The Manila Principles on Intermediary Liability Background Paper

Version 1.0, 30 May 2015

Table of Contents

1 Preface

This background paper describes six principles to guide government, industry and civil society in the development of best practices related to the regulation of online content through intermediaries.

These six principles are:

1. Intermediaries should be shielded by law from liability for third-party content
2. Content must not be required to be restricted without an order by a judicial authority
3. Requests for restrictions of content must be clear, be unambiguous, and follow due process
4. Laws and content restriction orders and practices must comply with the tests of necessity and proportionality
5. Laws and content restriction policies and practices must respect due process
6. Transparency and accountability must be built into laws and content restriction policies and practices

Each principle contains subsidiary points that expand upon the theme of the principle to cover more specific issues.

These Manila Principles were developed by an open, collaborative process conducted by a broad coalition of civil society groups and experts from around the world. This process was inspired in part by the International Principles on the Application of Human Rights to Communications Surveillance (the 13 Principles).[1]

Leading the work was a steering committee consisting of members from the Electronic Frontier Foundation (EFF, USA), the Centre for Internet and Society (CIS, India), Article 19 (UK), KICTANET (Kenya), Derechos Digitales (Chile), Asociación por los Derechos Civiles (ADC, Argentina) and Open Net (South Korea), who developed the first working draft of the background paper and principles, releasing it for broader public consultation and feedback in December 2014.

Over the following months the draft underwent further review by a diverse group of participants, over 30 of whom attended a face to face meeting in Manila, Philippines on 22-23 March 2015 where the principles were finalized. This background paper also takes into account comments from that broad group received during the public

[1] See *"International Principles on the Application of Human Rights to Communications Surveillance"*, accessed March 16, 2015. <https://necessaryandproportionate.net/>

consultation, however it has not undergone the same in-depth review by the broader public group, therefore it is being published by the steering committee alone, who take responsibility for any errors it may contain.

2 Introduction

All communication on the Internet requires a series of intermediaries to reach its audience. Their critical role in facilitating expression, and their ability to control and influence access to and availability of content, sometimes invites pressure from multiple actors who want to control, regulate, investigate, or silence online content and speech. Enforcing disproportionate or heavy-handed liability on intermediaries for the content of their users, including extending obligations that require them to monitor content and data being hosted or transmitted online, creates barriers to expression and innovation.[2] This hinders the right to freedom of expression as recognized at the international level.[3]

Thus, it is vital to maintain proportionate limits to intermediary liability for third party content as policies and laws are developed to defend and promote free expression and innovation. It is also valuable to encourage greater consistency in the laws and practices that apply to intermediaries. Such consistency is particularly needed given the borderless nature of the Internet and the global reach of intermediaries.

To this end, we have come together to develop a set of resources to help guide the development of intermediary liability policies that can foster and protect a free and open Internet. The resources that we are developing include a set of high-level principles on intermediary liability, a set of frequently asked questions about the principles, this background paper, and a jurisdictional analysis. These are intended as a civil society contribution to help guide companies, regulators and courts, as they continue to build out the legal landscape in which online intermediaries operate.

This background paper in turn explores emerging trends around intermediary liability and supports and expands upon each principle, drawing on existing international standards, human rights frameworks, jurisdictional jurisprudence, and research on intermediary liability laws, policies, and practices around the world.

---

[2] See Oxera Consulting LLP, *"The economic impact of safe harbours on Internet intermediary start-ups,"*, February 2015. <http://www.oxera.com/getmedia/cba1e897-be95-4a04-8ac3-869570df07b1/The-economic-impact-of-safe-harbours-on-Internet-intermediary-start-ups.pdf.aspx?ext=.pdf>

[3] See Article 19 International Covenant on Civil and Political Rights (ICCPR), Article 19 United Declaration of Human Rights (UDHR).

The background paper builds on reports at the international level published by the United Nations Organization for Education, Science and Culture (UNESCO), the World Intellectual Property Organization (WIPO), the United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, and the United Nations Department of Economic and Social Affairs (UNDESA). The paper also references best practice and case law from liability regimes across Argentina, Canada, Chile, India, Kenya, United Kingdom and USA. Lastly, the background paper draws upon research by the Centre for Democracy and Technology (CDT), the Association for Progressive Communications (APC), Article 19 and other civil society, academia, and domain experts.

3 Scope

This paper does not attempt to consider all aspects of the relationship between intermediaries and the users whose speech they help enable, their readers and other audiences of online speech (though a number of the projects that we draw upon and cite below, do have a broader scope than this one). Rather, we are concerned solely with the laws, policies, norms, and practices that relate to how intermediaries handle third-party Internet content that could raise criminal or civil liability issues for them or for their users. Specifically, the principles are meant to be directed at laws, policies, norms, practices, and private terms of service that relate to content restriction, including removal, blocking or filtering by intermediaries.

In general, we are not concerned here with the particular legal basis on which liability for content may arise or the reason why a party might want to have it restricted; that is, for example, whether the content may be allegedly defamatory, copyright-infringing, seditious or fraudulent. As will be explained below, this approach differs from that taken by the law of some countries, which may establish different intermediary liability rules in one or more of these cases.

Neither do we generally draw a distinction between restriction of content such as through blocking, and its outright removal. This approach too does not necessarily correspond to the approach taken by the law. In many jurisdictions governments with liability regimes that regulate the removal of content have separate legal provisions allowing for the blocking of content. These provisions often allow restriction based on different criteria and extent of liability, and courts can also establish their own standards on a case-by-case basis.

We have endeavoured to take a unified approach because no matter the ground of liability the content may attract, or the nature of the restriction, there are many common principles apportioning potential liability between the intermediary and

the user that can help ensure that the rights and interests of all parties are respected, as well as the public interest.

Other laws, policies, norms, and practices that intermediaries may adopt or enforce relating to Internet content, but which fall outside of the shadow of potential liability for that content, are not directly addressed here even though they too may have implications for users' freedom of expression online. This includes, in particular, issues of network neutrality. However, some closely associated issues, namely particular aspects of how user privacy is upheld in the implementation of a liability regime, are included within the scope of the principles where relevant.

We considered whether the principles should be addressed only to intermediaries, or only to governments, but limiting our audience to either option would have restricted the principles from addressing all appropriate targets capable of actioning the intended reforms. Our approach thus recognizes that specific stakeholders have distinctive roles in any intermediary liability regime, and hence we address some recommendations to all concerned stakeholders and others to specific actors. A similar hybrid audience is addressed in other international documents, such as the United Nations Guidelines for Consumer Protection.[4]

The principles that we put forward are not as prescriptive as laws or policies, although, in the context of content removal, their implications are not neutral as to the model of intermediary liability that is to be preferred. Without prescribing a single model for adoption in all cases, the application of the principles favors a model that provides expansive protections against liability, whilst recognizing that legal and operational considerations may require some intermediaries, particularly those who are not mere conduits, to assume greater responsibilities for content than others (see below under "Intermediary Liability Models" where these terms are explained). In general the greater the obligations that a model imposes upon intermediaries, the more human rights safeguards will be required to establish best practices for that model that are consistent with these principles.

Finally, it should be underlined that as a civil society document, the principles intended to be used in advocating for laws, policies and procedures that uphold the human rights of users. Due to the symbiotic relationship between intermediaries and their users, the limitation of intermediary liability naturally, also serves the intermediary's economic interests—but although this is important, the aim and objective of these principles are not to protect the economic interests of intermediaries themselves. To that extent, the principles that we develop can be

[4] See United Nations, *"United Nations Guidelines for Consumer Protection 1995"*, 2003, accessed March 16, 2015. <http://www.consumersinternational.org/media/33866/consumption_en.pdf>

distinguished from industry-developed principles such as the 2007 Principles for User Generated Content Services.[5]

4 Definitions

4.1 Intermediaries

In general terms, an intermediary is "any entity that enables the communication of information from one party to another"[6] As for online or Internet intermediaries (whom we will also refer to simply as "intermediaries" from this point), we have operated under a broad definition, shared with the recent UNESCO report "Fostering Freedom Online: The Role of Internet Intermediaries"[7] from which this background paper draws extensively (and which in turn draws on a meta-study of previous work, including reports from the OECD and CDT). This definition holds:

> *Internet intermediaries bring together or facilitate transactions between third parties on the Internet. They give access to, host, transmit and index content, products and services originated by third parties on the Internet or provide Internet-based services to third parties.*[8]

Examples of intermediaries falling within that definition would include:

- Internet Service Providers (ISPs)
- Search engines
- Social networks
- Cloud service providers
- E-commerce platforms
- Web hosting companies
- Domain name registrars
- Content aggregators
- Individuals who run open Wi-Fi hotspots, Tor nodes, etc.

There are some edge cases that do not clearly fall into this definition, depending on one's interpretation. These include manufacturers of products (rather than services)

---

[5] See *"Principles for User Generated Content Services"*, accessed March 16, 2015. <http://www.ugcprinciples.com/>

[6] T.F. Cotter, *"Some Observations on the Law and Economics of Intermediaries,"* Mich.St.L.Rev.67 (2006): 68-71.

[7] MacKinnon, R, Hickock, E, Bar, A and Lim, H, *"Fostering Freedom Online: The Role of Internet Intermediaries,"* Unesco, 2014, p.19, accessed March 16, 2015. <http://unesdoc.unesco.org/images/0023/002311/231162e.pdf>, henceforth "the UNESCO report".

[8] OECD, *"The Economic and Social Role of Internet Intermediaries,"* April 2010, p.9, accessed March 16, 2015. <http://www.oecd.org/internet/ieconomy/44949023.pdf>

that are used for accessing content, such as Web browser and Internet filtering software—though these are amongst the six classes of intermediary that APC identifies in a 2014 paper, "Internet Intermediary Liability: Identifying International Best Practices for Africa".[9] The scope of this paper is somewhat narrower than the APC research, in that we are only considering intermediaries' liability for third-party content, which will seldom apply to product vendors.

