9 September 2022

Review of Model Defamation Provisions
Policy, Reform & Legislation
NSW Department of Communities and Justice
Locked Bag 5000, Parramatta NSW 2124

By email: defamationreview@justice.nsw.gov.au

Dear Chair,

Thank you for the opportunity to provide a written submission for the Stage 2, Part A exposure draft Model Defamation Amendment Provisions.

Twitter supports smart regulation, and our focus is on working with governments to ensure that regulation of the digital industry is practical, effective, and feasible to implement while remaining inclusive and keeping core democratic values intact and promoting tech innovation, including Twitter's core commitment to an Open Internet worldwide.

In this vein, we support strong mechanisms to protect against defamation and assist in the swift removal of unlawful content, while balancing the need to protect principles of free expression to prevent a chilling effect on robust and open public discourse and avoid any unintended harmful consequences.

We trust this written submission will be a useful input to the Department's consultation process. Our submission stands together with the submission from the Digital Industry Group Inc., of which Twitter is a member. For clarity, and to complement and reinforce these statements, we've structured this submission to address the key issues within the MDAPs as they pertain to Twitter in Australia.

Working with the broader community we will continue to collaborate to create a safe and secure digital ecosystem. Twitter is committed to working with the government, our industry partners, and other stakeholders as we continue to build our shared understanding of the issues and find optimal ways to approach these together.

Thank you again for the opportunity to provide input as part of this important legislative reform process.

Kind regards,

[redacted]
[redacted]
Director of Public Policy
APAC

[redacted]
[redacted]
Senior Director of Public Policy
JAPAC

Introduction

In response to the Model Defamation Amendment Provisions 2022 ("**MDAPs**"), Twitter's submission will focus on key areas that relate to our business and operations within Australia.[1]

Recommendations 1 and 2

The definitions contained in Section 4 could benefit from further clarification with regards to specifying if it's the intention for a platform's internal search functionality to also be captured within the proposed definition of 'search engine.'[2]

On Twitter, there are many ways to use the Search function within the app or on desktop. People can find Tweets, keywords, or hashtags. We give users control over what they see in Search results through Safe Search mode, which filters potentially sensitive content, along with accounts that have been muted or blocked by a user, from their search results.[3] Additionally, Twitter Advanced Search is available when users logged in to the desktop at twitter.com, which allows people to tailor search results to specific date ranges, accounts, and more.[4]

In the MDAPs, Section 4 defines 'search engine' to mean a software application designed to enable its users to search for content on the internet.[5] As currently framed, this definition is unclear if it's intended to encompass internal functionality for users to search for information within that specific platform. The Background Paper explains that search engines "have no interest in the content," and "simply use an automated process to provide access to third-party content," but does not clarify if this is limited to search engines within a bounded platform or search engines that find content or websites across the World Wide Web.

While Twitter's internal search functionality is surfacing potentially relevant content that is available on the Twitter platform, we believe that the inclusion of internal search capabilities with the definitions contained in Section 4 is warranted, especially when considering that a platform's internal search functionality could be liable for defamation by indexing content, ranking results based on relevance or keywords, and featuring hyperlink snippets while a stand-alone search engine would be statutorily exempt.

Twitter submits that there should be greater clarity as to the circumstances in which a digital intermediary that provides internal search functionality would be covered either by the exemption for search engine providers.

Recommendations 3A and 3B

Twitter supports the introduction of a defence to be available to digital intermediaries where certain steps have been taken in response to a complaint within a specified period. With respect to the two models proposed by Recommendations 3A and 3B, Twitter supports the 'safe harbour' model proposed in Recommendation 3A.

Both models would require that digital intermediaries have in place a complaints mechanism that is reasonably accessible to the public, and to take steps with respect to complaints regarding digital content within 14 days, which is consistent with the purposes of the proposed reforms. The key point

---

[1]Model Defamation Amendment Provisions (2022) NSW. [https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions/draft-part-a-model-defamation-amendment-provisions-2022.pdf].
[2] *Ibid.*
[3] https://help.twitter.com/en/using-twitter/twitter-search
[4] https://help.twitter.com/en/using-twitter/twitter-advanced-search
[5]chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions/draft-part-a-model-defamation-amendment-provisions-2022.pdf

of difference between the models is that the Recommendation 3A 'safe harbour' model would automatically apply to digital intermediaries where a complainant has, or is provided with, sufficient identifying information about the originator to be able to pursue proceedings against them. This furthers the purpose of these reforms by focusing disputes between complainants and originators of defamatory matters.

In practice, we believe that the model proposed by Recommendation 3B would mean that digital intermediaries with limited or no knowledge of a dispute would be deprived of a defence in circumstances where it is plain that a complainant already has sufficient information to pursue an originator directly.

