TE MANA WHAKAATU
Classification Office

Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill Submission

April 2021

Introduction

1. The Classification Office welcomes the opportunity that this Bill presents to discuss much needed change in media regulation in New Zealand. Digital technologies such as livestreaming are emerging and evolving rapidly. At the same time, our understanding of the kind of threats presented by terrorism and violent extremism is undergoing significant change.

2. Accordingly, any legislative proposals to update our media regulatory framework will face the challenges presented by a highly dynamic and uncertain environment. For that reason, we very much agree with the Minister of Internal Affairs' open approach to inviting improvements at the Select Committee stage. We all want an Aotearoa that is free from harm, where abhorrent and illegal extremist and terrorist content online is dealt with quickly and appropriately and where rights to freedom of expression are limited only to the extent necessary in a free and democratic society.

3. In this area any legislative endeavour to balance the need to protect people from harm while protecting their essential human rights and freedoms are exactly that – a balancing act. The optimal point of balance will always be up for debate in a free and democratic society such as Aotearoa, but in our submission we outline how historically agreed approaches to regulating content have now been thrown into doubt with the inexorable rise of digital technologies and the global internet. This means that every opportunity should be taken to test and improve existing legislation (as is the objective of this Bill), while also looking to reform and renew the fundamentals of the media regulatory framework that in this country was largely designed in the 1980's and 90's, and which worked well for the types of media and content prevalent back then.

4. In considering the proposals contained in this Bill, the Classification Office has endeavoured to draw on four key touchstones, which we have used as a framework to guide our thinking. The touchstones are:
   - **Experience**: The Classification Office has been a central independent agency involved in the response to all of the major content harm events (whether terrorist, violent extremist, or other) that have substantially impacted New Zealanders in the past two years (and prior). Drawing on this experience, we have sought to ask the questions "*what is needed now*?" and "*what works*?"
   - **International Insight**: As part of our upskilling in the terrorist/violent extremist content area, we have been increasingly active in engaging with policy makers, regulators, NGOs and academics active in this space in New Zealand and around the globe. Most countries are wrestling with these issues, and to a large degree, we are talking about challenges that are inherent in an internet that mostly ignores national boundaries. We have sought to apply this insight to the current proposals.

- **The Christchurch Call**: The Christchurch Call is a ground-breaking approach to addressing the complex problem of online extremism, which was thrown into dreadful prominence by the attacks of March 15. The Call is a unique accord between sovereign states and global technology companies, drawing on the community of purpose and shared rejection of the terrorist's vicious and calculated weaponisation of digital technology. While the Call is not legislation, we consider that many of the principles enunciated in it provide a useful reference point for evaluating the Bill – such as the fact that the Call "rests on the conviction that a free, open and secure internet offers extraordinary benefits to society. Respect for freedom of expression is fundamental. However, no one has the right to create and share terrorist and violent extremist content online." One clear positive outcome from the Call is the GIFCT CIP protocol.[1]
- **The Royal Commission**: We also have the benefit of the Royal Commission of Inquiry into the Terrorist Attack on Christchurch Mosques ("**the Royal Commission**"). While the Royal Commission did not examine the follow-on impacts and consequences of the attacks (including the livestream "going viral" and being shared around the world), the Classification Office feels that many of the insights and recommendations contained in the Royal Commission's report[2] can be applied to testing the proposals contained in the Bill. We feel that the report supports a conclusion that an effective strategy in this area means having a whole-of-government approach to building social cohesion and inclusion, while closely collaborating with civil society, local government, and the private sector. It requires research, data and careful evaluation, as well as public education and discussion.

5. When we apply these four touchstones to the issues at hand, we feel that change can, and should, be made to our existing legislation in this area – but that such change should not only focus on making our existing legislation more effective in these challenging areas, but should also aim to improve transparency of decision-making and access to rights of independent review. It should build trust across industry, civil society, and vulnerable groups. Having those joint goals will put us in the best possible position to undertake the more fundamental media regulatory review that also needs to be undertaken. We have reviewed the proposals in the Bill in light of this philosophy.

6. All of our reference touchstones reinforce for us that the nature of this challenge means that it is not one that can be resolved by policy-makers without listening to those who have most at stake – those who have been subject to ideological extremism, those who have felt the consequences of unbridled hatred, or those who have suffered trauma vicariously. We know that representatives of these groups will be seeking to make submissions on this Bill and we would suggest that it is those voices that need to be paid closest attention to.

[1] This is a process by which GIFCT member companies become aware of, quickly assess, and act on potential content circulating online resulting from a real-world terrorism or violent extremist event (e.g. by removing it from the platform). GIFCT includes the major platforms, including Facebook, Microsoft, Twitter and YouTube. The CIP was created in April 2019 and announced in July 2019 in response to the tragedy in Christchurch (a follow on action from the Christchurch Call). We understand that its current use is limited to livestream footage of a terrorist event. We have seen it effectively activated after the Halle terrorist attack and the Glendale shooting livestream.

[2] https://christchurchattack.royalcommission.nz/the-report/.

About the Classification Office

7. The Classification Office Te Mana Whakaatu is an independent Crown entity responsible for classifying publications[3] that may need to be restricted or banned (an independent media regulator at arm's length from Government). Our work includes examination and decisions on both commercial content and criminal content, and a research, outreach and education function (including a public inquiries and complaints service).

