

Written Submission on the General Scheme of the Online Safety and Media Regulation Bill

Global Partners Digital response

March 2021

About Global Partners Digital

Global Partners Digital (GPD) is a social purpose company dedicated to fostering a digital environment underpinned by human rights.

Introduction

(1) We welcome the opportunity to provide a written submission on the General Scheme of the Online Safety and Media Regulation Bill to the Joint Committee on Media, Tourism, Arts, Culture, Sport and the Gaeltacht. In this submission, we examine elements of the General Scheme that relate to online safety (Part 4 - Heads 49A to 56) and do not comment on other sections, except to the extent that they are ancillary to or connected to Part 4. GPD recognises the legitimate desire of the Irish government to tackle unlawful and harmful content online, and we believe the majority of the proposals put forward in the General Scheme to be reasonable and sensible. Based on our analysis, however, we believe that particular aspects of the General Scheme, if taken forward in their current form, may still pose risks to individuals' right to freedom of expression and privacy online and could be inconsistent with Ireland's international human rights obligations.

(2) In this submission, we relay our concerns on a "Head by Head" basis and make a series of recommendations on how the proposals could be revised to mitigate these risks. We believe these considerations and recommendations, if incorporated into the final legislation, will help safeguard freedom of expression and privacy online.

Framework for analysis of the General Scheme

(3) Our analysis of the proposals in the General Scheme is based on international human rights law, primarily the International Covenant on Civil and Political Rights (ICCPR).¹ The most relevant human rights impacted by the proposals are the rights to freedom of expression and privacy. Article 19 of the ICCPR guarantees the right to freedom of expression, including the right to receive and impart information and ideas of all kinds regardless of frontiers. Article 17 of the ICCPR guarantees the right to privacy and provides that "no one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence". The rights to freedom of expression and privacy are also protected in other relevant treaties, such as

¹ Ireland ratified the International Covenant on Civil and Political Rights (ICCPR) in 1989.

Articles 13 and 16 of the Convention on the Rights of the Child and Articles 10 and 8 of the European Convention on Human Rights.²

(4) Restrictions on the rights to freedom of expression and privacy are only permissible under international human rights law when they can be justified. In order to be justified, a restriction must meet a three-part test, namely that: (i) it is provided by law; (ii) it pursues a legitimate aim; and (iii) it is necessary and proportionate, which requires that the restriction be the least restrictive means required to achieve the purported aim.

(5) It is important to remember that Ireland’s obligation to ensure that these rights are not unjustifiably restricted exists both in relation to restrictions which stem from the actions of the state itself as well as those caused by third parties, such as private companies. As such, it makes no difference from the perspective of the individual affected whether any restrictions are imposed and enforced directly by the state (e.g. through creating criminal offences which are enforced by the police and the courts) or through third parties, particularly when the third party is acting in order to comply with legal obligations.

Human rights analysis of the General Scheme

Part 2 - Media Commission

Objectives (Head 9)

(6) Head 9 of the General Scheme provides for the objectives of the Media Commission, including: to “ensure that democratic values enshrined in the Constitution, especially those relating to rightful liberty of expression are upheld”. We welcome the inclusion of this objective in Head 9. This reference is critical considering the potential negative impacts on freedom of expression posed by the proposed legislation. However, we recommend that Head 9 be amended to also reference Ireland’s international human rights obligations. Including specific reference to the International Covenant on Civil and Political Rights and the European Convention on Human Rights would further ensure that protecting and respecting the rights to freedom of expression and privacy is one of the Commission's statutory duties

(Recommendation 1).

Recommendation 1: Head 9 should be amended to include specific reference to Ireland’s international human rights obligations, for example by adding, as a further objective, “Ensure that Ireland’s international human rights obligations, as provided for in the International Covenant on Civil and Political Rights and the European Convention on Human Rights, are upheld, especially with regard to freedom of expression and privacy.”