To deny digital intermediaries a defence in these circumstances could generate frivolous or vexatious complaints where complaints notices are submitted to digital intermediaries in circumstances where a complainant has no intention of pursuing or contacting an originator.

Further, it may not be apparent to a digital intermediary within 14 days that a digital matter complained of is defamatory. The model proposed by Recommendation 3B would place pressure on digital intermediaries to remove content that a court may not consider defamatory simply to avoid liability or to retain the availability of a defence should a complainant pursue proceedings against the intermediary directly. Such a result is inconsistent with the objective of striking a balance between "protecting reputations and not unreasonably limiting freedom of expression."

Interaction between complaints notices and concerns notices

An intended purpose of the Recommendation 3A safe harbour model and the underlying complaints mechanism is to focus disputes between complainants and originators of defamatory content as much as possible.

At present, the distinction between a complaints notice in the form contemplated by section 31A and a concerns notice in the form described in section 12A is unclear. It appears that a complaints notice sent to a digital intermediary may satisfy the requirements for a concerns notice under section 12A, allowing proceedings to be commenced against the digital intermediary after 28 days, regardless of whether the complainant intended the complaints notice to serve this dual purpose.

A complaints notice should be sent for the purpose of pursuing the originator, or otherwise having defamatory material removed. It should not serve dual purposes and allow proceedings to be commenced against a digital intermediary in 28 days when the digital intermediary has spent time seeking consent from an originator to provide its contact information. Such an approach would not allow the digital intermediary to prepare and potentially defend itself against proceedings because its focus at this point should be to resolve the complaints notice for the complainant within 14 days. To allow a complaints notice to serve this secondary purpose would also undermine the purpose for which concerns notices were recently made a mandatory precursor to commencing proceedings: to facilitate efficient and cost effective resolutions of disputes prior to litigation.

Further to a complaints notice for the purpose of pursuing the originator, Twitter may or may not have email contact details for an account. We cannot force persons to provide an email and where they do, we can attempt to notify, but we do not control how quickly they decide to provide a relevant email address or not.

Twitter proposes that the MDAPs make clear that a complaints notice sent for the purpose of section 31A cannot serve as a concerns notice for the purpose of section 12A. This clarification would be similar to that in section 12A(2), which makes clear that a document that is to be filed to commence defamation proceedings cannot serve a dual purpose as a concerns notice.

The drafting should also be amended to make clear that where a complainant does not send a concerns notice to a digital intermediary (for example, where they are unable to do so because they have sufficient identifying information about the originator), the digital intermediary will have the benefit of the safe harbour defence. Twitter considers that the current drafting contains ambiguity and allows for a complainant to circumvent the complaints notice process to go straight to sending a concerns notice, thereby depriving digital intermediaries the ability to rely on the safe harbour. The drafting should be clarified in order to achieve the stated aim of the proposal set out in the Background Paper.

"Sufficient identifying information"

As noted above, a key objective of these proposed reforms is to focus disputes regarding defamatory material between complainants and the originators of the defamatory material. The intention of the proposed complaints mechanism is that digital intermediaries would be tasked with either facilitating contact between complainants and originators, or removing the matter complained of, and would in those circumstances have a complete defence. Where the complainant already has "sufficient identifying information" about the originator, a complaints notice cannot be sent, and the intention is that the digital would automatically receive the benefit of the safe harbour defence.

However, the proposed definition of "sufficient identifying information" requires information sufficient to enable proceedings to be commenced against an originator. In section 31A(4)(a), a complainant is not required to have applied for an order for substituted service or preliminary discovery to obtain such information before they are able to send a complaints notice. As currently drafted, it follows that:

- a complainant could be in possession of an originator's mobile number and email address (such that they are capable of sending a concerns notice, and could commence proceedings if an order for substituted service were obtained), but still be able to request that a digital intermediary provide further information required for traditional methods of service (namely, personal service); and
- a digital intermediary will not be taken to have met the requirements in section 31A(1)(c)(i) unless they can provide that specificity of information, even though it is not a kind of information ordinarily held by digital intermediaries, and would need to be provided by the originator. Further, it can be anticipated that originators who may be receptive to being put in contact with a complainant electronically may be hesitant to provide their physical address online, and would be likely to withhold consent altogether.

By compelling digital intermediaries to obtain from originators information they do not ordinarily hold, this part of the proposed section may undermine the objectives of focusing disputes between complainants and originators while also protecting reputations without unreasonably limiting freedom of expression.

