil its statutory functions at the same time as having the necessary independence from the Executive – especially in the context of regulating speech and the wider effects on rights under the NZBORA. Ultimate Parliamentary and judicial oversight may be a better model.

33. In this same context we think greater clarity is required on the specific role of government departments more generally. What is the proposed role of the Executive in day-to-day management of, or operating within, the system? On one view DIA's existing functions and powers would continue to exist, but it would have added oversight functions in respect of the regulator. If this is correct then we do not think consolidation of such powers and functions in the DIA is appropriate.

34. Similarly what role will the judiciary play in the context of these proposals? As set out above the Courts already have a role under the HDCA in the context of order making powers (including powers to order content be removed). But will courts also be able to judicially review the regulator and codes of practice? And what appeal routes would lie to the Courts?

35. We also consider further thought should be given to whether it really is appropriate to have a single regulator overseeing both online platforms and traditional media given the significant differences there are in regulating traditional media and online platforms. Is further thought being given to the option of a more constrained regulatory model similar to say the Commerce Commission and various subject matter regulators that are industry specific?

36. We also note the consultation paper says:

> "The current system is difficult to navigate and has big gaps. New Zealanders must figure out which of five industry complaint bodies to go to if they feel content is unsafe or breaches the conditions of the platform it is on. On top of that, not all forms of content are covered by those bodies."

37. However, the consultation paper does not actually appear to address this particular issue. In particular, there would still be four industry complaint bodies to go to at the end of these proposals. Netsafe does not necessarily endorse consolidation of all of the industry players into a single body. However, we could see a key role of the regulator being to join the system up more effectively, for example by coordinating a single reporting or complaints system and facilitating greater cooperation amongst all industry players.

Enforcement

38. We note the consultation proposals on enforcement are still embryonic and that responses are sought on an enforcement regime.

39. As regards takedown powers, given take down powers and enforcement of such powers is ultimately censorship with an effect on freedom of expression of the original

creator or poster there needs to be greater analysis and discussion and a clear justification on the interference with this right. At this stage we think the only take down powers the regulator should have should be in respect of powers and functions it would inherit under the classifications regime. No case has been made for broader take down powers whether under other criminal or civil laws. In particular, if there is a suggestion the regulator might have the power to issue take down notices in respect of civil matters that would, in our view, be a step too far and would appear to give the regulator unrivalled powers (and concomitant burdens).

40. We agree with the principle that any regulator should not have the power to undertake criminal prosecutions (especially if the regulator is also able to determine that content is objectionable or illegal) and that any criminal prosecutions should remain with current prosecutors.

41. As above we think further analysis is required on the interaction with the enforcement regime under the HDCA.

42. Further thought is also required on ultimate enforcement. The consultation paper suggests additional fines, and criminal prosecution of platforms for persistent non-compliance. The consultation document also suggests that the regulator should have the power to "initiate civil and criminal prosecutions". However, many regulated platforms may have no meaningful presence in NZ to be able to enforce civil or criminal penalties against them. What are the proposals to ensure meaningful enforcement? Will regulated platforms be required to have a registered New Zealand address, or to incorporate in New Zealand for example?

Education

43. We very much endorse the proposals for further education and awareness raising– both in schools and in the wider community. Despite drives by Government to encourage engagement with public services and officialdom through digital means (in other words essentially requiring the public to go online in order to interact with government), online safety is education and awareness raising remains significantly underfunded. Digital citizenship, media literacy, critical thinking and basic online safety knowledge are all vital skills. Unless funding were ringfenced, we remain to be convinced that it should be for the regulator to make decisions about who and what is funded in terms of education and awareness raising (especially as it would no doubt have a limited budget).

44. Thank you for the opportunity to comment. We look forward to continued engagement with DIA and other partners to help shape and perfect these proposals.

Netsafe
17 August 2023