8. The Classification Office is led by a Chief Censor and a Deputy Chief Censor (position currently vacant). Both are statutory appointments under our authorising legislation: the Films, Videos, and Publications Classification Act 1993 ("**the FVPCA**"). The Chief Censor has certain roles, responsibilities, and powers under the FVPCA and is the Chief Executive for the Classification Office. Classification decisions are deemed to be expert decisions under the FVPCA and there is an independent appeal process (see **Appendix A**).

9. Budget 2020 committed approximately an additional $17M over a four year period to build capability across the Department of Internal Affairs and the Classification Office in this area. Of this sum, just over $2M was allocated to the Classification Office for four years, enabling us to establish a dedicated three-person team working on countering violent extremism, with a focus on online content promoting terrorism and violence.

10. The new Countering Violent Extremism (CVE) team in our Office has a specialist classification role, as well as a research, education and outreach function and we're proactively engaging with New Zealand and overseas government agencies, academics, and experts at the forefront of countering violent extremism, to share insights and identify solutions. This year's major research project for the Classification Office is a nationally-representative survey of the New Zealand public on the topic of misinformation, disinformation, and the relationship to conspiracy theories, extremism and real-world harms.

11. On 15 March 2019, people around the world watched a livestream video of a terrorist killing 51 Muslim men, women and children in Christchurch mosques. The livestream video was amplified by algorithms on major platforms such as Facebook and YouTube, and even reached victims' family members and friends. It was a horrific wake-up call as to how digital technology can be weaponised in new and devastating ways. The Chief Censor and the Classification Office were responsible for deciding whether this livestream video (and related *Great Replacement* document) should be banned under the Act. We know that public statements around the classification of the livestream and the Great Replacement document served an important role in informing both public and industry handling of, and responses to, this material – and we have evolved our internal processes considerably in light of this and subsequent experiences with global online harm events. A discussion of this process and changes since this event are outlined further below.

[3] The legal definition of a 'publication' covers a wide range of mediums such as films, videos, music recordings, books, magazines, video games and online content.

Traditional Approaches to Regulating Harmful Content Online

12. Traditionally, media content has been produced, marketed and consumed in physical, broadcast and print formats. Regulation of potentially harmful effects of media has typically reflected those broad categories, with the FVPCA principally aimed at 'physical format' media such as film reels, VHS tapes and DVDs. This type of media is amenable to a detailed classification process – either ahead of publication (as in the case of cinema release films or commercial DVDs), or once an enforcement authority such as Police or the New Zealand Customs Service has encountered material that they consider may breach the harm provisions in the FVPCA.

13. Broadcast media is not amenable to a formal 'classification' process as the format is dynamic and often 'live'. The Broadcasting Standards Act 1989 reflects that reality by requiring that broadcasters comply with standards (dealing with aspects such as harm, balance, fairness, etc), with avenues for the public to pursue complaints in the event that they consider that a broadcaster may not have met those standards. Similar standards approaches are taken by the Media Council (formerly the Press Council) and the Advertising Standards Authority industry bodies responsible for undertaking self-regulation in their respective areas.

14. A slightly different approach is taken in the Harmful Digital Communications Act 2015 ("**HDCA**"), which is a relatively new (and digitally-focussed) piece of law. The HDCA uses a mixed model - both mediation and enforcement at the District Court, intended to be a faster redress process with low-level penalties for redressing harm against individuals.[4] Like the Privacy Act 1993, the HDCA is designed around 10 communication principles - for example, "a digital communication should not disclose sensitive personal facts about an individual, and a digital communication should not be threatening, intimidating or menacing."[5]

The Challenge

15. The challenge we face today is harmful content that is often not individually targeted (as per the HDCA), nor is it 'physical media' of the kind that the FVPCA is predominantly designed to address. It is also not broadcast media or press content that may be captured by one of the existing regulatory or industry bodies. We are talking about seriously harmful and objectionable (unlawful) content that is produced, shared and propagated on social media, internet forums and digital platforms of all kinds.

16. The challenge that viral, harmful online material presents for a traditional classification regulatory approach is significant. The issues are broadly the speed and reach of digital dissemination; the sheer volume of material that is involved; and the fact that a very significant volume of material may be harmful but not unlawful under the FVPCA.

Speed of Dissemination

17. Dangerous disinformation, terrorist/violent extremist content and other harmful publications are quickly amplified online. This is for a range of reasons, for example:

    - Algorithms: which recommend and expose users to targeted content that aligns with their demographics and content preferences, and influences who they engage with online - this

[4] Section 3, HDCA.
[5]

can be, for example, what content automatically populates on a person's profile page, directly recommends pages, groups or content to users, and in some cases has actually meant auto-creating pages for extremist groups or content;

- Technical functions on the platforms that accelerate the speed of exposure, such as the "Share", "Like" and livestream functions that can be operated by any user;
- "Influencers" and "subscribe" functions: for example, an influencer with a large group of followers on social media can potentially expose millions of other users to harmful content in one go. This sits alongside a commercial incentive for building followers (and advertising revenue), which can in turn encourage extreme content to be shared.