Twitter submits that it would support the objectives of this defence for the meaning of "sufficient identifying information" to more closely align with the types of information digital intermediaries ordinarily hold (e.g. email addresses and/or mobile phone numbers), so as not to require them to request new types of information from their users. Such regulatory proposals to collect additional information also run counter to the universally accepted privacy practice of data minimisation. Data minimisation requires goods or service providers to not seek to collect data beyond what is reasonably needed to provide the good or service. Data minimisation forms part of the existing APPs under the *Privacy Act 1988* (Cth), and is also a key principle of the Consumer Data Right.[6] This change would also allow complainants to send concerns notices electronically, and if necessary, seek other information from other intermediaries or orders for substituted service once proceedings

[6] OAIC, "Chapter 3: Privacy Safeguard 3 — Seeking to collect CDR data from CDR participants", accessed at https://www.oaic.gov.au/consumer-data-right/cdr-privacy-safeguard-guidelines/chapter-3-privacy-safeguard-3-seeking-to-collect-cdr-data-from-cdr-participants/.

have been commenced. This would also limit the extent to which a complaints mechanism may be abused for purposes other than genuine efforts to resolve disputes about defamatory material posted online.

Orders to have online content removed prior to trial

In line with the fundamental principles of privacy and free expression, Twitter submits that any model under the draft MDAPs allow an intermediary safe harbour to be available until platforms receive actual notice via a court order directing a platform to either release private information or remove specific content.

If the MDAPs were amended to provide for a statutory test for when a court may order that online content is removed prior to trial, or a regime for when a court may order that content that has been found to be defamatory is to be removed, a number of safeguards must be enshrined in the MDAPs to protect the interests of originators and digital intermediaries.

The current test for whether an interim injunction should be granted in defamation proceedings was established in the High Court decision of *Australian Broadcasting Corporation v O'Neill* (2006) 227 CLR 57 ("**O'Neill**").

In *O'Neill,* the court held that the power to order an interlocutory injunction in a defamation proceeding should be approached with 'exceptional caution.' The court must balance the value of free speech in considering whether to grant an interim injunction, and this is a significant consideration. Twitter submits that the common law test in *O'Neill* for whether content should be removed prior to trial by way of interim injunction remains appropriate. The common law test fairly balances important considerations of free speech with an individual's right to protect their reputation.

Therefore, it is unnecessary for amendments to be enacted to the MDAPs legislating a statutory test providing a court the power to remove online content prior to trial.

Situations may develop where a court determines in a preliminary hearing that a publication meets the serious harm threshold, and a court orders that particular allegedly defamatory material must be removed. That proceeding could then progress to trial and the defendant/originator could be successful by way of a defence. Then the situation would be that the publication has already been removed (due to the plaintiff's success in the serious harm proceeding), but with a successful defence subsequently established to the defamatory material. This raises questions regarding if and how a defendant could be compensated for infringement of their rights and wrongful takedown of their publication, as well as the possible liabilities of relevant intermediaries. Therefore, takedown orders should only be permitted after a final judgement is made.

Removing content that has been found by a court to be defamatory

During the course of this reformation process, it has been contemplated whether courts should have a clear regime to order Internet intermediaries to take-down or disable access to content that has been found to be defamatory regardless of whether the Internet intermediary would be liable as a defendant in the proceeding, and whether or not it is a party to a proceeding.

In Australia, there is currently no case law where a court has ordered an Internet intermediary, such as Twitter, to remove material that has been determined to be defamatory when that Internet intermediary is not a party to the proceeding. There is currently uncertainty as to whether courts have the power to order non-parties to remove defamatory content.

Twitter submits that it is unnecessary to make amendments to the MDAPs to provide power for a court to order a third-party remove material that has been determined to be defamatory. Twitter submits that drafting changes are needed to ensure that the material the subject of the takedown

order is precisely defined (e.g. by identification of specific website URLs) rather than being general in nature, which could create issues around whether the orders require active monitoring and searches for material which may potentially be captured, and may also restrain future publications. These issues are addressed further down in relation to Recommendation 5.

Circumstances would likely arise in which Twitter is required to determine whether content on its platform conveys similar imputations as those which they have been ordered to remove. If Twitter was then required to remove a substantial amount of content absent precise identification of the content that is the subject of the takedown order (e.g. by reference to the URL or another specific online location designation), then there could be a significant chilling effect on public discussion. When looking at these considerations, we believe that:

First, the right to freedom of expression and free speech must be a primary consideration when determining whether to make an order.

Second, the threshold for overcoming free speech primary consideration should be only in 'exceptional circumstances.'

Third, for orders prior to judgement, the applicant must show an extremely strong *prima facie* case that defamation has occurred and no defences apply.

Fourth, the party to which the prospective order applies (e.g. a digital intermediary) and the originator of the content must be given prior written notice of the relevant application or prospective order, and a statutory right to be heard. For example, the originator of the content may have a strong defence to their allegedly defamatory statement.

