SUBMISSION REGARDING THE REVIEW OF THE MODEL DEFAMATION PROVISIONS, AS PER PUBLIC NOTICE ISSUED BY THE NEW SOUTH WALES DEPARTMENT OF JUSTICE

Daphne Keller, PhD
Dr. Joan Barata Mir
Center for Internet and Society
Stanford Law School (USA)

The present submission has the object of providing comments and recommendations regarding the issues raised in Question 15 of the Discussion Paper which is the basis of this consultation:

> "(a) Does the innocent dissemination defence require amendment to better reflect the operation of Internet Service Providers, Internet Content Hosts, social media, search engines, and other digital content aggregators as publishers?
> (b) Are existing protections for digital publishers sufficient?
> (c) Would a specific 'safe harbour' provision be beneficial and consistent with the overall objectives of the Model Defamation Provisions?
> (d) Are clear 'takedown' procedures for digital publishers necessary, and, if so, how should any such provisions be expressed?"

The innocent dissemination defence is currently included in clause 32 of the Model Defamation Provisions and states the following:

> "(1) It is a defence to the publication of defamatory matter if the defendant proves that:
>
> (a) the defendant published the matter merely in the capacity, or as an employee or agent, of a subordinate distributor, and
>
> (b) the defendant neither knew, nor ought reasonably to have known, that the matter was defamatory, and
>
> (c) the defendant's lack of knowledge was not due to any negligence on the part of the defendant.
>
> (2) For the purposes of subsection (1), a person is a subordinate distributor of defamatory matter if the person:
>
> (a) was not the first or primary distributor of the matter, and
>
> (b) was not the author or originator of the matter, and
>
> (c) did not have any capacity to exercise editorial control over the content of the matter (or over the publication of the matter) before it was first published.
>
> (3) Without limiting subsection (2) (a), a person is not the first or primary distributor of matter merely because the person was involved in the publication of the matter in the capacity of:
>
> (a) a bookseller, newsagent or news-vendor, or
>
> (b) a librarian, or
>
> (c) a wholesaler or retailer of the matter, or
>
> (d) a provider of postal or similar services by means of which the matter is published, or
>
> (e) a broadcaster of a live programme (whether on television, radio or otherwise) containing the matter in circumstances in which the broadcaster has

no effective control over the person who makes the statements that comprise the matter, or
(f) a provider of services consisting of:
(i) the processing, copying, distributing or selling of any electronic medium in or on which the matter is recorded, or
(ii) the operation of, or the provision of any equipment, system or service, by means of which the matter is retrieved, copied, distributed or made available in electronic form, or
(g) an operator of, or a provider of access to, a communications system by means of which the matter is transmitted, or made available, by another person over whom the operator or provider has no effective control, or
(h) a person who, on the instructions or at the direction of another person, prints or produces, reprints or reproduces or distributes the matter for or on behalf of that other person."

This submission will be based on the most relevant international standards currently in place with regards to the role of private intermediaries vis-à-vis content moderation and intermediary liability for content distribution in general.

General Comment No. 34 concerning Article 19 of the International Covenant on Civil and Political Rights adopted on 29 June 2011, by the UN Human Rights Committee[1], states the following (para 39):

> "States parties should ensure that legislative and administrative frameworks for the regulation of the mass media are consistent with the provisions of paragraph 3. Regulatory systems should take into account the differences between the print and broadcast sectors and the internet, while also noting the manner in which various media converge".

The international mandate-holders on freedom of expression, including the UN Rapporteur on Freedom of Opinion and Freedom of Expression, the OSCE Representative on Freedom of the Media, the OAS Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights Special Rapporteur on Freedom of Expression, in their Joint Declaration of 1 June 2011 on freedom of expression and the Internet[2], stated the following:

> "(…) When assessing the proportionality of a restriction on freedom of expression on the Internet, the impact of that restriction on the ability of the Internet to deliver positive freedom of expression outcomes must be weighed against its benefits in terms of protecting other interests.
>
> (…) Approaches to regulation developed for other means of communication – such as telephony or broadcasting – cannot simply be transferred to the Internet but, rather, need to be specifically designed for it.
>
> (…) Greater attention should be given to developing alternative, tailored approaches, which are adapted to the unique characteristics of the Internet, for responding to illegal content, while recognising that no special content restrictions should be

[1] Available online at: http://www2.ohchr.org/english/bodies/hrc/docs/GC34.pdf.
[2] Available at: http://www.osce.org/fom/78309

established for material disseminated over the Internet".