Another edge case is that of content producers, who are normally excluded as intermediaries, but in some cases may fulfill both roles. For example Article 19 holds the position that online newspapers should be treated as intermediaries for the purposes of user-generated content (UGC), even while they also remain responsible for their own content.[10] A troublesome case illustrating this distinction is that of *Delfi AS v Estonia*, where the European Court of Human Rights[11] found no violation of the right to freedom of expression in a case where a newspaper was held liable for its users' comments. This was despite the fact that the newspaper had promptly removed the content at issue upon notice in compliance with the ECD.[12]

When developing liability rules for intermediaries, it is important that legal requirements are appropriate and proportional to the function and size of the intermediary. Thus the definition of an intermediary that we use may not coincide with the legal definition of an intermediary in a particular jurisdiction, which in any case, differs markedly from one country to another. For example, under Chilean net neutrality law, intermediaries are limited to commercial platforms,[13] while in Indian

---

[9] Zingales, Nicolo*, "Internet Intermediary Liability: Identifying International Best Practices for Africa*," November *2013,* p. 4, accessed March 16, 2015. <https://www.apc.org/en/pubs/internet-intermediary-liability-identifying-best-p>, (henceforth "APC report")

[10] See Article 19, *"Third Party Intervention Submissions by Article 19,"* accessed March 16, 2015. <http://www.article19.org/data/files/medialibrary/37592/Delfi-intervention-A19-30052014-FINAL.pdf>

[11] See European Court of Human Rights, *"Delfi vs. Estonia,"* October 10, 2013, accessed March 16, 2015. <http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-126635 - {"itemid":["001-126635"]}>

[12] See Article 19, *"Article 19 to European Court Online news sites should not be strictly liable for third party content,"* June 18, 2014, accessed March 16, 2015. <http://www.article19.org/resources.php/resource/37592/en/article-19-to-european-court:-online-news-sites-should-not-be-strictly-liable-for-third-party-comments>

[13] See *"Ley General de Telecomunicaciones No. 18.168 de 1982"*, Biblioteca del Congreso Nacional de Chile, accessed March 16, 2015. <http://www.leychile.cl/Navegar?idNorma=29591>

Internet law intermediaries are much more broadly defined.[14] APC's country studies in Africa also showed quite varied approaches.[15]

For simplicity of understanding, it can be useful to group intermediaries into broad categories, and naturally various approaches to this exercise have been adopted:

- A 2013 report[16] published by the Organization of American States (OAS) identifies the most relevant intermediaries as ISPs, website hosting providers, social networking platforms, and search engines.
- The UNESCO report[17] groups them into three general types: ISPs, search engines, and social networks.
- CDA 230 speaks of interactive computer services, information content providers and access software providers.
- The DMCA separates them into communications conduits, content hosts and search service and application service providers.
- The ECD categorizes intermediaries according to class based on their function and includes hosts, conduits, and caching. As a note, such a distinction does not address linking activities of search engines that may fall under different types of intermediaries such as host or conduit.

4.2 Intermediary Liability

The definition of "intermediary liability" in our context is not quite as broad as it sounds. It refers to the legal liability of Internet intermediaries for content contributed by, or activities carried out by, third parties.[18] It does not include liability that intermediaries may incur for their own content, or for other reasons altogether, such as taxation liability, liability for fraud or breach of contract, and liability of intermediaries to their users (eg. for custody of their data).

---

[14] See *"Information and Technology Act 2000,"* Department of electronics & Information Technology, Ministry of Communications & IT, Government of India, accessed March 16, 2015. <http://deity.gov.in/content/view-it-act-2000>

[15] See APC report.

[16] Inter-American Commission on Human Rights. Office of the Special Rapporteur for Freedom of Expression, *"Freedom of expression and the Internet,"* 2013, p. 40, accessed March 16, 2015. <http://www.oas.org/en/iachr/expression/docs/reports/2014_04_08_Internet_ENG_WEB.pdf> (henceforth "OAS report")

[17] UNESCO Report.

[18] See Edwards, Lilian, "*Role and Responsibility of Internet Intermediaries in the Field of Copyright and Related Rights*," WIPO, 2011, p. 3, accessed March 16, 2015. <http://www.wipo.int/export/sites/www/copyright/en/doc/role_and_responsibility_of_the_internet_intermediaries_final.pdf>, henceforth "the WIPO report."

4.3 Governments

Governments are the parties who issue content restriction orders, which have the force of law in a particular jurisdiction. Except where otherwise specified, references to governments include not only the executive branch of government, but also courts. Where this is not the intention (eg. see the discussion of Principle VI.d in this background paper), the two will be treated separately.

4.4 User Content Provider

The term "user content provider" refers to users who upload material or share material with others via an intermediary. This material may be user-generated content, or it may be content from a third-party that the user uploads or publishes to the intermediary's service. The user content provider is the person to whom primary liability may attach if the content is found unlawful by a court of law.

4.5 Content

Content includes include information, expression and communication provided by users. Content may be expressed in various forms including text, video, audio and any combination of these, and may be real-time or recorded. Content therefore includes blog posts, photographs and videos shared online, live voice or text communications made using telephony applications, and comment threads on online articles, amongst many other examples that could be given.

4.6 Content Restriction Orders

The Manila Principles refer to "content restriction orders" as a shorthand reference to requests issued by any branch or agency of the government. This includes court orders and executive orders which are legally binding for the restriction, including removal, blocking, or filtering of online content or platforms.

4.7 Content Restriction Requests

The Manila Principles refer to "restriction requests" as a shorthand to reference requests issued directly to intermediaries by private third parties for the restriction, including removal, blocking or filtering of information, or executive requests that seek to induce intermediaries to remove content under their terms of service or by an untested allegation that such content is illegal.

5 Legal Background

5.1 Human Rights Law

The standards from which a basic intermediary liability framework can be constructed already exist, most notably in the form of international and regional human rights instruments, as well as related soft law instruments and opinions such as the work of the UN Special Rapporteur on freedom of expression.[19]

The most relevant international legal human rights standard that underpins the principles, although it is not the only one, is the right to freedom of expression, as enshrined in the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and various regional instruments.

Amongst the most useful high-level commentaries illustrating the application of this right to online intermediaries was made in 2011 by the UN Human Rights Committee:

> *Any restrictions on the operation of websites, blogs or any other internet-based, electronic or other such information dissemination system, including systems to support such communication, such as internet service providers or search engines, are only permissible to the extent that they are compatible with paragraph 3. Permissible restrictions generally should be content-specific; generic bans on the operation of certain sites and systems are not compatible with paragraph 3. It is also inconsistent with paragraph 3 to prohibit a site or an information dissemination system from publishing material solely on the basis that it may be critical of the government or the political social system espoused by the government.*[20]

International corporate social responsibility, consumer law and competition law frameworks also help underpin existing and developing intermediary liability regimes and we have built on these in our report too. For example, the United Nations Guiding Principles on Business and Human Rights requires *inter alia* that "business enterprises should establish or participate in effective operational-level grievance mechanisms for individuals and communities who may be adversely

---

[19] La Rue, Frank, *"Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression,"* United Nations, Human Rights Council, April 17, 2013, accessed March 16, 2015. <http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf>

[20] UN Human Rights Committee, *"General comment no. 34 on Article 19: Freedoms of opinion and expression,"* International Covenant on Civil and Political Rights (ICCPR), 2011, paragraph 43, accessed March 16, 2015. <http://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf>

impacted", and the United Nations Guidelines for Consumer Protection aims to "assist countries in curbing abusive business practices by all enterprises at the national and international levels which adversely affect consumers".

In the longer term, these existing principles could be refined into a more specific global legal framework that establishes baseline limitations on intermediary liability, through an inclusive, multi-stakeholder process. Whilst such multi-stakeholder policy development processes are in their infancy, the NETmundial meeting of April 2014[21] is an early experiment along these lines.

Meanwhile, drawing on existing principles and research and the explication of the same by noted scholars and activists,[22] we propose a set of global baselines to guide the development and implementation of intermediary liability regimes and practice.

5.2 Trade and Competition

Another context in which intermediary liability rules are considered is that of trade and competition law and policy. This informed our analysis also, as broad variation amongst the legal regimes of the countries in which online intermediaries operate increases compliance costs for companies. It may discourage them from offering their services in some countries due to the high costs of localized compliance. A recent Oxera Consulting study found:

> *Legal clarity would be beneficial not only within, but also across, countries. Currently, intermediaries need to ensure compliance at the national level, and hence require legal expertise and compliance processes for each country. A more uniform approach across regions would allow companies to follow a clear legal framework, thereby lowering transaction costs and facilitating the expansion of intermediaries across jurisdictions. This is likely to benefit users by increasing choice and promoting competition between intermediaries.*[23]

This was found to lead to a possible increase in start-up success rates for intermediaries in countries adopting a liability regime with clearly-defined requirements, as well as increasing expected profits. Conversely, "Intermediary

[21] See NETmundial web site, accessed March 16, 2015. <http://www.netmundial.br/>

[22] See Association for Progressive Communications, *"UN encourages community responses to online hatred in new report,"* APC release, June 26, 2014, accessed March 16, 2015. <http://www.apc.org/en/press/un-encourages-community-responses-online-hatred-ne>

[23] Oxera, p.11.

start- ups are likely to be held back if the IIL [Internet Intermediary Liability] regime is not clear or entails complex compliance requirements."[24]

Similarly, the Internet Association has argued that differences in intermediary liability regimes can operate as a barrier to cross-border trade as bad laws are bad for local and foreign businesses.[25] Therefore, to ensure that citizens of a particular country have access to a robust range of speech platforms, each country should work to harmonize the requirements that it imposes upon online intermediaries with the requirements of other countries, where possible. While a certain degree of variation between what is permitted in one country as compared to another is inevitable, all countries should agree on certain limitations to intermediary liability.

In recognition of differences between regimes, a multi-stakeholder initiative called the Internet & Jurisdiction Project argues for the development of common principles of due process capable of application in multiple jurisdictions:

> *The Internet is transnational. Its cross-border nature challenges the international legal system that is based on a patchwork of separate national sovereignties. No reliable framework exists to handle this challenge. The resulting legal competition has unintended consequences including: increased jurisdictional conflicts, tensions between actors and a risk of fragmentation.*[26]

The Internet & Jurisdiction Project has identified as a challenge the lack of appropriate procedures to handle an increasing number of request from courts and authorities to ISPs in other jurisdictions, including attempted domain seizures, content takedowns and related access to user data. Their proposal attempts to address this through the definition of a draft architecture for how requests are submitted and how they are handled.[27] For the standardization of request submission, the proposal includes two parts: the development of standardized formats and the building of mutualized databases. For the request handling the proposal includes also two parts: rules to allow process predictability and dispute management.