Fifth, if an interlocutory order is made ordering removal of allegedly defamatory content, the parameters of any order must be clearly identified. A non-exhaustive list may include the specific URL(s) where the allegedly defamatory content is found, or the amount of time the Internet intermediary is required to 'block' content.

Sixth, in line with Twitter's global user notice policy, the MDAPs must clearly identify whether an order could be made if the originator of the content objects to the order upon notice being provided.[7]

Seventh, for orders prior to judgement, applicants who are unsuccessful in applications to have content removed prior to trial must be subject to costs consequences that are payable forthwith, to protect freedom of speech.

Eighth, for orders subsequent to content being found to be defamatory, the order must clearly specify the relevant defamatory material that requires removal through a specified URL. An Internet intermediary should not be required to remove content that is not specifically identified by the order, including cases of content with an equivalent meaning as this would put too much onus on the Internet intermediary to consider whether other content conveys certain imputations.

Twitter is also concerned that the complaints mechanism may be abused by persons seeking to obtain information about an originator's location for improper purposes.

For example, perpetrators of domestic violence seeking to obtain information about an originator's location, or suppress allegations posted online or to otherwise intimidate or harass the originator. The originator in this scenario would likely refuse to have information about their location disclosed, and the matters complained of may then be removed to preserve the defence. This outcome would not strike the appropriate balance between allowing genuine complainants to protect their reputation, and supporting freedom of expression.

---

[7] https://help.twitter.com/en/rules-and-policies/twitter-legal-faqs

In circumstances where the originator may not wish to provide their contact details to the complainant, the complaints process under Section 5 of the UK *Defamation Act 2013* enables the originator to provide their contact details to the intermediary, and the intermediary is not authorised to provide the contact details to the complainant unless served with a court order. Further consideration should be given to the options for the originator in such circumstances; we are concerned that the only option under Recommendation 3A in this circumstance is the removal of content and this outcome would not strike the appropriate balance between allowing genuine complainants to protect their reputation and supporting freedom of expression. Similar to the processes established in the UK through Norwich Pharmacal order, Twitter would recommend that a court order be issued prior to disclosure of user information for defamation proceedings.

Twitter welcomes the fact that, unlike the UK Section 5 regime, the Model 3A complaints notice process would provide internet intermediaries with flexibility as to the method they choose to seek the consent of the originator to pass on their details to the complaint. However, there are elements of the UK "Notice of Complaint" that may go some way to address this challenge. Unlike draft Recommendation 3A, the UK model also requires the complainant to provide their name, email address, and "confirmation that the complainant does not have sufficient information about the person who posted the statement to bring proceedings against that person, which would assist in the intermediary's efforts to connect them with the originator." We query why these elements have not been used in the current draft of Recommendation 3A and suggest this be considered.

"Access prevention steps"

This term is currently defined as "a step to remove, or to block, disable or otherwise prevent access by some or all persons to, the matter." The definition should clarify that it is sufficient to prevent access to a matter by persons in Australia. The absence of such a clarifying statement raises issues of purported extraterritorial application, which cannot be achieved by amendments to state defamation legislation.

Further, the drafting of section 31A(1)(c)(ii) currently requires that the digital intermediary itself took access prevention steps in order to receive the benefit of the defence. The drafting should make clear that if access prevention steps are taken by the originator or some other third party prior to the digital intermediary having a chance to do so, the defence should still apply.

Content of complaints notice

A digital intermediary that has received a complaints notice should be able to determine, based on the steps described and the information provided, whether a complainant has taken reasonable steps to obtain and/or already has sufficient identifying information to be able to pursue the originator directly. Digital intermediaries should not be burdened with the task of assessing information provided and trying to decide whether a complaints notice has been "duly" made, requiring their prompt action, or if no action need be taken because the complaints notice is bad in form.. Twitter suggests that the complaints notice be required to include, in addition to the requirements already proposed:

- The identifying information (if any) that the complainant has been able to find about a poster by taking reasonable steps (in addition to a description of the steps taken); and
- As noted above, a statement that the complainant has taken reasonable steps to obtain sufficient identifying information, but has been unable to obtain sufficient identifying information.

Time period

We understand that where a complaints notice is duly made, a digital intermediary would need to take action within 14 days to preserve the availability of the safe harbour defence. In circumstances

where, until a complaints notice is received, a digital intermediary generally has no background or context to the dispute and will likely need to make extensive enquiries to form a view as to the appropriate action to be taken in response to the complaint. Thus, we believe there needs to be a flexible standard in place to allow for the timeframe to be modified in circumstances where a platform is required to undertake further investigations or fact-finding with respect to the account or content in question.