Cooperation with other bodies (Head 29)

(7) Head 29 of the General Scheme provides for the Commission to enter into cooperation agreements with other bodies as it sees fit. We also welcome this provision as risks to human rights may emerge from how the Commission fulfils its mandate, and cooperating with other

² Ireland ratified the Convention on the Rights of the Child in 1992, and the European Convention on Human Rights in 1953.

bodies would help safeguard against these risks. It would therefore be beneficial for the legislation to provide for specific arrangements involving the Irish Human Rights and Equality Commission in the work of the Commission, particularly in the enforcement of its duties and functions (**Recommendation 2**).

Recommendation 2: Head 29 should be amended to include specific arrangements for cooperation between the Media Commission and the Irish Human Rights and Equality Commission. This would ensure the Irish Human Rights and Equality Commission is able to undertake their functions under the Irish Human Rights and Equality Commission Act 2014, and to further assist and review the work of the Media Commission when necessary.

Part 4 - Online Safety

Categories of harmful online content (Head 49A)

(8) Head 49A of the General Scheme sets out four categories of material that are considered to be harmful online content. With regards to category (a), “material which it is a criminal offence to disseminate under Irish (or Union law)” we do not consider the wording to in and of itself raise concerns from a freedom of expression perspective, but believe that it is unclear on the face of the Bill what criminal offences are captured. The explanatory notes state that it would include “a wide range of materials, including: child sexual abuse material; content containing or comprising incitement to violence or hatred; and public provocation to commit a terrorist offence”. We accept the argument that this wording is preferable to exhaustively listing the offences in the Bill itself (so that it can capture new criminal offences created in the future). However, to ensure clarity over the current scope of criminal offences, and therefore to ensure that none would raise concerns from a freedom of expression perspective, the government should be required under the legislation to publish and keep up-to-date a list of criminal offences under Irish and Union law which would be covered by Head 49A(a) (**Recommendation 3**).

Recommendation 3: The government should be required to publish and keep up-to-date, a list of criminal offences under Irish and Union law which would be covered by Head 49A(a). Head 49A should be amended to include such a requirement, for example by adding as a new subsection “The Minister shall publish, and keep up-to-date a publicly accessible list of criminal offences whose scope includes the dissemination of material for the purposes of this section”.

(9) We believe that the scope of category (b) “material which is likely to have the effect of intimidating, threatening, humiliating or persecuting a person to which it pertains and which a reasonable person would conclude was the intention of its dissemination” is too broad. We are especially concerned when it comes to material which might have the effect of “intimidating” or “humiliating” a person. Of these two, humiliating has a particularly low threshold of harm caused. While there will be examples of “humiliating” material which causes a high degree of harm (for example, when it relates to particular racial group and encourages violence or discrimination), there are many other examples of material which could be “humiliating” which cause far less harm, for example, embarrassing photographs or videos of individuals dancing or

under the influence of alcohol. Given that material which is likely to have the effect of persecuting or threatening a person is already captured, meaning that those greater harms will already be within scope, we do not see the need to include material which is merely “humiliating”. We would therefore suggest removing the term entirely from Head 49A (**Recommendation 4**).

(10) While the threshold of harm caused by speech which is “intimidating” is higher, it is still low. For example, a social media post directed toward a TD after a particular vote in the Dáil Éireann which states “I’ll remember what you did” could be considered as intimidating, whereas it might simply mean that the individual is saying that they will remember the TD’s voting record when choosing who to vote for at the next election. To avoid this risk, while still capturing material which might not reach the threshold of “threatening”, we would suggest replacing “intimidating” with wording which more clearly requires putting a person in fear for their physical safety (either imminently or generally) (**Recommendation 4**).

Recommendation 4: Head 49A(b) should be re-worded as follows (changes underlined):
“Material which is likely to have the effect of (i) threatening or persecuting a person to which it pertains, or (ii) putting that person in fear for their physical safety, and which a reasonable person would conclude was the intention of its dissemination.”