18. The livestream video from the March 15 attack is one of the most if not the most extreme known example of a terrorist publication that went viral quickly and in real time. We know that people from all over the world were exposed to this footage on digital platforms and social media sites such as Facebook, and that algorithms contributed to it being spread virally; it even reached victims' family members and friends. It was a horrific wake-up call to how digital technology can be weaponised in new and devastating ways.

Volume of material

19. Alongside the challenge of the speed and spread of digital viral reproduction, these technologies facilitate the production of huge volumes of material. We saw this in the hours following the initial March 15 livestream, where within a few hours of the event, hundreds (possibly thousands) of variations of the livestream video were shared, with edits, time distortions, sound track alterations and other changes – often designed to evade AI hash identification on the various platforms. We have seen this pattern continue with subsequent extremist events.

20. Each variant of the livestream video is technically a separate publication in its own right, with the potential for a separate classification decision under the FVPCA. Alongside the minor variants of the livestream, we saw innumerable media reports and private posts (many expressing shock and abhorrence at the attack) which included still images or short excerpts from the livestream. Each of these were also technically publications under the FVPCA, with the potential to be classified (although the context of this category of publication would mean that many or most would not be objectionable).

21. In short – the speed, spread and volume of content involved in a digital viral terrorist event presents fundamental challenges for a traditional, human-based classification regulatory model. Classifying content (particularly traumatising content recording real-world events that may have importance as evidence or news reporting) takes expertise, an understanding of context, and time.

Harmful but not unlawful

22. Since the March 15 attacks we have seen and assessed (if not formally classified) multiple other products involved violent extremism, terror, crime and suicide – all of which involving some online or digital element, and their own contextual complexities and issues. Examples include the livestreamed Halle synagogue extremist attack (October 2019); a Facebook post of a video of an attack on a cell phone tower in NZ (April 2020); a livestreamed attack in Glendale, Arizona (May 2020); the bystander video of the killing of George Floyd in Minneapolis (May 2020); a video of a

livestreamed suicide that went viral on the TikTok platform popular with NZ children (September 2020); the killing of Ashli Babbitt, filmed from four different devices during the storming of the Capitol in Washington (January 2021); 'bystander livestream' of victims of the mass murder in Boulder, Colorado (March, 2021). Some of these publications were objectionable under New Zealand law. Others were troubling, potentially harmful to children, but did not present the same immediate and serious risk of harm that would warrant an objectionable classification. Some had significant value in recording significant crimes or abuse. Many required some element of urgent response and assessment. All of this material is challenging, and demands a careful, considered approach.

Response to the Proposals in the Bill

23. The challenges outlined above show that broad media reform is needed and overdue. While this should not stop urgent and critical amendments happening in the interim (as is proposed in the current Bill), we recommend that overall progress is made intelligently with a connected strategy, drawing on the key principles of the Christchurch Call and the Royal Commission's report as well as insights drawn from experiences and initiatives currently being undertaken overseas. We have reviewed the proposals in the Bill in light of this philosophy.

Livestream offence and related clauses

24. The Bill proposes to introduce new definitions and a new offence targeting the action of livestreaming of objectionable content. The Explanatory Note to the Bill makes the following key points about this proposal:

    - "Livestreaming is the online transmission of events in real time. Any digital reproduction of a livestream is a recording and is therefore subject to existing provisions in the Act relating to publications."

    - "The criminal offence of livestreaming objectionable content only applies to the individual or group livestreaming the content. It does not apply to the online content hosts that provide the online infrastructure or platform for the livestream."

    - "The difference between a copy of livestreamed content and livestreaming itself means that the latter is not subject to the provisions of the principal Act except where expressly included by the Bill. Livestreaming is expressly included for the purposes of *new Part 7A* (*see* definition of online publication in *new section 119A* in *clause 9)* and in *new section 124AB*, which establishes a new offence of livestreaming objectionable content (see *clause 10*). There's a connected amendment to section 131(2B) (*clause 11*)."

25. The criminal offence relates to the physical action of livestreaming content (as opposed to sharing that content) and the offence has a mens rea of "knowing or having reasonable cause to believe that the content was objectionable". Clause 10, new section 124AB.

26. The penalty for the offence is a maximum term of imprisonment of 14 years for an individual and a fine not exceeding $200,000 for a body corporate. Note that this correlates with the penalties for existing offences, such as to distribute a publication knowing (or having reasonable cause to believe) that the publication is objectionable (section 124 of the Act).

Classification Office Comments

27. We understand that this proposal is intended to clarify the law and specifically ring-fence the action of livestreaming in addition to the existing law which regulates sharing of livestreamed footage. The Classification Office notes that in our experience the act of livestreaming to a digital or social media platform (such as Facebook) ordinarily creates multiple digital copies of the livestream, across multiple devices. That was the basis for legal classification of the March 15 livestream (and subsequent livestreams) which was upheld on de novo review by the Film and Literature Board of Review, and which has also been relied on in multiple subsequent prosecutions for possession of a copy of the livestream (typically digital).

28. Accordingly the Classification Office submits that, in our interpretation of the definition of publication under the FVPCA conjoined with the offence provisions contained in that Act (notably section 124) in the vast majority of cases a livestream that is objectionable will also create an unlawful digital publication. We would note that typically the Classification Office will not be viewing an objectionable publication as it is streamed in real time, and therefore we will be classifying and dealing with some form of digital copy (which would also serve to activate the offence provisions under the FVPCA).