Whilst the trade and competition dimension of intermediary liability is therefore acknowledged, nevertheless the development of intermediary liability regimes as a

---

[24] Oxera, pp. 2-3.

[25] Internet Association, "*Harmonizing Intermediary Immunity for Modern Trade Policy,"* May 2014, accessed March 16, 2015. <http://internetassociation.org/wp-content/uploads/2014/05/May-2014-Section230.pdf>

[26] Internet & Jurisdiction Project*, "Progress Report 2013/14,"* p. 5, accessed March 16, 2015. <http://www.internetjurisdiction.net/progress-report-2013-14/>

[27] *Id.*

response to pressure via international trade agreements, such as the Trans-Pacific Partnership, is seen as not a legitimate or inclusive process, and these principles aim for a more holistic treatment of their subject matter, as well as specifically critiquing exclusionary mechanisms of intermediary policy development in principle VI.

6 Intermediary Liability Practices

6.1 Intermediary Liability Models

Intermediary liability for third-party content occurs "where governments or private litigants can hold technological intermediaries such as ISPs and websites liable for unlawful or harmful content created by users of those services"[28]—including for their failure to block or filter such content. Intermediary liability in this sense can arise from a multitude of issues such as copyright infringements, digital piracy, trademark disputes, network management, spamming and phishing, "cybercrime", defamation, hate speech, and child pornography, as well as covering both illegal content and offensive but legal content, and engaging areas of law ranging from censorship, to broadcasting and telecommunications laws and regulations, and privacy law.[29]

The Manila Principles also cover circumstances in which intermediaries may restrict content in anticipation of possible liability (or for other reasons) pursuant to their own terms of service; an important inclusion because of the trend for intermediaries to be pushed to take "voluntary measures" against users, as explained further below. To omit the mechanism of terms of service based content restriction from consideration would therefore leave a grave gap in the principles.

The Manila Principles have been developed in the context of existing intermediary liability regimes, without in any way being constrained to remain compatible with these regimes. In this regard, there are three general approaches to intermediary liability that have been discussed in much of the recent work in this area, including the Centre for Democracy and Technology's 2012 report, "Shielding the Messengers: Protecting Platforms for Expression and Innovation." These approaches are:

1. Expansive protections against liability
2. Conditional immunity from liability

---

[28] Center for Democracy and Technology, "*Intermediary Liability: Protecting Internet Platforms for Expression and Innovation*," April 2010, p.1, accessed March 16, 2015. <https://cdt.org/insight/protecting-internet-platforms-for-expression-and-innovation/>.

[29] See Comninos, Alex, "*The Liability of internet intermediaries in Nigeria, Kenya, South Africa and Uganda: An uncertain terrain*," Association for Progressive Communications, 2012, p.6, accessed March 16, 2015. <http://www.apc.org/en/pubs/liability-internet-intermediaries-nigeria-kenya-so>

3. Primary liability for third-party content[30]

Expansive protections are provided for intermediaries, for example, in the regime established under section 230(c) of the Communications Decency Act (CDA) in the United States which establishes that intermediaries should not be considered as publishers and exempts them from liability for most types of third-party content (although not, notably, for intellectual property infringements).

The conditional immunity from liability approach, which CDT terms "conditional safe harbor", seeks to balance protection of intermediaries from liability while defining certain roles for them with respect to unlawful content. Under this approach an intermediary receives protection from liability for user conduct, only if the intermediary meets certain conditions such as compliance with a statutory "notice and notice" or "notice and takedown" system. The Canadian liability framework creates conditional safe harbor for intermediaries by establishing a "notice and notice" system for copyright infringements, with effect from 2015. Under the "notice and notice" system, the primary responsibility of the intermediary upon receiving a removal request is to forward the notification to subscriber (or explain to the claimant why they cannot forward it). This enables the dispute to be directly resolved between the complainant and the content producer and no content is taken down by the intermediary. England's Defamation Act 2013 also establishes a "notice and notice" system. Drawing upon Canada's intermediary liability framework and the English Defamation Act, Article 19 has developed safeguards targeted at the liability regime of "notice and notice".[31]

The conditional immunity approach also corresponds to the regime established under the Digital Millennium Copyright Act (DMCA), Title 512, which exempts intermediaries from liability for copyright infringements if they comply with certain conditions, such as compliance with a statutory "notice and take-down" system. Under such a system, intermediaries need to respond to take-down requests and take down copyright infringing content in order to keep their protection from liability. In such regimes problems arise due to ambiguity over what is unlawful and an incentive structure skewed towards content removal.

---

[30] See Center for Democracy and Technology, *"Shielding the Messengers: Protecting Platforms for Expression and Innovation,"* 2012, pp.4-15, accessed March 16, 2015. <https://www.cdt.org/files/pdfs/CDT-Intermediary-Liability-2012.pdf>

[31] For more detail see Article 19, *"Internet Intermediaries: Dilemma of Liability Q and A,"* August 2013, accessed March 16, 2015. <http://www.article19.org/resources.php/resource/37243/en/internet-intermediaries:-dilemma-of-liability-q-and-a>

In Europe internet intermediaries[32] are afforded protection from liability for all types of content (including intellectual property) on an equal basis, under what in practice amounts to a conditional immunity model, but which differentiates between different classes of intermediaries. Under the E-Commerce Directive (ECD) Article 14 Internet intermediaries are afforded protection from intermediary liability for being a mere conduit for information, for caching information, or for hosting information.[33] Provided these activities are "of a mere technical, automatic and passive nature" and the intermediary "has neither knowledge of nor control over the information which is transmitted or stored" they are afforded protection from liability.[34] For those who operate as mere conduits, or as caching service providers, they are protected from liability for that content, as long as they did not modify transmitted information, and did not collaborate with recipients of its services in order to undertake illegal activity. Protection from liability for hosting content is conditional on the service provider having been unaware of content on its network, and once becoming aware of unlawful content on its network, acting expeditiously to remove it.

The primary liability approach is described in the CDT report as "blanket or strict liability for intermediaries", though this is not a comprehensive description because it also refers to a situation where there is an onerous and/or vague negligence standard for intermediaries. This is the case in China and Thailand, for example, where intermediaries are frequently held liable for third-party content, thereby providing them with a strong incentive to pre-emptively censor that content.[35] This can reflect a conscious policy on the part of the government or other actors to control certain illegal, unlawful and undesirable content on the Internet by specifically holding intermediaries responsible for such content, because they

[32] Termed "information society services". For further explanation see University of Oslo, "The E-Commerce Directive Article 14:Liability exemptions for hosting third party content," 2011, accessed March 16, 2015. <https://www.duo.uio.no/bitstream/handle/10852/19450/117618.pdf>

[33] European Parliament and the Council of the European Union, *"European E-Commerce Directive 2000/31 (ECD),"* articles 12-15, accessed March 16, 2015. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32000L0031&from=EN>This does not include liability for the protection of individuals with regard to the processing of personal data which "is solely governed by Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (2) and Directive 97/66/EC of the European Parliament and of the Council of 15 December 1997 concerning the processing of personal data and the protection of privacy in the telecommunications sector (3)"

[34] Id., paragraph 42.

[35] Tao, Qian, *"Legal framework of online intermediaries' liability in China"*, Info 14, 6 (2012): 59-72; *"The knowledge standard for the Internet Intermediary Liability in China,"* International Journal on Law Information Technology 20, 1 (2012): 1-18.

provide a more convenient locus of control than end users.[36] Similarly, intermediary liability for third party content can be the default position in the context of laws, such as those of Kenya and Nigeria, that do not differentiate between an intermediary and an author and publisher of original content.[37] In such cases, determining when intermediaries should be held liable or not entails much of the same comprehensive review of the law as would apply to the original author of the content.

6.2 Approaches to Content Restriction

There is a continuum of content restriction mechanisms that have been adopted by intermediaries or governments and that range from being the least balanced and accountable in protecting users freedom of expression and intermediaries from heavy or disproportionate liability, and those that are more so. These do not quite overlap with the three models described above, because intermediaries can and do employ mechanisms that go beyond their legal obligations. The below mechanisms range from 1 as the least accountable to 5 as the most accountable:

1. **The restriction of content that does not involve independent human review,** for example through the use of automated tools that flag and remove content or restrict its accessibility without express user consent. This is to be distinguished from systems that allow users of social networks or ISPs to opt-in to the use of automated obscenity, spam, and abuse filters, but an important feature of these is that they only affect the user who consents to them.[38] In practice, the application of such systems can be multi-layered—with the initial identification of content being through an automated process, and subsequent actions reviewed by humans, which merges into the second mechanism. Such automated filters can be developed by private companies or in collaboration with governments.
2. **The unilateral removal of content** by the intermediary without legal compulsion in response to a private third party request received, without affording the uploader of the content the right to be heard or access to remedy. Although an intermediary may be within their rights to act on a private request if the content is in violation of their content policy, such action taken without providing notice, the right to be heard, or access to

[36] McKinnon, Rebecca, "Are China's demands for self-discipline spreading to the West?" McClatchy, January 18, 2010, accessed March 16, 2015. <http://www.mcclatchydc.com/2010/01/18/82469_commentary-are-chinas-demands.html?rh=1>

[37] *Id.*

[38] Some intermediaries may be able to claim the implied consent of their users for filtering of malware, though we do not express a view on that.

remedy to the content uploader, raises accountability issues and impinges on free speech.

3. **Notice and takedown mechanisms** in which content orders are not assessed by an independent authority, but instead incorporate, as the DMCA attempts to do, an effective appeal and counter-notice mechanism in which intermediaries remove content upon receiving a take down request, but provide the uploader of the content notice of its removal and a right of appeal. Where this breaks down is that the cost and incentive structure is weighted towards removal of content in the case of doubt or dispute, resulting in more content, including legitimate content, being taken down and staying down.
4. **Notice and notice regimes** in which the intermediary passes on a removal request to the uploader of information. Notice and notice regimes typically provide strong social incentives for those whose content is reported to be unlawful to remove the content, but do not legally compel them to do so. If legal compulsion is required, a court order must then be separately obtained. As noted above, Canada has followed this approach for copyright works, and England for defamation.
5. **Notice and judicial takedown regimes** require a complaining party to obtain a judicial order for the removal of content before the intermediary will respond by taking the content down, and with the possibility of appeal or judicial review. This model, adopted in jurisdictions like Chile in the context of copyright, balances the rights of the user and the interests of the party requesting content restriction in many cases, but has been criticized by some on practicality and efficiency grounds.