Provisions for further categories of harmful online content (Head 49B)

(11) Head 49B provides that the Media Commission may propose to include or exclude further categories of material from the definition of harmful online content, and sets out the procedure for the Minister to follow when considering such proposals received by the Commission. While the procedure to expand the definition of harmful online content would be consultative in nature, requiring the Minister to consult with the Joint Oireachtas Committee when considering proposals, and then require approval by the Government, the Government is only able to approve or disapprove the proposal in its entirety. Furthermore, any proposals adopted by the Government would be taken forward through secondary legislation, in the form of regulations laid before the Houses of the Oireachtas. This type of secondary legislation tends to receive far less scrutiny than primary legislation. Given that the expansion of categories of harmful online content will likely result in impacts on what people are able to say online, any such expansion should, ideally, be made via primary legislation to ensure sufficient democratic oversight (see **Recommendation 5A**).

(12) If, however, the existing process of expanding what is considered harmful content via secondary legislation is retained, it would be beneficial to note explicitly that the right to freedom of expression should be considered by the Minister when considering any amendments, rather than a general reference to “the fundamental rights of users and operators of relevant online services” as currently provided in Head 49B(8)(g) (**Recommendation 5B**).

Recommendation 5A: Head 49B should be amended to require that any potential expansion of harmful online content be made via primary legislation to ensure sufficient democratic oversight.

(OR)

Recommendation 5B: Head 49B(8)(g) should be re-worded as follows (changes underlined): “the fundamental rights of users and operators of relevant online services, including, in particular, the right to freedom of expression”.

Definition of age inappropriate online content (Head 49C)

(13) Head 49C defines “age inappropriate online content” and then sets out a non-exhaustive list of examples. While the examples of age inappropriate online content (gross or gratuitous violence; cruelty, including mutilation and torture, towards humans or animals; and pornography) are all clear, the overall definition (“material which may be unsuitable for exposure to minors and that they should not normally see or hear and which may impair their development, taking into account the best interests of minors, their evolving capacities and their full array of rights”) is unclear. By defining it in such a way that it encompasses material which “may” be unsuitable, that should not “normally” be seen or heard, and which “may” impair a child’s development, vast quantities of material which may in practice cause no harm at all will be caught up, potentially interference with children’s right to freedom of expression under Article 13 of the Convention on the Rights of the Child (CRC).

(14) While we recognise that it might not be possible to set out an exhaustive list of types of material that are age inappropriate, we believe that the overall definition could be tightened, so as still to include the examples provided. While we understand that this definition is substantially based on wording derived from the revised Directive, and contains elements from the Council of Europe Recommendation on the Rights of the Child in the Digital Environment (CM/Rec(2018)7), we suggest that this definition also contain a specific reference to a child’s right to freedom of expression. This would mitigate potential risks to a child’s right to freedom of expression as provided for in Article 13 of the CRC (**Recommendation 6**).

Recommendation 6: Head 49C should be reworded as follows (changes underlined): ““age inappropriate online content” means material which is unsuitable for exposure to minors on the basis that it is likely to impair their development, taking into account the best interests of minors, their evolving capacities and their full array of rights, including the right to freedom of expression, and includes:

- (a) material containing or comprising gross or gratuitous violence,
- (b) material containing or comprising cruelty, including mutilation and torture, towards humans or animals, and,
- (c) material containing or comprising pornography.”

Online safety codes (Head 50A)

(15) Head 50A provides that the Media Commission shall prepare online safety codes governing standards and practices that shall be observed by designated online services. While the details of the codes themselves would need to be assessed once published, we suggest that any codes relating to user complaints and/or issues handling mechanisms allow users to be able to notify the designated online services in a simple and straightforward way, of content which they consider to be prohibited. These codes should also require services to, at a minimum, (i) inform affected users of content that has been flagged for removal, restriction or moderation; (ii) create an opportunity for that user to be able to input into the moderation process; and (iii) provide an

appeal mechanism for affected users to challenge decisions. This would reflect best practice as these elements are found within other proposed online safety frameworks, including the UK's Online Harms White Paper and Australia's Online Safety Bill.