Additionally, we believe that clarification is needed on the bounds of any designated time period. This should address uncertainty in relation to weekends and public holidays, which also differ for the complainant and the digital intermediary if they are not situated in the same time zone. For example, the UK legislation clarifies that the designated "time period does not include any time falling on a non-business day in England and Wales (i.e. Saturday, Sunday, Good Friday, Christmas Day or a Bank Holiday)."[8]

Additionally, we would recommend that the period could be extended with the complainant's agreement, depending on the circumstances needed to contact the poster.

Complaints mechanism to be "easily accessible"

Twitter submits that the requirement for a complaints mechanism to be "easily accessible by members of the public" should be rephrased to more clearly set an objective test that would employ a reasonable person standard.

Recommendation 4: Interaction with *Online Safety Act* immunity

Interaction with the Online Safety Act generally

Currently under the *Online Safety Act 2021* (Cth) ("**OSA**"), section 235, certain 'Internet service providers' and 'Australian hosting service providers' are afforded protection from liability. Specifically, Internet service providers and Australian hosting service providers can rely on section 235 as a defence where they were not aware of the nature of online content and where a service provider or hosting service provider would be required to monitor, make inquiries about, or keep records of online content.

For the purposes of section 235 of the OSA as applied to State and Territory defamation laws, it is not clear whether a general complaint to a digital intermediary is sufficient to make it 'aware of the nature of the online content' or whether a complaint must specify the defamatory nature of the content. It is also not clear whether a court judgement finding the material in question defamatory is required before the digital intermediary loses the immunity in section 235.

There is also uncertainty as to the territorial reach of section 235. In *Fairfax Media Publications; Nationwide News Pty Ltd; Australian News Channel Pty Ltd v Voller* [2020] NSWCA 102 ("***Voller***"), Basten JA (in obiter) considered that the better view of clause 91(1) of Schedule 5 of the *Broadcasting Services Act 1992* (Cth) (the predecessor to section 235 of the BSA) is that it applies only to those internet content hosts which host content on servers located in Australia. The definition of "Australian hosting service provider" in the OSA, which came into effect on 23 January 2022, is a "person who provides a hosting service that involves hosting material in Australia". This suggests that the immunity will not apply to hosting service providers which do not have servers physically present in Australia.

The MDAPs need to consider the application of section 235, including with respect to the innocent dissemination defence in section 32 of the Uniform Defamation Acts, to determine when and how

[8] Defamation Act 2013 [https://www.legislation.gov.uk/ukpga/2013/26/contents/enacted].

basic Internet services fall within the definitions of 'Internet service providers' and 'Australian hosting service provider'.

Any amendments to the innocent dissemination defence in section 32 of the Uniform Defamation Acts should be reconciled with section 235 of the OSA to ensure there is a consistent approach that media publishers who monitor reader comments are not liable for defamatory comments that they do not endorse or adopt.

Interaction with Online Safety Act immunity

Twitter believes that the exemption from section 235(1) of the *Online Safety Act* (OSA) should be applicable for intermediaries in the defamation context. Given the potential interaction and related purposes, and noting that it is vital for digital intermediaries to have sufficient clarity regarding their legal obligations, Twitter considers this a topic worthy of examination to ensure there is no unintended inconsistency or overlap. Twitter does not believe that there should be an exemption as digital intermediaries could then be required to adhere to a general monitoring obligation.

Furthermore, consideration should be given to the interaction with the OSA's adult cyber abuse scheme as there will be material that could be considered both defamatory and cyber abuse. Part 3 of the OSA provides a complaints system for cyber-abuse material targeted at an Australian adult, enabling that person to make a complaint to the e-Safety Commissioner. Part 7 of the OSA outlines how the provider of a social media service, relevant electronic service or designated internet service may be given a removal notice by the eSafety Commissioner. It is unclear the extent to which this process will interact with the proposed complaints notice process contemplated in relation to allegedly defamatory material.

The digital content regimes under the OSA and any new regimes created by the MDAPs will result in a duplication of obligations and create significant operational and administrative burdens for intermediaries. The approach adopted for a complaints notice procedure should consider the process under the OSA to avoid confusion regarding any conflicts of law.

Recommendation 5: New court powers for non-party orders to remove online content

Twitter acknowledges and respects local law and understands there are existing court powers by which a digital intermediary may be ordered to facilitate the removal of defamatory matter as is acknowledged in the Background Paper.

Against that background, Twitter opposes the introduction of new court powers that would expose non-parties to potentially broad, unlimited orders requiring them to expend significant time and resources to identify and monitor the republication of digital matters. Even if it were possible for intermediaries to comply with such orders (in many cases it will not be), it has the potential to impose on a court the significant burden of supervising non-parties to litigation. This would also be problematic in the context of interlocutory applications as there has not been an opportunity for the user to put forward a defence and as such, a court could make a finding that the content is not defamatory at a later date.