29. However, the Classification Office has not had the benefit of seeing any legal analysis relied on by officials in recommending this change, and we accept that circumstances could conceivably arise that would mean that the position in relation to an objectionable livestream should be put beyond doubt. We note some commentators have suggested that explicitly making livestreaming an offence risks criminalising innocent individuals who inadvertently capture objectionable footage, but we do not think that the proposed change would result in a meaningful increase in such a risk, given our interpretation of the FVPCA as it stands, and the low likelihood that such inadvertent or legitimate bystander footage would have the promotional elements we look for in an objectionable publication.

Urgent Interim Classification Assessment

- The Bill proposes to introduce a new urgent interim classification process and assessment. It is proposed that this new process could apply to all publications covered by the principal Act not just online publications.
- Under this proposal, the Chief Censor would have the ability to make an urgent interim classification assessment, which would have the full legal effect of a full classification decision (e.g. the penalty for distributing a publication knowing that it has been assessed as objectionable under this interim process would be a maximum term of imprisonment of 14 years for an individual and a fine not exceeding $200,000 for a body corporate).
- "The Chief Censor can exercise this power if he or she believes that there is an urgent need to notify the public that the content of the publication is likely to be objectionable (on the basis of the interim assessment) and to limit harm".[6]
- The urgent interim assessment expires after the earliest of 20 working days or when a classification decision has been made.

[6] Explanatory Note to the Bill, pg. 5.

- There is proposed immunity from civil/criminal liability provided to officials for actions taken in making an interim assessment.
- There is proposed immunity from civil/criminal liability provided to ISPs and online content hosts who remove or prevent access to an online publication that is the subject of an interim assessment.
- The reasons for the interim classification assessment must be provided in the notice of the subsequent classification decision.

This policy proposal for urgent interim classifications appears to be based on an assumption that the process for classifying the livestream footage and Great Replacement manifesto was unduly cumbersome for digital media in an emergency situation, though we note that the urgent interim proposal is intended to apply to both traditional and digital media. The Regulatory Impact Assessment to the Bill states that:

> "… the Chief Censor must publish a written decision within five working days of classifying a publication, which can delay initial decisions on objectionable content… following these procedures exactly can take time and does not suit situations where the availability of a publication is likely to be injurious to the public good and there is an urgent need to notify the public of this harm – particularly in the online sphere where media can 'go viral' quickly. In these situations the Chief Censor may not necessarily have the time, for example, to fully justify, present all evidence, and document the decision as is required by the current statutory processes or may need to reallocate resources to do this that would otherwise be employed elsewhere, thus compromising other decisions OFLC is required to make. As a consequence, the Chief Censor may have to delay public notification in relation to a given publication. The 15 March terrorist's manifesto was a lengthy complex document and the Chief Censor had to consider delaying his classification to meet the five-day requirement."[7]

We have provided a diagram of our understanding of this proposed process in **Appendix A.** We have checked this process diagram with the Department.

Classification Office Comments

30. The Classification Office submits that the FVPCA should be amended to extend the time required to provide a written classification decision. However, we suggest changes be made to the Bill to provide for maximum flexibility and certainty, as circumstances require.

    The Classification Office has faced and continues to face a wide variety of challenges involving critical, real-time unfolding online events that have the potential for real harm, and which also may be classified as objectionable. There is a crucial balance to be struck between acting swiftly enough to be timely and effective, and avoiding acting instinctively or capriciously. In the hours following the attacks on March 15, the Chief Censor personally reached out to front-line agencies and enforcement to provide support and advice. Coordination of the response continued through following days, with the livestream being provided to the Chief Censor for initial review on Saturday 16 March, with the decision by the Chief Censor to "call in" the livestream under the

[7] *Regulatory Impact Assessment: Countering violent extremism online – changes to censorship legislation to better protect New Zealanders from online harm* (Department of Internal Affairs, 1 June 2020), pp 9-10, https://www.dia.govt.nz/diawebsite.nsf/Files/Proactive-releases/$file/regulatory-impact-assessment-countering-violent-extremism-online.pdf.

FVPCA occurring on Sunday 17 March. The Chief Censor convened an emergency classification meeting with the Deputy Chief Censor and Senior Advisers, on Monday 18 March, when the decision to classify the livestream as objectionable was discussed, finalised and publicly announced on the same day – as far as is known this was at the time the fastest classification ever completed. Some subsequent urgent classifications and advice on terrorist publications and 'livestreams' have happened even more quickly – for example the classification decision on the objectionable Halle attack livestream was completed within hours of initial notification.

31. The critical learning of the Classification Office over the past two years is that there are some requirements for providing a swift, but properly robust and considered, final classification decision in cases such as these. Initial advice can properly be provided to help guide enforcement and response from agencies in real time. A final decision can be provided extremely quickly, once the complete publication (whatever it may be) has been received, reviewed and discussed by the Chief Censor and/or Deputy Chief Censor and at least one other senior classification adviser.

32. Given that (if those core requirements are met) a final classification decision can be provided extremely quickly, the preference of the Classification Office will always be to provide a final classification decision, over an interim assessment. We have seen how, in the case of March 15 and subsequently, a clear communication about a formal classification can help the public to quickly understand the situation and what is expected of them.