Recommendation 7: Head 50A should be amended to ensure that online safety codes relating to user complaint and/or issues handling mechanisms reflect international best practice on transparency and due process.

A new subclause should be added to Head 50A as follows:

(x) Online safety codes which require designated online services to establish user complaint and/or issues handling mechanisms should allow users to notify designated online services in a simple and straightforward way, of content which they consider to be prohibited. In addition, such codes should require designated online service providers to, at a minimum, (i) inform affected users of content that has been flagged for removal, restriction or moderation; (ii) create an opportunity for that user to be able to input into the moderation process; and (iii) provide an appeal mechanism for affected users to challenge decision.

(16) Head 50A would also require the Media Commission to consider various matters in preparing online safety codes. We have suggestions in relation to two of the matters listed in Head 50A, and particularly on Head 50A(3)(f) - "the impact of automated decision making in relation to [content delivery and content moderation] by designated online services". This is because the scale of content which is generated and shared online will likely require companies to turn to automated processes, including AI, to meet codes that require services to minimise the availability of unlawful and harmful content on their services. There is a real risk that these automated processes will detect and remove content that is not actually unlawful or harmful in a particular context. Automated processes have had some success in relation to content moderation with types of images, including the ability to identify copies of images that have already identified by humans as constituting child sexual abuse and exploitation. However, automated processing has been less effective when identifying speech or less specific forms of unlawful or harmful content.³ As noted by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression:

"AI-driven content moderation has several limitations, including the challenge of assessing context and taking into account widespread variation of language cues, meaning and linguistic and cultural particularities. Because AI applications are often grounded in datasets that incorporate discriminatory assumptions, and under circumstances in which the cost of over- moderation is low, there is a high risk that such systems will default to the removal of online content or suspension of accounts that are not problematic and that content will be removed in accordance with biased

³ See, Cambridge Consultants, "Use of AI in Online Content Moderation: 2019 Report Produced on Behalf of OfCom", available at: https://www.ofcom.org.uk/data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf

or discriminatory concepts. As a result, vulnerable groups are the most likely to be disadvantaged by AI content moderation systems.”⁴

(17) This limitation is particularly acute given the breadth of some of the forms of harmful content set out in Head 49A, and particularly material which is likely to have the effect of “intimidating” or “humiliating” a person. If a human would have difficulty making such an assessment, then an automated tool would certainly be unable to make a proper assessment. For example, a human might have the ability to discern the difference between a joke made to a friend and a truly “intimidating” threat between two individuals, but even the most advanced automated processing tool is unlikely to be able to differentiate between the two forms of expression.

(18) Due to this inability to recognise context, and the evidence of inaccurate decision making by automated tools when it comes to many forms of content moderation, the use of automated content moderation tools would risk the inadvertent removal of content which is lawful and harmless. The online safety codes should therefore not require the use of automated processes to proactively monitor and remove content (**Recommendation 8**). Where automated decision-making is undertaken on a voluntary basis, the online safety codes should promote the use of open source tools, transparency around standards, and appropriate appeals mechanisms (**Recommendation 9**).

Recommendation 8: Head 50A(3)(f) should be re-worded as follows (changes underlined): “the impact of automated decision making, and its limitations, in relation to content delivery and content moderation by designated online services.”

Recommendation 9: New subclauses should be added to Head 50A as follows:

(x) Online safety codes may not mandate the use of automated processes, including automated decision making, by designated online services or categories thereof.

(x) Online safety codes which make reference to automated processes, including automated decision making, must promote the use of open source tools, transparency over the use of automated processes, and providing appropriate appeals mechanisms to challenge decisions made by automated processes.