The first three mechanisms, which involve content being restricted by intermediaries unilaterally, or by unadjudicated allegations of illegality (eg. notices from private parties) are not supported by the standards in the Manila Principles, although principle V.a does permit a non-judicially ordered removal in the most clear and serious exceptional circumstances provided by law, generally involving *manifest illegality*, and/or where the harm to the victim is otherwise irreparable—and then only with necessary safeguards against abuse as further set out in the Manila Principles.

As we will see, it is the latter two mechanisms that the Manila Principles support in most cases.

7 Manila Principles for Intermediary Liability

Principle I: Intermediaries should be shielded by law from liability for third party content

In order to preserve the right to freedom of expression, while enabling an environment for innovation by users and organizations, governments should provide legal protections exempting intermediaries from liability for third party content on their networks or platforms.

The subsections below define some key aspects that any legal regime addressing intermediary liability should address:

I.a. Any rules governing intermediary liability must be provided by laws, which must be precise, clear, and accessible.

The rules and obligations that governments impose on intermediaries should be constitutionally valid and in compliance with all applicable legal norms of due process. Intermediaries should resist restricting content where such criteria of constitutionality and due process have not been followed. Imposing liability on internet intermediaries without providing clear and accessible guidance as to the precise type of content that is not lawful and the precise requirements of a legally sufficient notice encourages intermediaries to over-remove content.

This principle also encompasses the principle of legality, a fundamental aspect of all international human rights instruments, which is a basic guarantee against the state's arbitrary exercise of its powers. For this reason, any restriction on human rights, including the right to free expression, must be "provided" or "prescribed" by law.[39]

Furthermore, the meaning of "law" implies certain minimum qualitative requirements of clarity, accessibility, and predictability as well as democratic process.

The OAS Report states:

> *the first condition of the legitimacy of any restriction of freedom of expression—on the Internet or in any other area—is the need for the*

[39] The meaning of legality has been derived from the principle of legality, as defined in the International Principles on the Application of Human Rights to Communications Surveillance. "*International Principles on the Application of Human Rights to Communications Surveillance,"* May 2014, accessed March 16, 2015. <https://en.necessaryandproportionate.org/>

> *restrictions to be established by law, formerly [sic] and in practice, and that the laws in question be clear and precise.* [40]

The Human Rights Committee has clarified the meaning of "law" for the purposes of Article 19 ICCPR stating that:

> *A 'law', must be formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly and it must be made accessible to the public. A law may not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution. Laws must provide sufficient guidance to those charged with their execution to enable them to ascertain what sorts of expression are properly restricted and what sorts are not.*[41]

The European Court of Human Rights has followed a similar approach in its jurisprudence. In particular, it has held that the expression "prescribed by law" implies the following requirements:

> *Firstly, the law must be adequately accessible: the citizen must be able to have an indication that is adequate in the circumstances of the legal rules applicable to a given case. Secondly, a norm cannot be regarded as a "law", unless it is formulated with sufficient precision to enable the citizen to regulate his conduct; he must be able—if need be with appropriate advice—to foresee, to a degree that is reasonable in the circumstances, the consequences which a given action may entail.*[42]

The UNESCO report points out:

> *Policy, legal, and regulatory goals affecting intermediaries must be consistent with universal human rights norms if states are to protect online freedom of expression and if companies are to respect it to the maximum degree possible. Governments need to ensure that legal frameworks and policies are in place to address issues arising out of intermediary liability and absence of liability. Legal frameworks and policies affecting freedom of expression and privacy should be contextually adapted without transgressing universal standards, be consistent with human rights norms including the right to freedom of expression, and contain a commitment to principles of due process and fairness.*

---

[40] OAS report, pp. 26-27.

[41] See UN Human Rights Committee, *"General comment no. 34 on Article 19: Freedoms of opinion and expression."*

[42] European Court of Human Rights, *"Sunday Times vs. United Kingdom,"* April 26, 1979, paragraph 49, accessed March 16, 2015. <http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-57584 - {"itemid":%5B"001-57584"%5D}>

> *Legal and regulatory frameworks should also be precise and grounded in a clear understanding of the technology they are meant to address, removing legal uncertainty that would otherwise provide opportunity for abuse or for intermediaries to operate in ways that restrict freedom of expression for fear of liability.*[43]

Constitutional law in each country determines the precise requirements that underpin the legality of a law made by its legislature, and administrative law determines the legality of lawmaking by the executive branch. Even when a law is constitutional, this does not necessarily mean that an intermediary should comply with it. If that law contravenes international human rights standards, and if the intermediary does not operate from that country or otherwise subject to its jurisdiction, then the intermediary is both legally and ethically justified in declining to enforce laws that would restrict the availability of its content within its borders.

Article 19 noted in its 2013 report on intermediary liability:

> *Approximately 30 participating States have laws based on the EU E-Commerce Directive. However, the EU Directive provisions rather than aligning state level policies, created differences in interpretation during the national implementation process. These differences emerged once the national courts applied the provisions. These procedures have also been criticized for being unfair. Rather than obtaining a court order requiring the host to remove unlawful material (which, in principle at least, would involve an independent judicial determination that the material is indeed unlawful), hosts are required to act merely on the say-so of a private party or public body. This is problematic because hosts tend to err on the side of caution and therefore take down material that may be perfectly legitimate and lawful.*[44]

I.b. Intermediaries should be immune from liability for third-party content in circumstances where they have not been involved in modifying that content.

The essence of the first principle is expressed here, recommending the adoption of an intermediary liability regime that provides expansive protection from liability. The proviso that the intermediary's immunity from liability applies when they have not modified the content in question is intended to be a narrow one. The intermediary should be liable, in that case, only for the modifications that they have

[43] UNESCO report, p. 186.

[44] See Article 19, *"Internet Intermediaries: Dilemma of Liability,"* 2013, accessed March 16, 2015. <http://www.article19.org/data/files/Intermediaries_ENGLISH.pdf>

made—for example, if an user's comment on a news story is edited by an intermediary in a way that makes it defamatory, liability for those edits would lie on the intermediary. It is not intended that technical modifications, such as the addition of HTTP headers by an caching intermediary, would give rise to liability for the cached content.

This is consistent with the 2011 Joint Declaration on Freedom of Expression and the Internet, which provides:

> *No one who simply provides technical Internet services such as providing access, or searching for, or transmission or caching of information, should be liable for content generated by others, which is disseminated using those services, as long as they do not specifically intervene in that content or refuse to obey a court order to remove that content, where they have the capacity to do so ("mere conduit principle").* [45]

The Manila Principles do recognize one other qualification on the intermediary's immunity from liability in III.d below, namely the law may (but not must) require the intermediary to pass on qualifying notices of the illegality of content to the user who provided that content. This is the can be regarded as a minimal safe harbor requirement compatible with the Manila Principles.

I.c. Intermediaries must not be held liable for failing to restrict lawful content.

This is closely related to the previous point, but specifies that not only must an intermediary not be liable for third-party content in circumstances where they have not been involved in modifying that content, but must also not be made liable for failing to restrict lawful content. This principle is particularly important in light of the emerging issue of the right to be de-indexed by search engines, an application of the European Data Protection Directive commonly known as the right to be forgotten.

[45] The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, Article 19, Global Campaign for Free Expression, and the Centre for Law and Democracy, "*Joint Declaration on Freedom of Expression and the Internet,"* June 1, 2011, p. 2, accessed March 16, 2015. <http://www.osce.org/fom/78309> (Hereinafter "Joint Declaration on Freedom of Expression and the Internet).

In May 2013, the European Court of Justice (CJEU) in the case *Google Spain SL and Google Inc V. AEPD* (C-131/12),[46] ruled that individuals have the right to request search engines to remove links to websites that are displayed when queried for using the name of an individual. Importantly, the judgment further clarified that search engines do constitute a "data controller" as defined in the European Union Data Protection Directive. Though the judgment provided individuals with the ability to request removal, it also defined safeguards that must be adhered to including that the removal of content must comply with requirements under the Directive i.e the right to erasure and the right to object,[47] and that if an individual's request is not granted by the search engine - they have the right to take the complaint to the competent authority.[48]

Although the CJEU ruling failed to take freedom of expression properly into account, it did recognize the potential interference that the removal of links from the list of results could have on the legitimate interest of Internet users "potentially interested in having access to that information". The ruling recommended that a fair balance should be sought between that interest and the data subject's Fundamental rights under Articles 7 and 8 of the Charter, recognizing that balance may depend on consideration of the nature of the information in question and its sensitivity for the data subject's private life and on the interest of the public in having that information.