Twitter is concerned about the wording of a proposed new power for courts to order a non-party to "to prevent or limit the continued publication or republication of the [defamatory] matter..." While we note that companies would get a chance to make submissions before such an order is made, we are concerned that this might lead to a general monitoring obligation that is not feasible with current technology. The proposal differs significantly from the status quo which is the usual practice of digital intermediaries to limit access to digital matter on their platforms once a court has found the matter to be defamatory of a complainant (where the originator of the matter has not already removed it. The wording should be clarified to ensure that this doesn't equate to a general monitoring obligation.

Were a court to be empowered to make orders of this kind notwithstanding the imposition it places on the court to supervise non-parties, it is critical that the scope be very clearly and narrowly defined, so as to avoid unduly burdening the non-party, acknowledge the nature of their involvement in the publication, and take into account what steps they are capable of taking.

With regard to the form proposed in section 39A of the draft Part A MDAPs, Twitter has very real concerns. The section would empower a court to make orders, without express geographical or temporal limitations, requiring non-parties to take steps that may be beyond their technical capacity, exposing them to the risk of being found in contempt of orders made in proceedings in which no relief was originally sought against them. We make the following specific comments on section 39A:

> The section contains no geographical or temporal limitations. On the current wording of section 39A, a court may make an order that would require a digital intermediary to:
>
> - monitor, indefinitely, all subject matter relating to the complainant to ensure the matter(s) the subject of the court's findings are not republished;
> - remove matter(s) from being accessed anywhere, not just Australia, despite no finding having been made that the matter(s) being defamatory under local laws other than in Australia.

As noted in the Background Paper, it is not clear that the section, to the extent it seeks to empower courts to make orders concerning extraterritorial publication, is of a kind that can be made by amendments to the MDAPs.

Section 39A would empower a court to make non-party orders in circumstances where it has not first made equivalent orders against the originator of the defamatory matter. Where proceedings have been commenced against the originator only, it is appropriate that relief against ongoing publication or republication be sought in the first instance from the originator.

Twitter submits that only once an originator has failed to comply with those orders should a non-party be required to expend time and resources to consider the proposed order and become involved in the proceedings. To do otherwise may create a practice of complainants seeking relief against non-parties only, or against non-parties and originators concurrently, pulling non-parties into proceedings where the originator is capable of and intends to prevent or limit ongoing publication or republication.

Section 39A(2)(a) contemplates non-parties being ordered to prevent or limit the "continued publication or republication" of a matter. It is one thing to require a non-party to remove digital matter or posts that were expressly the subject of the court's decision and are clearly identified to the non-party (for example, by reference to a URL, date and time, and content of the post). However, orders regarding "republication" may pose an impossible burden on non-parties and on the court to supervise compliance with its orders.

While it is not clear from the wording of the section, Twitter understands from the Background Paper that the section is intended to be directed to instances where a matter has gone "viral" on a platform, republished many times by persons other than the originator(s) involved in the proceedings. An order to prevent or limit any republication may require monitoring by a non-party to ensure that neither the defendant, nor any other person, republishes the matter in question.

To comply with such an order would demand significant resources, is not reflective of the non-party's limited knowledge of and involvement with the matter, and may in some circumstances be in conflict with the OSA. Even if significant resources were expended, any instance of republication of which the non-party may not be aware could expose the non-party to being found in contempt of the order.

Twitter submits that it would be more appropriate that:

- a non-party be the subject of orders only with respect to the precise matters that were the subject of the court's findings, identified as precisely as possible, as published by the originator(s) who were named as defendants in the proceedings in question;
- the defendant(s) alone be the subject of orders prohibiting republication of a matter;
- to the extent that a defendant does not comply with such orders, the plaintiff is responsible for bringing the non-compliance to the court's attention, and obtaining additional non-party orders (specifically identifying any re-publications) as appropriate.

The Background Paper includes the following: "Where a complainant has obtained judgement against an originator, the court has awarded a remedy but in some circumstances, enforcement of the remedy can be elusive. Where an originator is unable to remove the content (for example because it has been copied and shared by others using new hyperlinks or on other platforms and therefore has 'gone viral') or refuses to do so, there may be a role for non-parties (often comprising internet intermediaries which host or otherwise provide access to the content) to play." It appears that section 39A may see digital intermediaries ordered to remove any "republication" (e.g. re-post or screenshot) of a digital matter, including matters that have gone 'viral'. In practicality, this would create considerable difficulties complying with such an order, and considerable resources required to comply.

Section 39A(3) refers to non-parties being ordered to take "one or more access prevention steps," or "a step to be taken in relation to all, or only some, of the users of an online service." There is ambiguity in the draft Part A MDAPs as to what is meant by an "access prevention step."