(19) The second, as noted above in relation to Head 49B, is the importance of noting explicitly that the right to freedom of expression should be considered by the Media Commission when developing online safety codes, rather than a general reference to “fundamental rights” as currently included in Head 50A(3)(m) (**Recommendation 10**).

Recommendation 10: Head 50A(3)(m) should be re-worded as follows (changes underlined): “the fundamental rights of users and operators of designated online services, including, in particular, the right to freedom of expression”.

⁴ Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc. A/73/348, 29 August 2018. Para. 15.

Systematic complaints scheme (Head 52B)

(20) Head 52B would require the Media Commission to establish a scheme to receive notices about systematic issues with online services from nominated bodies such as non-governmental organisations, and would be able to request information, investigate or audit an online service on the basis of information received. We have no particular comments about how to ensure the effectiveness of the systemic complaints scheme, but do note that the, due to the framework as a whole, the scheme is likely only to encourage complaints to be made that online service providers are failing to do enough to remove certain types of material, and not that they are doing too much (for example taking down lawful and harmless content in efforts to comply with the online service codes). This risks creating an imbalanced picture of what online service providers are doing when it comes to compliance, focusing solely on under-compliance rather than over-compliance. While, in many areas of life, over-compliance with regulatory requirements does create any risks to individuals, here, over-compliance would create risks that individuals' right to freedom of expression online was being unnecessarily restricted.

(21) Under Principle 3 of the United Nations Guiding Principles on Business and Human Rights, states should "(a) enforce laws that are aimed at, or have the effect of, requiring business enterprises to respect human rights" and "(b) ensure that other laws and policies governing the creation and ongoing operation of business enterprises, such as corporate law, do not constrain but enable business respect for human rights". By encouraging systemic complaints to be made only when it comes to under-compliance, rather than over-compliance, we are concerned that the scheme would in fact constrain online service providers' ability to respect freedom of expression. This could be remedied by clarifying in Head 52B that complaints can be made that online service providers are taking steps not required by the online service codes. This has been proposed by the UK government in its full response to the Online Harms White Paper, which specifies that "*Ofcom will accept super-complaints demonstrating substantial evidence of a systemic issue that is causing harm, or risks causing harm, including about limits on freedom of expression*".⁵ **(Recommendation 11)**

Recommendation 11: A new subclause should be added to Head 52B as follows:

(x) For the purposes of section 52B(1), "systemic issues" include: (i) a failure by an online service provider to comply with any online safety code, and (ii) an online service provider taking action beyond that required by any online safety code and which interferes with the fundamental rights of users including, in particular, the right to freedom of expression.

Sanctions for non-compliance and senior management liability (Head 54A and Head 54B)

(22) We welcome that all sanctions provided for under Head 54A must be approved by a court and that a number of relevant factors must be taken into account when determining such sanctions. For example, Head 16A(1)(a) requires that any administrative financial sanction "(i) is appropriate and proportionate to the breach or the failure to cooperate with an investigation". Despite the high maximum administrative financial sanction (up to €20,000,000

⁵ UK Online Harms White Paper: Full Government Response to the Consultation (Dec 2020):

<https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response>

or, in the case of an undertaking, up to 10% of relevant turnover of the preceding financial year), we believe that the safeguards in place will help ensure that any financial penalties will, in practice, be proportionate.

(23) We are also pleased that senior management liability, as provided for in Head 54B, is limited to a secondary form of liability, restricts the scope of persons that could be held liable, and requires a particular burden of proof to be met - the consent, connivance, or neglect model. It is important that the available sanctions remain proportionate to the actual harm caused, since disproportionate sanctions may skew incentives to remain in compliance, leading to an abundance of caution and the over-removal of content, creating risks to freedom of expression. We consider the procedures and sanctions outlined in Heads 16, 54A and 54B to be sufficient in this regard.