In the context of intermediary liability, this judgment has a number of implications and highlights an intersection between intermediary liability and data protection. The decision provides European citizens the right to request that links to "inadequate, irrelevant or no longer relevant" information be removed from search results, however, the information does not disappear from the internet altogether. The ruling establishes a regime similar to notice and takedown and while pages that are de-linked will still be available in their original forms online, search engines are put in the position of having to review and take action (by either removing or

---

[46] See Court of Justice of the European Union, "*Google Spain SL, Google Inc. v Agencia Española de Protección de Datos (AEPD), Mario Costeja González*," May 13, 2014, accessed March 16, 2015. <http://curia.europa.eu/juris/document/document.jsf?text=&docid=152065&pageIndex=0&doclang=en&mode=req&dir=&occ=first&part=1&cid=264438>

[47] See Ausloos, Jef, "*European Court Rules against Google, in Favour of Right to be Forgotten*," LSE Media Policy Project Blog, accessed March 16, 2015. <http://blogs.lse.ac.uk/mediapolicyproject/2014/05/13/european-court-rules-against-google-in-favour-of-right-to-be-forgotten/>

[48] See Court of Justice of the European Union, "*An internet search engine operator is responsible for the processing that it carries out of personal data which appear on web pages published by third parties,*" Press Release 70/14, May 13, 2014, accessed March 16, 2015. <http://curia.europa.eu/jcms/upload/docs/application/pdf/2014-05/cp140070en.pdf>

maintaining information) on private removal requests. As a response to the judgment, Google has put in place a request mechanism for its European user base.[49]

A number of concerns[50] have also been raised on the impact of the judgment on free expression and the right to access. Will such a right enable individuals to remove speech pertaining to them that they disagree with? Or, as the judgment is limited to search engines and search queries, is the obligation limited to search engines not directing to the original site of the information? Further, will such a right create an access asymmetry—as information is never deleted from the internet and this increases the gap between those who know where to find information and those who need a search engine to do so. Even as there are talks to extend the ruling outside the European Union[51] the ruling has been found inconsistent with systems such as the Inter-American Commission on Human Rights.[52]

The implementation of the ruling so far has also raised issues that remain unresolved such as information inequality,[53] private censorship, the impact on public discourse about political issues, that it may be used by people in positions of power to manipulate press coverage and that "outdated" information about an individual, removed in part because they are not a public figure may in due course, become very relevant if that individual immediately goes on to seek public office.[54] These concerns are magnified in contexts where the right of the public to have information about the present and the past is threatened.

A key aspect of this ruling is that it doesn't relate to libelous or defamatory information. It censors lawful content that contains personal information because it may yet cause detriment to individuals when processed by search engines because they can combine lawful information to generate completely new insight and

---

[49] See Lomas, Natasha, "*Google Offers Webform To Comply With Europe's 'Right To Be Forgotten' Ruling*," TechCrunch, May 30, 2014, accessed March 16, 2015. <http://techcrunch.com/2014/05/30/right-to-be-forgotten-webform/>

[50] See Mansoori, Sara and Eloise Le Santo, "*Over half a million Google URLs removal requests to date; the 'Right to be Forgotten' in practice*," International Forum for Responsible Media, November 14, 2014, accessed March 16, 2015. <https://inforrm.wordpress.com/2014/11/14/over-half-a-million-google-urls-removal-requests-to-date-the-right-to-be-forgotten-in-practice-sara-mansoori-and-eloise-le-santo/>

[51] See Vijayan, Jaikumar, "*EU May Ask Google to Extend 'Right to Be Forgotten' Beyond Europe*," November 26, 2014 accessed March 16, 2015. <http://www.eweek.com/security/eu-may-ask-google-to-extend-right-to-be-forgotten-beyond-europe.html - sthash.xmOOEcn6.dpuf>

[52] See Inter-American Commission on Human Rights, "*Principles in the Declaration of Principles of Freedom of Expression*," OAS, 2000, accessed March 16, 2015. <http://www.oas.org/en/iachr/mandate/Basics/principlesfreedom.asp>

[53] See Bertoni, Eduardo, "*The Right to Be Forgotten: An Insult to Latin American History*," Huffington Post Technology Blog, November 24, 2014, accessed March 16, 2015. <http://www.huffingtonpost.com/eduardo-bertoni/the-right-to-be-forgotten_b_5870664.html>

[54] *Id.*

provide access to outdated lawful information that would simply disappear or not be accessible easily otherwise.[55] Further, under the ruling, European citizens may seek removal of links from search results however, it does not lead to the removal of the content itself, which in many instances may be both legal and accurate. Targeting intermediaries such as search engines does not fully address concerns about third party content.[56]

I.d. Intermediaries must never be made strictly liable for hosting unlawful third-party content, nor should they ever be required to monitor content proactively as part of an intermediary liability regime.

Governments should refrain from incorporating into liability regimes requirements that go beyond forwarding any notice received on a retroactive basis. In particular governments should refrain from requiring intermediaries to proactively monitor and report content as such requirements negatively impact the right to free speech and the right to privacy.[57]

Imposing strict liability on intermediaries for illegal content is especially pernicious, because it requires them to proactively screen for such content and make a determination of its legality in order to avoid direct liability. Because such screening mechanisms are slow and expensive, the incentivize intermediaries simply to withdraw or to sharply limit access to their services.

On this point the 2011 Joint Declaration states that:

> *intermediaries should not be required to monitor user-generated content and should not be subject to extrajudicial content takedown rules which fail to provide sufficient protection for freedom of expression (which is the case with many of the 'notice and takedown' rules currently being applied).*[58]

The Joint Declaration states:

---

[55] See Ruiz, Javier "*Landmark ruling by European Court on Google and the 'Right to be Forgotten'*", Open Rights Group, May 15, 2015, accessed March 16, 2015.<https://www.openrightsgroup.org/blog/2014/landmark-ruling-by-european-court-on-google-and-the-right-to-be-forgotten>

[56] See Geist, Michael, "*Right to be forgotten ruling lacks balance: Geist*," TheStar.com, Tech News, May 16, 2014, accessed March 16, 2015. <http://www.thestar.com/business/tech_news/2014/05/16/right_to_be_forgotten_ruling_lacks_balance_geist.htm>

[57] See European Information Society Institute (EISi), "*Third Party Intervention Submission In re Delfi AS v. Estonia, App. no. 64569/09,"* June 15, 2014, p. 17, accessed March 16, 2015. <http://www.eisionline.org/images/EISi-Delfi-Intervention.pdf>

[58] *Id.*

At a minimum, intermediaries should not be required to monitor user-generated content and should not be subject to extrajudicial content takedown rules which fail to provide sufficient protection for freedom of expression (which is the case with many of the 'notice and takedown' rules currently being applied). [59]

Principle II: Content must not be required to be restricted without an order by a judicial authority

Laws should not require the intermediary to take action on a content restriction order or request without the consent of the person who put the content in question online, unless the party requesting the takedown is an independent and impartial judicial authority.

II.a. Intermediaries must not be required to restrict content unless an order has been issued by an independent and impartial judicial authority that has determined that the material at issue is unlawful.

Although we recommend that any restriction of content should be authorized by an impartial judiciary as the best qualified authority to determine validity or harm of information, we also recognize the need to balance this ideal against the need for expedited action in exceptional circumstances, and also that other legitimate interests that may be impacted by the administrative and financial burden that large quantities of content restriction requests may create.

This is because judicial review of content restriction requests does impose a significant burden upon the complainant, whose cause may also be legitimate. This burden cannot be dismissed even from a human rights standpoint. If each content restriction request was required to be reviewed individually by a judge, this would have one of two outcomes:

1. Drastically reducing the number of complaints that could be reviewed, thereby implicitly leaving a large number of potentially objectionable materials online.
2. Overburdening the intermediary, and/or the judicial process set in place to deal with such requests, to the extent that either the intermediary withdraws its services, or that judicial resources are over-allocated to content restriction requests.

[59] Joint Declaration on Freedom of Expression and the Internet, p.2.

Either of these outcomes could be suboptimal from a human rights standpoint. If no action at all were taken on the majority of content complaints, the rights and interests of complainants could suffer. Similarly, if intermediaries, under the burden of too many requests, limited or withdrew their services, users would clearly suffer. And if courts were overburdened, the result could be to limit the resources available to deal with other civil and criminal justice issues.

Thus as in many other areas, there is the need to find a balance, so that illegality can be reduced, but with safeguards to avoid causing or encouraging private censorship.[60] As expressed in the Rapporteurs' Joint Declaration:

> *On evaluating the proportionality of a restriction to freedom of expression on the Internet, one must weigh the impact that the restriction could have on the Internet's capacity to guarantee and promote freedom of expression against the benefits that the restriction would have in protecting other interests.*[61]

The answer to this dilemma that we propose is threefold:

1. The cost burden on the intermediary and/or on the justice system can be shifted to the party requesting content restriction.
2. Part of the burden can be shifted, in part, to the user of the intermediary's services, through a notice and notice system.
3. The burden of a full judicial hearing can be reduced by instituting an expedited judicial process, subject to due legal safeguards.

Taking the first two points together, principle III.d authorizes governments to require intermediaries to forward notices of claimed illegality to their users, in certain cases. There is nothing in the Manila Principles that prevents such a regime to provide for cost recovery by intermediaries for forwarding those notices. For example in New Zealand, ISPs are permitted to charge rightsholders $25 for each notice of claimed copyright infringement forwarded to users under that country's graduated response regime.[62]

---

[60] La Rue, Frank, *"Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression,"* United Nations, Human Rights Council, May 16, 2011, Paragraphs 42, 43 and 75, accessed March 16, 2015. <http://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf>

[61] Joint Declaration, point 1 b).

[62] Copyright (Infringing File Sharing) Regulations 2011, accessed March 16, 2015. <http://www.legislation.govt.nz/regulation/public/2011/0252/latest/DLM3886623.html>

Moving to the third point, Chile is an example of a country with an expedited judicial notice and takedown regime for copyright works.[63] In response to its Free Trade Agreement with the United States, the system introduced in 2010 is broadly similar to the DMCA, with the critical difference that intermediaries are not required to take material down in order to benefit from a liability safe harbor, until such time as a court order for removal of the material is made, under a special expedited legal process. Responsibility for evaluating the copyright claims made is therefore shifted from intermediaries onto the courts, which is a major difference.

Although this requirement does impose a burden on the rights holder, this serves a purpose by dis-incentivizing the issue of automated or otherwise unjustified notices that are more likely to restrict or chill freedom of expression. In cases where there is no serious dispute about the legality of the content, it is unlikely that the lawsuit would be defended. In any case, the Chilean legislation authorizes the court to issue a preliminary or interim injunction on an ex parte basis, on condition of payment of a bond where serious grounds exist.

II.b. Orders for the restriction of content must:

1. Provide a determination that the content is unlawful in the jurisdiction.
2. Indicate the Internet identifier and description of the unlawful content.
3. Provide evidence sufficient to document the legal basis of the order.
4. Where applicable, indicate the time period for which the content should be restricted.

A judicial order will trigger the intermediary's obligation to respond exclusively according the information and instruction provided by the judicial authority.

First, an order issued to an intermediary must therefore be clear about the jurisdiction in which the content is unlawful. However, due to the propensity of some courts to exercise long-arm jurisdiction over intermediaries that may operate outside of their physical jurisdiction, this provision of the Manila Principles does not suggest that every such claim of jurisdiction need be taken at face value.[64] Instead, we suggest that only if the intermediary offers the content from the same physical jurisdiction as the court that makes the order, are they required to comply.