33. The difficulty that then arises however is the current requirements under the FVPCA that the Classification Office provide written reasons within five days of making a (final) decision. As discussed above, we have learned that crisis events may often involve multiple publications or products, and the reasons for classification, while clear to us, may necessitate a detailed written opinion. In those circumstances, multiple decisions relating to a single event may place unreasonable pressure on the Classification Office to produce multiple high-quality written decisions within a five day window.

34. Our reservation therefore is that the Bill provides for an additional 20 day period for producing the written reasons for the decision, but this applies only when an interim assessment is made. This potentially creates a perverse incentive to opt to make an interim assessment, even where the Office would otherwise make an ordinary, final decision - if only to take advantage of the additional time attached to provisional decisions. This is undesirable, in our view. If it is in a position to make a swift, final decision – the Classification Office should. Particularly in critical or urgent situations of the nature under discussion here.

35. Drawing on our working level of understanding of international response process and empowering legislation in these areas, we have not identified any analogous interim classification power either operating overseas, or currently under consideration. As outlined above, formal legal classification processes are fundamentally predicated on an expectation that they will involve human beings making often complex assessments of material that may be ambiguous in presentation and intent. The context for material is also extremely important and often needs to be paid due attention (which is something we feel the work of the Royal Commission also supports). All of this takes time, but as has been demonstrated in responses to recent critical events, can be undertaken effectively and rapidly if circumstances allow.

36. The Classification Office accordingly recommends that the 20 day period for providing written decisions be extended to 20 days for final decisions. With that change, the Classification Office does not see a need to provide for interim assessments, and we have identified a risk that conferring an interim assessment power could potentially result in undue pressure being placed on the Office to make capricious or partly-informed decisions in the height of a crisis response. That must not happen. In the event that the power to make interim decisions is retained in the Bill, we would strongly suggest that such a power also be applied with clear protocols between the Classification Office and submitting/enforcement agencies to ensure that the application and use of such a power is fully understood and subject to clear safeguards to protect the integrity of the classification process.

Take-down Notices

37. The Regulatory Impact Assessment states that:

> "DIA has no explicit power to request and / or enforce online content hosts to remove objectionable content from their platforms. Neither a take-down power for objectionable content, nor compliance measures to enforce it, were necessary in a pre-Internet age. The current process for requesting the removal of objectionable content is to advise online content hosts that they may be committing an offence under New Zealand law if they do not remove the content. To date, this process is generally effective, but does not provide certainty for either Government or online content hosts and relies on goodwill and cooperation which may not necessarily be the case should similar events occur again. Following the Terror Attacks, the Government could not point to clear legislation stating that online content hosts' failure to remove the video was illegal and had no legislative mandate to compel removal of such content to prevent its spread causing harm. Companies that did comply with requests to remove content were operating without legal certainty."[8]

38. The Bill proposes to introduce a new "take-down" power for enforcement officers. The Explanatory Note states that:

> "The take-down powers are aligned with current powers of seizure of objectionable publications under the Act… The Bill authorises an Inspector of Publications (excluding constables) to issue take-down notices for particular online publications."

39. A take-down notice would be issued to an "online content host", and direct the removal of a specific link so that it is no longer viewable in New Zealand. This must be complied with as soon "as reasonably practicable".

40. A take-down notice can be issued for an online publication if:
    - An interim classification assessment has been made that the publication is likely to be objectionable;
    - The online publication has been classified as objectionable; or
    - The Inspector believes, on reasonable grounds, that the online publication is objectionable.

[8] *Regulatory Impact Assessment: Countering violent extremism online – changes to censorship legislation to better protect New Zealanders from online harm* (Department of Internal Affairs, 1 June 2020), pg. 10, https://www.dia.govt.nz/diawebsite.nsf/Files/Proactive-releases/$file/regulatory-impact-assessment-countering-violent-extremism-online.pdf.

41. A take-down notice, connected to an interim assessment, has effect for 20 working days but becomes permanent if a classification decision is made that the publication is objectionable. A take-down notice based on the other two grounds becomes permanent immediately.

42. There is proposed to be explicit immunity from civil and criminal liability for officials and online content hosts relating to actions when issuing or complying with a take-down notice. Online content hosts may be required, under the notice, to retain a copy of the publication for the purposes of an investigation or proceeding.

43. The Explanatory Note to the Bill states that:

> "… it is intended (but not required by the Bill) that the authority to issue a take-down notice will only be exercised in situations where other options for seeking the removal of objectionable content online has proven ineffective. The current collaborative practice of requesting online content hosts to voluntarily remove identified objectionable content will continue to be the first and preferred approach."

44. A takedown notice is only reviewable under the Act through the classification process (eg if a classification decision found that the content was not objectionable). New section 119I sets out proposed remedies and costs for failing to comply with a take-down notice, including a pecuniary penalty of up to $200,000.

45. There are certain reporting requirements for the Secretary of Internal Affairs listed in new section 119K, including making publicly available a list of all take-down notices that have been complied with, and publishing the number of notices issued and the number complied with in the annual report.