Further guidance regarding regulation of speech on Internet platforms has been provided by reports of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, as well as in the widely endorsed Manila Principles for content takedown.[3] According to the Manila Principles, human rights-compliant judicial orders "must be limited to specific content violating the law authorizing the order, employ the least restrictive technical means, and be limited in duration and geographic scope."[4]

In addition to this, the same mandate-holders, in their Joint Declaration of 3 May 2017 on freedom of expression, "fake news", disinformation and propaganda[5], stated the following:

> "Intermediaries should never be liable for any third party content relating to those services unless they specifically intervene in that content or refuse to obey an order adopted in accordance with due process guarantees by an independent, impartial, authoritative oversight body (such as a court) to remove it and they have the technical capacity to do that."

From a national point of view, it is also important to keep in mind the already existing case law regarding intermediary liability for defamatory content posted by third parties, in particular, the Google Inc. v Duffy case[6]. In a nutshell, Google was found to be liable as a secondary publisher of defamatory material for reasons including that Google had intentionally designed its search engine to produce results in the way it did and had facilitated the reading of the defamatory material in an indispensable, substantial and proximate way. In addition to this, the Court also observed that the claimant had notified Google of the defamatory materials and had given Google a reasonable timeframe to remedy the situation.

Last but not least it also needs to be added, in terms of context, that hosting platforms like Google, Facebook, Instagram, Twitter and others distribute an important amount of third-party content on a daily basis. Platforms, particularly those operating as social media, typically have in place terms of service (TOS) to deal with inappropriate content once it has been uploaded. TOS usually deal with issues related to hate speech, protection of minors, and terrorist online content among others, and are enforced on the basis of a combined use of automated moderation mechanisms, human moderators, and reports from users. However, TOS do not usually include rules banning defamatory content as such. As the Discussion Paper states, the purpose of defamation laws is "to protect people's reputation and provide a dispute resolution framework to vindicate a defamed person's reputation". In addition, resolution of defamation claims requires "balancing freedom of expression and freedom to

---

[3] David Kaye (Special Rapporteur), Human Rights Council, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression at 6, U.N. Doc. A/HRC/32/38 (May 11, 2016), http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/32/38 [https://perma.cc/4RMF-U8EJ] (explaining that the Manila Principles "establish baseline protection for intermediaries in accordance with freedom of expression standards"); Manila Principles, https://www.manilaprinciples.org/.

[4] Principle 5, Manila Principles. Human rights guidance reflects a concern to avoid incentivizing platforms to avoid risk by removing lawful but controversial speech. Empirical research suggests that such"over-removal" is commonplace in notice-and-takedown systems. Daphne Keller, "Empirical Evidence of 'Over-Removal' by Internet Companies under Intermediary Liability Laws," Stanford Center for Internet and Society, October 12, 2015, http://cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws.

[5] https://www.osce.org/fom/302796

[6] Google Inc. v Duffy [2017] SASCFC 130.

publish information in the public interest on the one hand with the right of individuals to have their reputations protected from defamatory publications and the right to remedies for such publications on the other hand". This balance needs to be struck by a judiciary authority according to the specific criteria established by national legislation. Therefore, the labeling of a certain piece of content as defamatory requires a complex process of assessment, compared to many examples in the areas mentioned above. This assessment is linked, first of all, to an individual action against an expression that is considered to be harmful, and the possible concurrence of a wide array of defences that may deem such harm justified in certain cases.

On the basis of the abovementioned parameters we comment and recommend the following, as the approach most consistent with international human rights principles:

a) Online platforms that aggregate and distribute third-party content which is unmoderated or only moderated ex-post on the basis of their TOS cannot be tasked with the responsibility of evaluating the defamatory nature of a certain piece of content, even in cases where there is a complaint or takedown request conveyed by the affected person.

b) Online platforms that aggregate and distribute third-party content which is unmoderated or only moderated ex-post on the basis of their TOS must only have the legal obligation and responsibility to take down a specific piece of content of defamatory nature when requested by the competent Court after evaluating the circumstances of the case.

c) It is therefore recommended that, for the purposes of the innocent dissemination defence, intermediaries mentioned above are explicitly identified in the Model Defamation Provisions, and automatically considered as subordinate distributors in any case.

d) In addition, competent Courts should limit the geographic scope of injunctive relief in defamation cases involving platforms.[7]

[7] Professor Dan Svantesson of Bond University, a world-renowned expert on Internet Jurisdiction, lists doctrinal tools and legal considerations regarding cross-border enforcement requests in his 2017 book *Solving the Internet Jurisdiction Puzzle*. These include in particular "dis-targeting" approaches (p. 101), geolocation technologies (p. 200-214), and relevance of both rights and duties for courts assessing conflicts of national law (p. 218).