See under III.b below for discussion of the second requirement, that an Internet identifier and description of the unlawful content be provided.

---

[63] Center for Democracy and Technology, "*Chile's Notice-and-Takedown System for Copyright Protection: An Alternative Approach,"* August 2012, accessed March 16, 2015. <https://www.cdt.org/files/pdfs/Chile-notice-takedown.pdf>

[64] See IV.c below

Third, evidence of the legal basis of the order will include the law allegedly being violated, and the legal basis for the authority of the court issuing that order to enforce that law.

Finally, as intermediaries should not have to assume that the order is to remain in effect for an unlimited duration, we also require that the order should explicitly specify this. For example, it may be an interlocutory order that will remain in place only until a final determination is made at trial.

A related issue not expressly treated in the text of this principle concerns how an intermediary should respond to a court order that is not directed at the intermediary directly, but rather a third party such as an individual user who posted an allegedly defamatory remark on the intermediary's platform. As a matter of principle, these should not generally be actioned by an intermediary because a legal obligation to remove content is not created by court orders directed to third parties. In practice however, some intermediaries also restrict content in response to such third party orders, which is usually explicable on the basis that a terms of service infringement has also occurred.

The rationale is that findings of liability of third parties may establish knowledge when communicated to the intermediary, and in some intermediary liability regimes, actual knowledge of illegality of content can expose the intermediary to its own primary liability in separate proceedings.[65] In some regimes, intermediaries can even be held liable despite the lack of notice about the illegality of content (because it is a "red flag" case), or knowledge can be imputed by other means than notice (such as for example, publication of decisions in major newspapers or other types of constructive awareness). Therefore, to avoid likely direct liability in the future, the intermediary may take action despite not being a direct party to the original court proceedings.

It is important to note that the Manila Principles do not sanction this practice because we recommend against the imposition of primary liability on intermediaries; however, we do recognize it as a reality in some jurisdictions. Therefore we suggest that third party court orders that otherwise comply with the criteria set out in II.b should only be accepted by an intermediary in lieu of a direct order against the intermediary itself where there has been a terms of service infringement, and where the intermediary is liable to attract direct liability if it fails

[65] For example, a German domain registrar was held liable for a torrent site's copyright infringement because it was "obvious" that the site was used for infringements: see Essers, Loek, "*German court finds domain registrar liable for torrent site's copyright infringement,*" PC World, February 7, 2014, accessed March 16, 2015. <http://www.pcworld.com/article/2095740/german-court-finds-domain-registrar-liable-for-torrent-sites-copyright-infringement.html>.

to act. Note, again, that this circumstance will never arise in a jurisdiction whose intermediary liability regime complies with the Manila Principles.

II.c. Any liability imposed on an intermediary must be proportionate and directly correlated to the intermediary's wrongful behavior in failing to appropriately comply with the content restriction order.

Governments should not impose disproportionate penalties on intermediaries. In the case that an intermediary fails to comply with a legal obligation, for example to pass on a notice for content restriction, this should be proportionate to the infraction committed by the intermediary (for example, a penalty for contempt of court, if applicable in the case of a court order), rather than being the same penalty that would apply to the author of the content.

In other words, where safe harbors for the protection of intermediaries from third party liability are provided in law, it should be understood that failure of an intermediary to abide by the conditions for the safe harbor will result only in loss of the safe harbor and not in an automatic finding of liability. In addition, no ISP or speaker should be liable or lose safe harbor protection for any act which would not attract liability if done offline.

In particular, intermediaries should not face criminal penalty for failing to comply with a content restriction order. Heavy or disproportionate penalties push intermediaries to remove content, even when it may be lawful, in order to avoid penalty. This overblocking has a negative impact on freedom of expression and disincentivizes intermediaries from challenging extra-legal content orders.

II.d. Intermediaries must not be liable for non-compliance with any order that does not comply with this principle.

Although this may seem an obvious corollary of the foregoing points, and perhaps it is, this point clarifies that non-compliance by an intermediary with a content restriction order should not in itself give rise to liability, where that order does not follow from a judicial finding that content is actually illegal.

The point is worth making explicitly because in some jurisdictions, such as South Korea, liability has been imposed upon intermediaries even in cases where no content has actually been found illegal.[66]

[66] Park, Kyung-Sin, "*Intermediary liability: Not Just Backward but Going Back*", 2014, accessed March 16, 2015. <http://opennetkorea.org/en/wp/main-free-speech/intermediary-liability-korea-2014>

Principle III. Requests for restrictions of content must be clear, be unambiguous, and follow due process

Consistent with Principle II, intermediaries should not be required to restrict content without an order from a judicial authority. However in the common event that governments or private complainants request content restriction in advance of a court order being issued, the following principles apply.

III.a. Intermediaries must not be required to substantively evaluate the legality of third-party content.

Any procedures for content restriction that require intermediaries to make determinations on content legality, without any oversight or accountability, or those which only respond to the interests of the party requesting removal, are unlikely to balance public and private interests. A better balance can be obtained through a mechanism where power is distributed between the parties involved, and where an impartial, independent, and accountable oversight mechanism exists.

The OAS report states on this topic:

> *save for in extraordinarily exceptional cases, this type of mechanism puts private intermediaries in the position of having to make decisions about the lawfulness or unlawfulness of the content, and for the reasons explained above, create incentives for private censorship. Indeed, extrajudicial notice and takedown mechanisms have frequently been cause for the removal of legitimate content, including specially protected content. ... Specifically, the requirement that intermediaries remove content, as a condition of exemption from liability for an unlawful expression, could be imposed only when ordered by a court or similar authority that operates with sufficient safeguards for independence, autonomy, and impartiality, and that has the capacity to evaluate the rights at stake and offer the necessary assurances to the user.* [67]

III.b. A content restriction request pertaining to unlawful content must, at a minimum, contain the following:

1. The legal basis for the assertion that the content is unlawful.
2. The Internet identifier and description of the allegedly unlawful content.
3. The consideration provided to limitations, exceptions, and defenses available to the user content provider.

[67] OAS report, pp.47-48.

4. Contact details of the issuing party or their agent, unless this is prohibited by law.
5. Evidence sufficient to document legal standing to issue the request.
6. A declaration of good faith that the information provided is accurate.

Private parties issuing content restriction requests on the basis of claimed illegality should also ensure that such requests are adequate and clear, so that the intermediary can adequately respond, including taking any action that may be required by law in order for it to be granted immunity from liability. (Under the Manila Principles, the only such action would be forwarding a notice of illegality to the user under III.d below, but more extensive obligations are imposed under many existing legal regimes, as described in the introduction.)

The specifications give here expand upon those in II.b above that court orders should contain. They require firstly that content restriction requests must include a detailed description of the specific content alleged to be illegal and to make specific reference to the law allegedly being violated, and the country where that law applies.

Secondly, content restriction requests should be required to specify the exact location of the material—such as a specific URL—in order to be valid. This is perhaps the most important requirement, in that it allows hosts to take targeted action against identified illegal material without having to engage in burdensome search or monitoring. An intermediary cannot be imputed with knowledge if information is missing, such as a valid URL.

Next, senders should be required to certify that they have considered in good faith whether any limitations, exceptions, or defenses apply to the material in question. This is particularly relevant for copyright and other areas of law in which exceptions are specifically described in law.

Requests should also be required to contain contact information for the sender. This facilitates assessment of notices' validity, feedback to senders regarding invalid notices, sanctions for abusive notices, and communication or legal action between the sending party and the poster of the material in question. We do allow that contact details may be the details of an agent; for example, to address cases where harassment of a legitimate complainant may occur if their direct contact details are given.

The requirement to document the complainant's standing to issue the request also helps to reduce the incidence of bogus notices. Notices should be issued only by or

on behalf of the party harmed by the content. For copyright, this would be the rights-holder or an agent acting on the rights-holder's behalf.

Finally a sender of a notice should be required to attest under legal penalty to a good-faith belief of the truth of the facts stated. This kind of formal certification requirement signals to request-senders that they should view misrepresentation or inaccuracies on notices as akin to making false or inaccurate statements to a court or administrative body. This helps to limit bad faith restriction requests, and can provide the basis for sanctions against those who send false notices (see Principle III.g).

These requirements expand upon those that CDT has recommended for clear notices in a notice and action system, in response a European Commission public comment period on a revised notice and action regime.[68]

III.c. Content restriction requests pertaining to an intermediary's content restriction policies must, at the minimum, contain the following:.

1. The reasons why the content at issue is in breach of the intermediary's content restriction policies.
2. The Internet identifier and description of the alleged violation of the content restriction policies.
3. Contact details of the issuing party or their agent, unless this is prohibited by law.
4. A declaration of good faith that the information provided is accurate.

In cases where a content restriction request is issued pursuant to an alleged infraction of the intermediary's own content restriction policies, the information to be provided is the relevant subset of those required in respect of allegedly illegal content in III.b above.

[68] *Id.*

III.d. Intermediaries who host content may be required by law to respond to content restriction requests pertaining to unlawful content by either forwarding lawful and compliant requests to the user content provider, or by notifying the complainant of the reason it is not possible to do so ('notice and notice'). Intermediaries should not be required to ensure they have the capacity to identify users.

This paragraph permits the law to institute a "notice and notice" regime, requiring intermediaries to pass on content restriction requests to the uploader of the information in question. Such a mechanism ensures that the intermediary is not placed in a quasi judicial position—making determinations regarding the legality or illegality of content. The Manila Principles do not prevent intermediaries from passing on notices voluntarily, even in the absence of a legal mandate, though neither do they require this.

Note that this paragraph explicitly only applies to content hosts, not to intermediaries such as ISPs who are mere conduits. Thus, the Manila Principles do not support a "graduated response" regime against those who merely *access* allegedly unlawful content.

As noted earlier, Canada is an example of a jurisdiction with a notice and notice regime, though limited to copyright content disputes. Although this regime is now established in legislation, it formalizes a previous voluntary regime, whereby major ISPs would forward copyright infringement notifications received from rights-holders to subscribers, but without removing any content and without releasing subscriber data to the rights-holders absent a court order. Under the new legislation additional record-keeping requirements are imposed on ISPs, but otherwise the essential features of the regime remain unchanged.