46. We have provided a diagram of our understanding of this proposed process in **Appendix A.**

Classification Office Comments

47. We note that this proposal has been developed to provide tools for enforcement officers to restrict objectionable content that is available online (analogous to the existing power for an inspector to go into a shop and seize an objectionable publication). Obviously the task of trying to remove objectionable content from the internet is several orders of magnitude beyond that of policing traditional media at physical premises in New Zealand. In this context the balancing act between acting effectively to protect people from harm while also protecting freedom of speech and preserving the ideal of a free, open and secure internet is more challenging. Bearing that in mind, the Classification Office supports in principle the proposal to establish a clear and prescribed take down notice power, given the current realities of objectionable content on the Internet. We know of Internet sites who make a feature of hosting unquestionably objectionable content (such as the March 15 livestream video) and who will make a point of refusing reasonable requests by New Zealand enforcement agencies to remove this content, even if only for New Zealand viewers. This power, alongside a well-configured and specified filter provision, could provide a suitable escalation pathway and response to that difficult issue.

48. Earlier in the policy development process the Classification Office advised that we thought it important that the primary legislation explicitly include a right of review/appeal, and that information about these rights should be included in any take-down notice that is issued. We support the specification of these rights and the requirement to provide notice of them in the current Bill. We do, however, consider that further enhancements can be made in this area:
    - there could be additional and clearer explanation in the primary legislation specifying what this right of review is, as opposed to relying on the recipient reading Part 4 of the FVPCA (which relates to classification decisions, rather than take-down notices);
    - As it is currently drafted in new section 119J, a person who receives a takedown notice would only have a right of review to the Film and Literature Review Board after a classification decision has been made. We note that it is not a given that a classification decision will ever be made in relation to the material to which the take down notice relates. The Bill should explicitly provide for a review pathway in that eventuality.
    - We recommend that the new section 119C(6) should be amended to put it beyond doubt that the ability to use the "call in" power under section 13(3) will not be limited to situations where an Inspector issuing a take-down notice is based on "reasonable grounds". We can forsee situations where, for example, publications that may have been historically classified as objectionable will merit reconsideration via the call-in power. A number of publications historically classified as objectionable might be reclassified if assessed under the current framework.
    - We recommend that the Bill specify additional baseline information to be provided to the Chief Censor regarding take down notices issued. An URL link (that may no longer be able to be accessed) and notification to the Classification Office without a copy of the publication or reasons why it was subject to a take-down notice will be insufficient information to make an informed choice about whether the publication(s) should be called in for classification. Again we see this as supporting robust process and basic levels of transparency to help build trust in the system and ensure that interventions are informed. We see this approach as consistent with both the principles set out in the Christchurch Call and the comments on the need for active and less siloed approaches made in the Royal Commission's report.

Electronic System – "the filter"

49. The Bill proposes to introduce a new scheme, giving the Government the power to block objectionable content online. The Regulatory Impact Assessment states that:

> "Government currently works cooperatively with ISPs to block child sexual exploitation material – an obvious form of objectionable online content where censorship decisions are clear-cut – via the Digital Child Exploitation Filtering System (DCEFS). Consideration of the broader spectrum of objectionable online content, all of which is illegal, deserves a clear legal framework. There is currently no statutory authority in primary legislation to support robust and transparent consideration and development of mechanisms to filter and / or block objectionable online content.
>
> The absence of a legal framework also puts ISPs in an uncertain legal position, where currently they are making voluntary decisions to remove content from their customers' services and can be seen by their customers as censoring content. ISPs have stated they feel uncomfortable with potential legal liability for operating in this role. In the wake of the attacks, some ISPs raised concerns on this

issue and continue to request greater Government support to identify what, and how, content should be blocked."[9]

50. The Bill essentially sets out a "skeleton framework", enabling an internet filtering system to be established, with the detail to be in Regulations. The filter would be operated by the Department of Internal Affairs and would potentially mean that the New Zealand public could be blocked access to a particular online publication, website/webpage hosting a publication that:
    - Has an interim assessment as objectionable
    - Has been classified as objectionable
    - An inspector of Publications believes on reasonable grounds to be objectionable.

51. The filter regulations would set out:
    - The criteria for identifying online publications that are likely to be objectionable and for preventing access to those publications by the filter;
    - Governance arrangements for the filter;
    - Requirements for the administration and technical oversight of the filter, including data security and privacy;
    - Reporting requirements for the filter system;
    - Obligations of ISPs relating to the operation of the filter;
    - Review and appeal processes for the decision approving the operation of the filter; and
    - Review and appeal processes for decisions and actions relating to the operation of the filter.

When deciding on the design and form of the filter, the Secretary of Internal Affairs must consider:
- The need to balance any likely impact on public access to non-objectionable online publications and the protection of the public from harm from objectionable online publications;
- Any likely impact on performance for all other network traffic;
- Departmental and technical capacity to operate the system
- Likely compliance costs.

Under the Bill, the filter must be able to both identify and block access to a particular publication with reasonable reliability – based on criteria in regulations.