'Access prevention step' is defined as "a step to remove, or to block, disable or otherwise prevent access by some or all persons, to the matter." We believe there would be technical limitations on what access prevention steps can be taken, particularly where the matter in question is the subject of a temporary injunction. It is worth noting that intermediaries may not be able to restore digital content if the court concluded that the defendant was not liable at a later date. Thus, we suggest clarification that there is no expectation to restore content in circumstances where the court concluded that the defendant was not liable.

Recommendation 6: Considerations when making preliminary discovery orders about originators

Twitter is broadly supportive of the proposal advanced in Recommendation 6 to amend the MDAPs to provide that, in any preliminary discovery application, the court must take into account privacy, safety, or public interest considerations that may arise should the order be made. However, Twitter considers that the amendments should go further than this current threshold and should also require that a complainant prove that they have a prima facie case.

Recommendation 6 would require any court considering a preliminary discovery application to take into account, among the other relevant factors, the objects of the *Defamation Act*, and "privacy, safety or other public interest considerations that may arise if the order is made."

The most likely scenarios, following these reforms, where a complainant would seek preliminary discovery from a digital intermediary to identify a poster and/or obtain their contact information, are:

- the complainant engaged with the complaints mechanism, but the digital intermediary did not provide the poster's contact information (e.g. because the poster did not consent and/or the digital intermediary took access prevention steps.)
- the complainant did not send a complaints notice before seeking preliminary discovery.

We note that in the second scenario, the complainant would have skipped the complaints notice mechanism (and in theory, could have obtained the information they seek via that mechanism).

Twitter would advocate that the complainant be required to go through the complaints notice mechanism before they seek preliminary discovery. If complainants are prevented from seeking preliminary discovery where they have not first sent a complaints notice, this will encourage efficiency in the process and avoid circumvention of the process. It would also ensure that digital platforms of all sizes and scale, including new market entrants and startups, would be better equipped to adopt a compliance posture with this legislation as complainants would be encouraged to use the complaints notice mechanism as their first resort.

Twitter is broadly supportive of the recommendation to require a court to take into account certain considerations before making an order for preliminary discovery of a poster's identity or contact information. As noted in the Background Paper, the section proposed in the draft MDAPs would co-exist with requirements in Australian jurisdictions that courts consider all relevant circumstances. In the context of an application for preliminary discovery from a digital intermediary, seeking to identify a poster and/or obtain sufficient contact information for them, we would expect that a court would take into account the factors listed in section 23A(2).

Revealing the identities of originators in relation to a potential defamation action

Twitter's mission is to provide a platform where people have the opportunity to exchange ideas and information, and to express their opinions and beliefs. There is a concern that if preliminary discovery applications become more prevalent in Australia, then it may have a chilling effect on freedom of expression. It may be a disincentive for users to engage in public debate and express their opinions and beliefs if they believe their identity and contact details may be revealed to a complainant in the future without strong safeguards in place.

A serious concern is the potential for such orders to be abused. Specifically, Twitter is concerned that the low threshold for preliminary discovery orders does not currently ascertain when a prospective complainant's primary aim is to unmask an pseudonymous originator with the potential to harass or intimidate that person (e.g. especially if personal details, such as a mobile number, are ordered to be handed over as to a legitimate application).

The current threshold for granting orders for preliminary discovery to identify a prospective defendant is relatively low in Australia compared with the United Kingdom (UK) and Ontario, Canada, discussed in further detail below. For example, in Australia the complainant is not required to demonstrate a prima facie cause of action in defamation, nor are they required to demonstrate they are acting in good faith. Recent Australian Federal Court decisions in *Kukulka v Google LLC* [2020] FCA 1229 and *Kabbabe v Google LLC* [2020] FCA 126 highlight the ease with which complainants (otherwise known as prospective applicants) have successfully sought provision of information as to the identity of an unknown originator of allegedly defamatory material as part of preliminary discovery against the Internet intermediary.

As is evident from Federal Rule 7.22, there is currently no express requirement for the court to have regard to competing considerations such as privacy of users, freedom of expression, or the protection of whistleblowers when considering pre-action discovery applications. This is of great concern to Twitter as there are likely to be issues regarding the conflicts of laws if the originator is not a citizen or resides outside of Australia.

The UK approach

In the UK, orders to 'innocent' third parties to disclose the identity of alleged originators of defamatory content are known as 'Norwich Pharmacal' order (NPO) which derived from the UK case of *Norwich Pharmacal v Commissioners of Customs and Excise* [1974] AC 133. Twitter is supportive of the NPO process as it helps protect user privacy and provide a protected, legal avenue for the disclosure of private data.