Analysis of data collected during this voluntary regime indicates that it has been effective in changing the behavior of allegedly infringing subscribers. A 2010 study by the Entertainment Software Association of Canada (ESAC) found that 71% of notice recipients did not infringe again, whereas a similar 2011 study by Canadian ISP Rogers found 68% only received one notice, and 89% received no more than two notices, with only 1 subscriber in 800,000 receiving numerous notices.[69] However, in cases where a subscriber has a strong good faith belief that the notice they received was wrong, there is no risk to them in disregarding the erroneous notice—a feature that does not apply to notice and takedown.

[69] Geist, Michael, "*Rogers Provides New Evidence on Effectiveness of Notice-and-Notice System"* March 23, 2011, accessed March 16, 2015. <http://www.michaelgeist.ca/2011/03/effectiveness-of-notice-and-notice/>

In the Canadian notice and notice system, some of the notices requesting content restriction that intermediaries have been required to send have contained misleading information.[70] We do not suggest that intermediaries should be required to vet the accuracy of all notices that they pass on, because that in itself would require them to exercise a level of legal judgment that could be both burdensome and inappropriate to entrust to them. However what can be required is that standard-form information can be included with all notices giving details of the rights of the recipient of the notice to contest or challenge the facts that it states, and the intermediary could at least ensure that this information is passed along.

Much of this information will be similar from one notice to another, but some details may differ depending on the grounds on which the content restriction is made. The challenge is how to achieve the desired level of clarity if the claimant, and perhaps the intermediary, are unaware of all the applicable legal rules. APC has suggested the use of standard forms raising the questions relevant for such determinations.[71] Similarly, Google has a form that guides the claimant through a series of questions, before ultimately recommending the appropriate legal form for their complaint to take.[72]

III.e. When forwarding the request, the intermediary must provide a clear and accessible explanation of the user content provider's rights, including in all cases where the intermediary is compelled by law to restrict the content a description of any available counter-notice or appeal mechanisms.

Although the Manila Principles do not support notice-and-action regimes, this principle does have particular application in such a regime, by requiring that a forwarded notice should include any available counter-notice procedures so that content providers can contest mistaken and abusive notices and have their content reinstated if the law has compelled its removal prior to a judicial order being made. Under such notice and action regimes, users should be entitled to raise defences and in the event of disputes, they should be referred to low cost arbitration, and the notice will include details of these procedures.

---

[70] Geist, Michael, "*Rightscorp and BMG Exploiting Copyright Notice-and-Notice System: Citing False Legal Information in Payment Demands*" January 8, 2015, accessed March 16, 2015. <http://www.michaelgeist.ca/2015/01/rightscorp-bmg-exploiting-copyright-notice-notice-system-citing-false-legal-information-payment-demands/>

[71] APC report, p. 28.

[72] See Google, *"Removing Content From Google,"* accessed March 16, 2015. <https://support.google.com/legal/troubleshooter/1114905?hl=en>

III.f. If intermediaries restrict content hosted by them on the basis of a content restriction request, they must comply with Principle VI on transparency and accountability below.

The intermediary is at liberty to ignore content restriction requests that it receives based on its own policies, but in many cases intermediaries do solicit reports of infringing content from their users, as a way of crowd-sourced content moderation. Provided that the intermediary's policies are clear and transparently applied (see VI.c and VI.e), the Manila Principles do permit this (see also V.f). For example, a social media platform intended for users who are children might have a policy against hosting images that depict sexual activity or violence.

Note however that the legitimate scope of content restriction available by conduits such as ISPs is much more limited than that available to content hosts, because ISPs are required under net neutrality principles to treat all Internet content equally, whereas content hosts have much more discretion about what content they do and do not choose to host.

III.g. Abusive or bad faith content restriction requests should be penalized.

In order to deter abuse, senders of erroneous or abusive notices should face possible sanctions. For example a sender could be held liable for damages or attorneys' fees for making improper misrepresentations (or for repeatedly making improper misrepresentations).[73] In the United States for example, senders may face penalties for misrepresentations of infringement.[74] Such penalties may not be sufficient, however, as takedown abuse continues to occur.

Principle IV. Laws and content restriction orders and practices must comply with the tests of necessity and proportionality

IV.a. Any restriction of content should be limited to the specific content at issue.

Judicial orders determining the unlawfulness of specific content and mandating its restriction should be clear, specific, and limited to avoid over-removal of content. To this end, courts should only order the removal of the bare minimum of content that is necessary to remedy the harm identified and nothing more.

As CDT asserts in its 2012 intermediary liability report:

---

[73] *Id.*

[74] 17 U.S.C. § 512(f)

> *Actions required of intermediaries must be narrowly tailored and proportionate, to protect the fundamental rights of Internet users. Any actions that a safe-harbor regime requires intermediaries to take must be evaluated in terms of the principle of proportionality and their impact on Internet users' fundamental rights, including rights to freedom of expression, access to information, and protection of personal data. Laws that encourage intermediaries to take down or block certain content have the potential to impair online expression or access to information. Such laws must therefore ensure that the actions they call for are proportional to a legitimate aim, no more restrictive than is required for achievement of the aim, and effective for achieving the aim. In particular, intermediary action requirements should be narrowly drawn, targeting specific unlawful content rather than entire websites or other Internet resources that may support both lawful and unlawful uses.*[75]

Whilst this recommendation addresses courts, it also can apply to intermediaries. This is because, depending on jurisdiction and location of where the content order is originating from, intermediaries may retain discretion about how to respond to that order.

IV.b. When restricting content, the least restrictive technical means must be adopted.

Intermediaries implementing court orders, private third party requests, or enforcing their terms of service should adopt the least restrictive means of doing so. This determination should take into consideration the proportionality of the harm caused/to be caused by the content, the nature of the content, the class of intermediary, the impact on affected users, and the proximity to the content uploader.

There are a number of different ways that access to content can be restricted. Examples applicable to content hosts include:

- hard deletion of the content from all of a company's servers,
- blocking the download of an app or other software program in a particular country,
- blocking the content on all IP addresses affiliated with a particular country ("IP blocking"),

[75] Center for Democracy and Technology, "*Shielding the Messengers: Protecting Platforms for Expression and Innovation"* p. 12. Supra.

- removing the content from a particular domain of a product (eg, removing from a link from the .fr version of a search engine that remains accessible from the .com version),
- blocking content from a 'version' of an online product that is accessible through a 'country' or 'language' setting on that product, or
- some combination of the last three options (i.e., an online product that directs the user to a version of the product based on the country that their IP address is coming from, but where the user can alter a URL or manipulate a drop-down menu to show her a different 'country version' of the product, providing access to content that may otherwise be inaccessible).

Examples applicable to conduits include:

- filtering based on full URL, destination DNS or IP address, and
- content filtering based on type of traffic, content of traffic (eg. keywords revealed by deep packet inspection).

While almost all of the different types of content restrictions described above can be circumvented by technical means such as the use of proxies, IP-cloaking, or Tor, the average Internet user does not know that these techniques exist, much less how to use them. Of the different types of content restrictions described above, a domain removal, for example, is easier for an individual user to circumvent than IP blocked content because you only have to change the URL of the product you are using to, i.e. ".com" to see content that has been locally restricted. To get around an IP block, you would have to be sufficiently savvy to employ a proxy or cloak your true IP address.

Therefore, the technical means used to restrict access to controversial content has a direct impact on the magnitude of the actual restriction on speech as well as the extent to which an individual's privacy is infringed upon. The more restrictive the technical removal method, the fewer people that will have access to that content. To preserve access to lawful content, online intermediaries should choose the least restrictive means of complying with removal requests, especially when the removal request is based on the law of a particular country that makes certain content unlawful that is not unlawful in other countries. Further, when building new products and services, intermediaries should build in removal capability that minimally restricts access to controversial content.

The 2011 *Joint Declaration on Freedom of Expression and the Internet* issued by the four rapporteurs on freedom of expression made the following points about the dangers of allowing filtering technology:

Mandatory blocking of entire websites, IP addresses, ports, network protocols or types of uses (such as social networking) is an extreme measure—analogous to banning a newspaper or broadcaster—which can only be justified in accordance with international standards, for example where necessary to protect children against sexual abuse.

Content filtering systems which are imposed by a government or commercial service provider and which are not end-user controlled are a form of prior censorship and are not justifiable as a restriction on freedom of expression. There has been a problem of over-removal, such as child safety filters that also remove safe sex information.

Products designed to facilitate end-user filtering should be required to be accompanied by clear information to end-users about how they work and their potential pitfalls in terms of over-inclusive filtering.[76]

Similarly, the Council of Europe has suggested a number of safeguards in relation to the use of filters by intermediaries, recommending that states should:

- introduce regulations where necessary to prevent the intentional use of filters to restrict access to lawful content
- assess filters both before and during their implementation to ensure their effects are appropriate and proportional and avoid unreasonable blocking of content
- Provide for effective means of recourse including suspension of filters where users claim lawful content or access is being blocked.[77]

In short, filtering at the conduit level is a blunt instrument that should be avoided whenever possible. Similarly to how conduits should not be legally responsible for content that they neither host nor modify (the "mere conduit" rule discussed *supra*), mere conduits are not able to assess the context surrounding the controversial content that they are asked to remove and are therefore not the appropriate party to receive takedown requests. Therefore, governments should not require conduits to build in the capability to filter content.[78]

---

[76] Joint Declaration *on Freedom of Expression and the Internet*, pp. 2-3. Supra.

[77] Council of Europe, "*Recommendation on freedom of expression and information with regard to Internet filters*," March 26, 2008, accessed March 16, 2015. <https://wcd.coe.int/ViewDoc.jsp?id=1266285&Site=CM>

[78] See "*International Principles on the Application of Human Rights to Communications Surveillance*", Supra. "'States should not compel service providers or hardware or software vendors to build surveillance or monitoring capability into their systems".