Classification Office Comments

52. We support the concept of a properly specified and carefully applied internet filter that could serve to block New Zealand access to known and classified objectionable publications. As noted above, our experience shows that there are websites whose sole purpose is to be part of an ecosystem promoting publications with violent extremist and terrorist content classified as objectionable, such as the March 15 livestream video. We know that these websites are unlikely to respond favourably to a request from a DIA Inspector to remove the content, either informally or via a take-down notice, so a filter becomes an escalation point to ISPs to block that content. These websites are not on the dark web but are able to be located via a simple search. In these

[9] *Regulatory Impact Assessment: Countering violent extremism online – changes to censorship legislation to better protect New Zealanders from online harm* (Department of Internal Affairs, 1 June 2020), pg. 11, https://www.dia.govt.nz/diawebsite.nsf/Files/Proactive-releases/$file/regulatory-impact-assessment-countering-violent-extremism-online.pdf.

situations, we think that there is a strong argument that a carefully applied filter will help to reduce harm to the New Zealand public, including young people, of inadvertently stumbling across this objectionable content (while accepting that a few determined seekers will continue to find it, e.g. via a VPN). We have seen this filter model work effectively as a tool for an independent media regulator in Australia (ie the e-Safety Commissioner).

53. However, the Classification Office believes that the Bill requires much more detail in the primary legislation, as opposed to delegated to the Executive to choose in Regulations, for how this filter is going to operate in practice and what protections are guaranteed to be built into the system to guide both the design and the practice. We think this is necessary both to structurally build in essential limits and safeguards, ensure protection of fundamental rights for New Zealanders into the future, and to build public trust and confidence that the system will only be used when necessary and will operate in a transparent and easily reviewable way. Our view is that at a basic level there needs to be transparency about when this potentially very powerful tool is used, for what content, and that a clear review and appeal pathway should be provided for in primary legislation. We note that this is approach is also consistent with the advice from the Legislation Design Advisory Committee[10].

54. The Classification Office is not aware of an analogous approach to that proposed in this Bill adopted in any overseas jurisdiction. We would further observe that the placement of key checks and balances, transparency and review mechanisms in primary legislation appears consistent with the principle set out in the Christchurch Call that governments should enact regulation that is consistent with a free, open and secure Internet as well as international human rights law.

Removing the "safe harbour" provisions – Harmful Digital Communications Act 2015

55. The Bill proposes to remove the immunity that exists for online content hosts because the current law:

> "…creates potential for online content hosts being exempt from any criminal or civil liability if they break the law under the FVPC Act but follow the procedures in the HDC Act. This undermines enforcement efforts around objectionable content. Analysis following the Terror Attacks identified that online content hosts could simply have notified that uploader of the Terrorist's video, waited two days to take it down, and be exempted from criminal liability under the FVPC Act."[11]

56. The Explanatory Note to the Bill notes that:

> "this will mean that enforcing the new offence or modified offences in the [FVPCA] will not be limited by the HDC Act safe harbour provisions for online content hosts. It will ensure that online content hosts can be prosecuted for hosting objectionable content if they are liable for doing so."

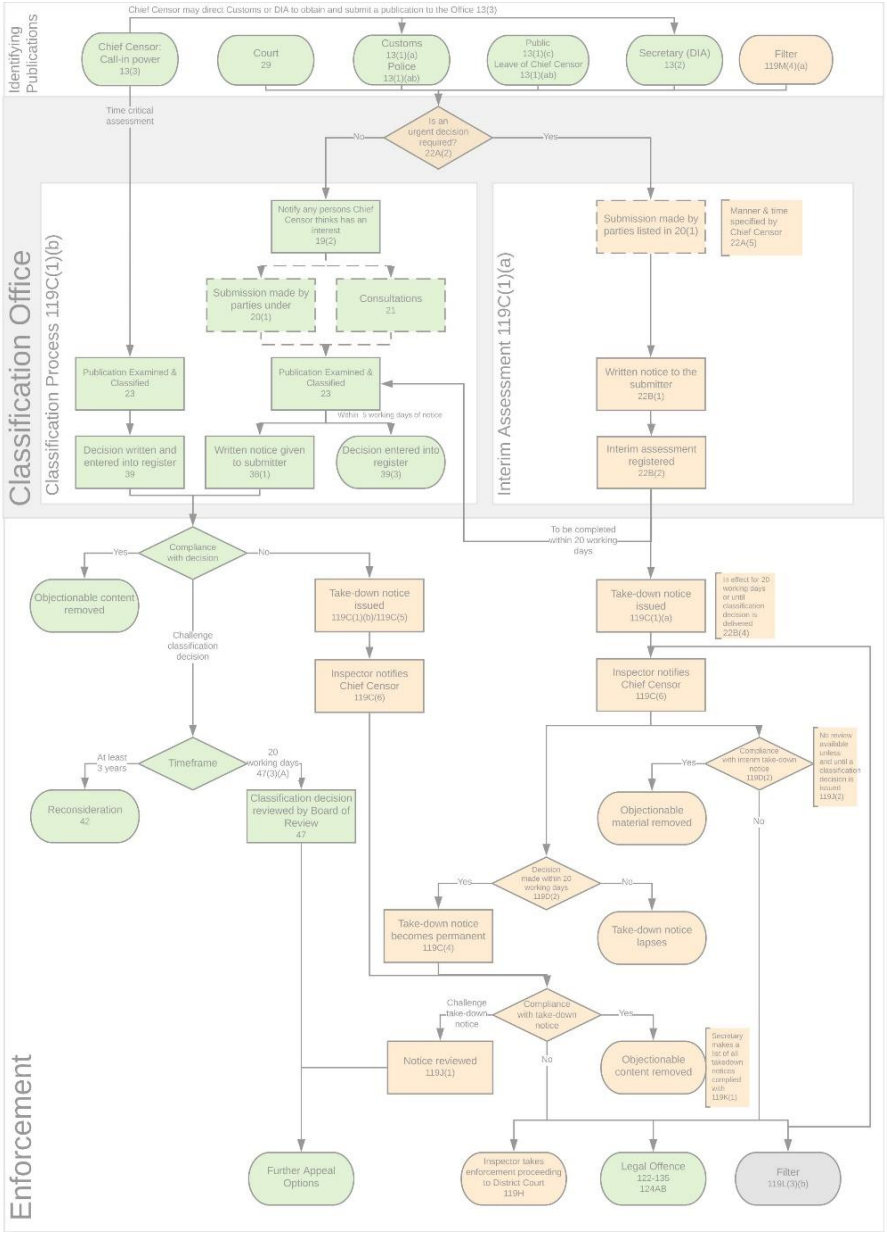
[10] www.ldac.org.nz/guidelines/legislation-guidelines-2018-edition.
[11] *Regulatory Impact Assessment: Countering violent extremism online – changes to censorship legislation to better protect New Zealanders from online harm* (Department of Internal Affairs, 1 June 2020), pg. 10, https://www.dia.govt.nz/diawebsite.nsf/Files/Proactive-releases/$file/regulatory-impact-assessment-countering-violent-extremism-online.pdf