Recommendation 12: We suggest that the procedures and sanctions outlined in Head 16, Head 54A and Head 54B are retained in their present form in the final version of the Bill.

Designation of relevant online services (Head 56)

(24) Head 56 would require the Media Commission to consider various matters when designating online service providers and we have suggestions in relation to these two matters. The first is “the impact of automated decision making in relation to [content delivery and content moderation] by relevant online services”, and we refer to our earlier reasoning set out above in relation to Head 50A(3)(f) (**Recommendation 13**).

Recommendation 13: Head 56(2)(g) should be re-worded as follows: “the impact of automated decision making, and its limitations, in relation to content delivery and content moderation by relevant online services.

(25) The second, as noted above in relation to Head 49B and Head 50A, is the importance of noting explicitly that the right to freedom of expression should be considered by the Media Commission when designating relevant online services, rather than a general reference to “fundamental rights” (**Recommendation 14**).

Recommendation 14: Head 56(2)(m) should be re-worded as follows: “the fundamental rights of users and operators of designated relevant services, including, in particular, the right to freedom of expression”.

(26) Head 56(13) provides that the Media Commission would not be able to require online services to comply with an online safety code that relates to material which is not a criminal offence to disseminate if the service is either an “interpersonal communications service” or a “private online storage service”. While we welcome these limitations, we are still concerned about the potential content of the online safety codes as they relate to encrypted services. While the exact details of the codes have yet to materialise, the inclusion of obligations for encrypted services to filter or monitor material would almost certainly amount to an unjustifiable restriction on individuals’ right to communicate privately. This is because such services use, almost universally, end-to-end encryption, which means those who develop and provide such

services are unable to filter or monitor, or otherwise access or moderate content which is generated or shared using them. Any filtering, monitoring or similar provisions would simply be unfeasible unless these services ceased to use end-to-end encryption, or weakened it in some way.

(27) The current wording in Head 56 still leaves open the possibility of codes containing obligations on “interpersonal communications service” or a “private online storage service” in relation to material which is a criminal offence to disseminate, even if those services use end-to-end encryption. There is therefore the potential for the codes to impose obligations to filter or monitor content on services which use end-to-end encryption. Compliance with such an obligation by service providers who currently use end-to-end encryption would force them either to cease using end-to-end encryption, or introduce some form of “backdoors” so as to be able to filter or monitor content. The ability to communicate using end-to-end encryption is critical for the security and safety of many who rely on the privacy that it provides, such as human rights defenders, journalists and minorities vulnerable to persecution. As the UN Special Rapporteur has noted:

“A state’s obligations to respect and ensure the rights to freedom of opinion and expression and to privacy include the responsibility to protect encryption (...) Because of the roles played by encryption, restrictions on their use must satisfy the requirements of legality, necessity and proportionality, and legitimacy. Blanket prohibitions of encryption plainly fail these conditions. Measures that systematically weaken encryption and digital security more generally, such as backdoors, key escrows, and data localization requirements, also interfere with rights to opinion, expression and privacy”.⁶

(28) We do recognise that it would be possible to impose certain specific obligations on services which use end-to-end encryption in relation to material which is a criminal offence to disseminate which would not undermine their use of encryption. For example, an obligation to enable users to be able to report content shared via such services which they believe is illegal, would not raise any concerns. However, Head 56, as currently worded, does not sufficiently limit obligations which would ensure that services would not have to cease, restrict or in any way weaken their use of encryption, and should be amended. We believe that this can be addressed through amendments either to Head 56 or Head 53A (**Recommendation 15**).

Recommendation 15: A new subclause should be added to Head 56 or Head 53A as follows:

(x) Online safety codes may not contain any obligations on online services or categories thereof which would have the purpose or effect of requiring or encouraging them to cease, restrict or in any way weaken their use of encryption.

⁶ Research Paper 1/2018 of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, “Encryption and Anonymity follow-up report”, June 2018. Paras 5-8.