Through this process, a court exercises its equitable jurisdiction and therefore under the common law a complainant is required to prove:

a. *a wrong must have been carried out, or arguably carried out, by an prospective defendant;*
b. *there must be the need for an order to enable action to be brought against the prospective defendant; and*
c. *the person/company against whom the order is sought must:*
    i. *be mixed up in so as to have facilitated the wrongdoing; and*
    ii. *be able or likely to be able to provide the information necessary to enable the prospective defendant to be sued.*

Unlike in Australia, UK courts are expressly required to take into account countervailing human rights considerations, such as data rights, rights of privacy, and the right of freedom of expression. Such considerations have set the threshold for the test. Under the first limb of the test, the complainant must show that the prospective defendant 'arguably', or (as a recent case has put it) 'well arguably', committed wrongdoing, and, under the second limb, the making of the order must be a 'necessity', which introduces considerations of alternative mechanisms and of proportionality.

As mentioned above, the current threshold for granting orders for preliminary discovery to identify a prospective defendant in Australia is relatively low compared to other similar jurisdictions. There is no requirement in Australia for a complainant to demonstrate a prima facie, or at least an arguable, cause of action in defamation, as required under the UK approach.

The Ontario approach

In March 2020, the Law Commission of Ontario, Canada, released its final report following its Defamation Law in the Internet Age review process ("**LCO Report**").

According to the LCO Report, Norwich Pharmacal type orders are often directed to Internet intermediaries. The test was established by the Ontario Divisional Court in *Warman v. Wilkins-Fournier*[9]. In this case, the Court held that the Rules of Civil Procedure must be interpreted in a manner consistent with Charter rights and values, including the right of freedom of expression and privacy interests.

The Court established a four-part test for determining whether a third party must disclose the identity of an anonymous online user. The court must consider whether:

a. *the unknown alleged wrongdoer had a reasonable expectation of anonymity;*
b. *the applicant had a prima facie case of defamation and was acting in good faith;*
c. *the applicant had taken reasonable steps to identify the anonymous party and had been unable to do so; and*
d. *the public interest favouring disclosure outweighed the freedom of expression and privacy interests of the unknown alleged wrongdoers.*

The Court also opined that anonymous speech should be afforded some degree of protection as a component of freedom of expression as protection for anonymous speech encourages more speech.[10] The Court also considered that it enhances public discourse, particularly in cases where public interest speech is motivated by fear of persecution or social ostracism. Anonymous speech also allows the author's message to be heard without being coloured by the author's identity and permits sensitive information to be conveyed without embarrassment.

Recommendations with respect to power of courts to order removal of content and reveal identity of originator

[9] *Warman v. Wilkins-Fournier*, 2010 ONSC 2126.
[10] *Ibid.*

Twitter considers the privacy and protection of its users to be of the utmost importance. It considers the application of the Federal Court Rules, and also the similar provisions within the state civil procedure rules, in pre-action discovery application where a complainant is seeking the user's details to be an insufficient protection of user's private information, and a threat to the general public's confidence in an ability to publish sensitive information, opinions, and beliefs without sufficient safeguards in place.

Twitter believes the safest and most effective regime would be for a prospective complainant to make an application to an Australian court, which satisfies a proscribed test, and then act in accordance with an order that is duly granted.

Twitter is concerned to limit such an order being made to Australian-based users only. It is submitted that Australia should consider adopting a similar test to that established in the *Warman v. Wilkins-Fournier* case. Twitter considers this test to appropriately balance the interests in pseudonymous/anonymous free speech and privacy on the one hand, and reputation and the administration of justice on the other hand.

Recommendation 7: Mandatory requirements for an offer to make amends for online publications

The amendments proposed by Section 15 regarding the requirements for an offer to make amends are somewhat unsuited for digital mediums. The current draft does recognise both the role of a digital intermediary (including its limited knowledge and involvement in the publication), and the recourse it can offer with respect to a matter (in circumstances where digital platforms are unfamiliar with the subject matter), and are therefore unable to offer the publication of corrective or clarifying material. It is important to note, however, that there is currently no convenient or reasonable location where intermediaries can publish corrective or clarifying material. Thus, there is no plausible way for an intermediary to make an offer to make amends, which would fail to obtain the intended objective with this recommendation.

Conclusion

A platform such as Twitter – which empowers originators to be in complete control of what is posted on their account – does not influence or control the originator. The right to free speech will be undermined by a complaints procedure that does not allow for consideration of the merits of the complaint. Prioritising expediency over substance would be a mistake. The MDAPs need to carefully balance a complainant's wish to have content removed by a digital intermediary with the public interest in continuing to allow digital intermediaries to facilitate the free flow of ideas and allow for robust discussions and debates.

Twitter is committed to working with the NSW Government, our industry partners, and other stakeholders to ensure that we have a better understanding of the issues at stake and can find the best way to approach this together. Working with the broader community, we will continue to test, to learn, to share, and to improve, so that our platform remains effective and safe for everyone.