Classification Office Comments

57. While the Classification Office is not aware of any content host invoking the HDC Act 'safe harbour' provisions as a defence to prosecution under the FVPCA, we accept that the changes in this Bill and in the internet regulatory landscape generally appear to increase the risk that these provisions may be invoked. We make no further comment on this part of the Bill.

Oral Submission

58. The Classification Office requests the opportunity to appear before the Select Committee to present an oral submission.


Identifying Publications
Chief Censor may direct Customs or DIA to obtain and submit a publication to the Office 13(3)
Chief Censor: Call-in power 13(3)
Court 29
Customs 13(1)(a) Police 13(1)(ab)
Public 13(1)(c) Leave of Chief Censor 13(1)(ab)
Secretary (DIA) 13(2)
Filter 119M(4)(a)
Classification Office
Classification Process 119C(1)(b)
Time critical assessment
Is an urgent decision required? 22A(2)
No
Yes
Notify any persons Chief Censor thinks has an interest 19(2)
Submission made by parties under 20(1)
Consultations 21
Publication Examined & Classified 23
Publication Examined & Classified 23
Within 5 working days of notice
Decision written and entered into register 39
Written notice given to submitter 38(1)
Decision entered into register 39(3)
Interim Assessment 119C(1)(a)
Submission made by parties listed in 20(1)
Manner & time specified by Chief Censor 22A(5)
Written notice to the submitter 22B(1)
Interim assessment registered 22B(2)
To be completed within 20 working days
Enforcement
Compliance with decision
Yes
No
Objectionable content removed
Challenge classification decision
Take-down notice issued 119C(1)(b)/119C(5)
Inspector notifies Chief Censor 119C(6)
Timeframe
At least 3 years
20 working days 47(3)(A)
Reconsideration 42
Classification decision reviewed by Board of Review 47
Further Appeal Options
Take-down notice issued 119C(1)(a)
Is effect for 20 working days or until classification decision is delivered 22B(4)
Inspector notifies Chief Censor 119C(6)
Compliance with interim take-down notice 119D(1)
No review available unless and until a classification decision is issued 119J(2)
Yes
No
Objectionable material removed
Decision made within 20 working days 119D(2)
Yes
No
Take-down notice becomes permanent 119C(4)
Take-down notice lapses
Challenge take-down notice
Compliance with take-down notice
Yes
No
Notice reviewed 119J(1)
Objectionable content removed
Secretary makes a list of all takedown notices complied with 119K(1)
Further Appeal Options
Inspector takes enforcement proceeding to District Court 119H
Legal Offence 122-135 124AB
Filter 119L(3)(b)
Identifying Publications
Public 13(1)(c)
Secretary (DIA) 13(1)(b)
Enforcement & Security Agencies
Filter 119M(4)(a)
Inspector of Publications
Reasonable grounds that the publication is objectionable 119C(1)(c)
Seek voluntary removal of the publication 119C(2)
Permanent Take-down notice issued 119C(1)(c)
No review is available unless and until a classification decision is issued 119J(2)
Inspector notifies Chief Censor 119C(6)
Chief Censor uses power under s13(9) 119C(6)
Yes
No
Classification process
No classification
Written notice given to submitter 38(1)
Decision entered into register 39(3)
If notice challenged
Notice reviewed 119J(2)
Classification decision reviewed by Board of Review 47
Further Appeal Options
Compliance with take-down notice
Yes
No
Objectionable material removed
A copy of the publication retained by host 119E(2) Host must destroy all copies of publication except in situations listed in 119E(3)
Inspector takes enforcement proceeding to District Court 119H
Legal Offence
Filter 119L(3)(c)
Key
Existing in FVPCA
Proposed changes to FVPCA
More Info Needed
Optional Process
Terminator
Note