On the part of governments, a key element of due process lies within the legal system itself. An independent and impartial judiciary exists, at least in part, to preserve the citizen's due process rights. Many have called for an increased reliance on courts to make determinations about the legality of content posted online in order to both shift the censorship function from unaccountable private actors and to ensure that courts only order the removal of content that is actually unlawful. However, when courts do not have an adequate technical understanding of how content is created and shared on the internet, the rights of the intermediaries that facilitate the posting of the content, and who should be ordered to remove unlawful content, they can damage the online ecosystem. Therefore, we recommend that courts seek expertise about the technical feasibility of restriction measures to ensure that the least restrictive technical means be adopted.

IV.c. If content is restricted because it is unlawful in a particular geographical region, and if the intermediary offers a geographically variegated service, then the geographical scope of the content restriction must be so limited.

Intermediaries should restrict content in the most limited way possible in order to preserve freedom of expression. This includes, when possible, intermediaries applying geographical filters to content based on the jurisdiction in which the content is allegedly unlawful.

A user should be able to access content that is lawful in her country even if it is unlawful in another country. Different countries have different laws and it is often difficult for intermediaries to determine how to effectively respond to requests and reconcile the inherent conflicts that results from jurisdictional differences. For example, content that denies the holocaust is illegal in certain countries, but not in others. If an intermediary receives a request to remove content based on the laws of a particular country and determines that it will comply because the content is not lawful in that country, it should not restrict access to the content such that it cannot be accessed by users in other countries where the content is lawful.

To respond to a request based on the law of a particular country by blocking access to that content for users around the world, or even users of more than one country, essentially allows for extraterritorial application of the laws of the country that the request came from. A current example of this is in the case of *Equustek Solutions v. Morgan Jack*, in which a Canadian trial judge ruled that Google must remove links to full websites that contained pages selling a product that allegedly infringed trade secret rights, not only from its Canadian search pages, but around the world. (Google appealed, and EFF has intervened in that appeal, which remains pending.)

While it is preferable to standardize and limit the legal requirements imposed on online intermediaries throughout the world, to the extent that this is not possible, the next-best option is to limit the application of laws that are interpreted to declare certain content unlawful to the users that live in that country. Therefore, intermediaries should choose the technical means of content restriction that is most narrowly tailored to limit the geographical scope and impact of the removal.

IV.d. If content is restricted owing to its unlawfulness for a limited duration, the restriction must not last beyond this duration, and the restriction order must be reviewed periodically to ensure it remains valid.

Similarly, the temporal scope of restriction should be as limited as necessary to comply with the law. See also Principle II.b above.

Principle V. Laws and content restriction policies and practices must respect due process

This is the second principle (after principle III) referencing due process. Due process is a legal entitlement of democratic citizenship, but intermediaries too are called upon to adopt processes that afford users a right to be heard before their content is restricted, as it is by adopting such processes that illegitimate restrictions can be minimized.

The adherence of intermediaries to due process norms is especially important to close the loophole whereby authorities may apply pressure upon intermediaries to restrict content voluntarily, outside of the rule of law. By ensuring that intermediaries' policies also contain analogous protections against arbitrary content restriction, this danger of extra-legal restriction is curtailed.

V.a. Before any content is restricted on the basis of an order or a request, the intermediary and the user content provider must be provided an effective right to be heard except in exceptional circumstances, in which case a *post facto* review of the order and its implementation must take place as soon as practicable.

Courts should ensure that any proceeding deliberating on a content restriction is done in the presence of the author or the person who uploaded the content, providing him or her the right to be heard. Recognizing that there are exceptional circumstances that may require the government or law enforcement to restrict

content as soon as possible, without the time or ability to locate an author for a proceeding—such circumstances should be permitted, but an ex post facto review must take place as soon as possible.

This is an important part of due process, and is particularly important to protect against the abuse of ex-parte injunctions.

We do not attempt to list the sorts of exceptional circumstances that may justify deviation from the normal role of a judicial hearing prior to content restriction. However, as an illustration only, the UN Special Rapporteur set out four categories of content that must be prohibited under international law and that States are required to prohibit domestically. These include: (1) child pornography, (2) direct and public incitement to commit genocide, (3) advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence, (4) and incitement to terrorism. [79]

The Special Rapporteur recognizes that access to these four categories of content may be restricted, and in case of child pornography and incitement to commit genocide, underscores that the use of blocking and filtering technologies must be sufficiently precise, and that there must be adequate and effective safeguards against abuse or misuse including oversight and review. However, he stresses that while the four types of expression constitute offences under international criminal law and/or international human rights law and which States are required to prohibit at the domestic level, they all also constitute restrictions to the right to freedom of expression. He further reiterates, that all content restriction and blocking practices and policies must comply with the three-part test of prescription by: unambiguous law; pursuance of a legitimate purpose; and respect for the principles of necessity and proportionality.

A similar (but broader) test of "manifest illegality" has been applied in several jurisdictions (eg. France), [80]and similar language ("manifestly ill-founded") has been interpreted by the European Court of Human Rights (ECHR). [81]However it has been noted by Article 19 that the concept of manifest illegality is too broad and vague in relation to the types of content and thus introduces a level of ambiguity that many

[79] La Rue, Frank, 2011, pp.8-13. Supra.

[80] APC report, p.13; See also Bits of Freedom, *"A clean and open Internet: Public consultation on procedures for notifying and acting on illegal content hosted by online intermediaries,"* Response to the EU notice and action consultation, accessed March 16, 2015. <https://www.bof.nl/live/wp-content/uploads/040912-submissiontoformofconsultationeuropeancommission.pdf>

[81] See *"European Convention on Human Rights,"* Article 35(3) under (2) of the Admissibility Criteria, accessed March 16, 2015. <http://www.echr.coe.int/Documents/Convention_ENG.pdf>

intermediaries (particularly small intermediaries) will be less qualified to judge and act on without an authoritative determination.

Beyond content that can be categorized as manifestly illegal, content owners have argued that content that infringes copyright should be removed without judicial authorization from intermediaries' networks, on the grounds that judicial content restriction orders do not scale to the level required to address the allegedly rampant infringement of copyright works on intermediaries' networks and platforms.

One of the problems with this argument is that intellectual property infringements are rarely so legally unambiguous as the exceptional cases set out by the Special Rapporteur. As Bits of Freedom noted in its submission to the European Commission public consultation on procedures for notifying and acting on illegal content hosted by online intermediaries:

> *A one-approach-fits-all will not work. As indicated under (5), content that is unmistakably lawful and depicting or describing criminal behavior, should be dealt with differently from content that is unlawful because it infringes a trademark, copyright or other rights of intellectual property. Such infringements must of course be terminated, but the unlawfulness will be harder to assess.*

Another reason to differentiate intellectual property infringements from the exceptional cases set out by the Special Rapporteur is that the former normally only impact on the economic interests of rightsholders, whereas the latter can impact on more fundamental rights such as the right to personal integrity and the right to life.

V.b. Any law regulating intermediaries must provide both user content providers and intermediaries the right of appeal against content restriction orders.

Both user content providers and intermediaries should have access to remedy when the application of the intermediary liability regime results in a decision that affects them negatively. From governments, this most importantly involves ensuring that mechanisms of appeal exist when content is wrongly restricted—or when it is wrongly not restricted.

For example, the lack of judicial review was the constitutional flaw in the original HADOPI legislation in France, which sought to address the related issue of intellectual property enforcement against Internet end-users.[82]

Government should also ensure that they do not interfere with the ability for intermediaries to remediate the wrongful restriction of content when a content reinstatement request is upheld. This is particularly relevant in the case of data erasure requests under laws that recognize what has become popularly known as a "right to be forgotten", as explained in the discussion of principle I.c above. [83]

Whilst the provision of access to remedy by intermediaries has to be subordinated to other laws, including data protection laws, such laws should not require the intermediary to permanently erase content while the removal request remains subject to review. In order to ensure that this is not the case, there is an urgent need to harmonize between data protection and intermediary liability laws.

V.c. Intermediaries should provide user content providers with mechanisms to review decisions to restrict content in violation of the intermediary's content restriction policies.

Content providers cannot rely on the court system to resolve disputes over content that has been restricted based on a violation of terms of service, because many content restriction requests never reach the judicial system before the intermediary takes action on them under terms of service. It is incumbent on the intermediary in such cases to provide and to communicate a clear mechanism for review and appeal of the content restriction decision. (In cases where the legal system does provide a further or ultimate mechanism of recourse, this principle does not of course preclude or supersede it.)

As noted in the UNESCO Report:

> *Remedy is the third central pillar of the UN Guiding Principles on Business and Human Rights, placing an obligation on governments and companies to provide individuals access to effective remedy. This area is where both governments and companies have much room for improvement. Across intermediary types, across jurisdictions and across the types of restriction, individuals whose content or publishing access is restricted as well as individuals who wish to*

---

[82] Lovejoy, Nathan *"Procedural Concerns with the HADOPI Graduated Response Model*," Harvard Journal of Law and Technology, JOLT Digest, January 13, 2011, accessed March 16, 2015. <http://jolt.law.harvard.edu/digest/copyright/procedural-concerns-with-the-hadopi-graduated-response-model>

[83] *Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González.* Supra.

> *access such content had inconsistent, limited, or no effective recourse to appeal restriction decisions, whether in response to government orders, third party requests or in accordance with company policy. While some companies have recently increased efforts to provide appeal and grievance mechanisms and communicate their existence to users, researchers identified examples of rules being inconsistently enforced and enforced in a manner also not consistent with the principles of due process.*[84]

Similarly the WIPO report states:

> *Particular attention must be paid to some kind of independent scrutiny of accusations of alleged copyright infringement before any sanctions are imposed, as well as access to review afterwards, as the Internet Freedom clause demands. Users should have access to redress for economic, reputational and privacy harms caused by false or negligent allegations in a way that effectively discourages such.*[85]

Access to remedy may require more than merely a remedy for wrongful content removal, but also for any associated privacy violations, defamation, etc. Sometimes this may require immediate removal and only a possibility of subsequent reinstatement, yet other times reinstatement would be less significant and the ability to receive notice and to be heard are more important.

V.d. In case a user content provider wins an appeal under (b) or review under (c) against the restriction of content, intermediaries should reinstate the content.

Whenever an intermediary restricts content, there should be a clear mechanism through which users can request reinstatement of content. When an intermediary decides to remove content, it should be immediately clear to the user that content has been removed and why it was removed. If the user disagrees with the content removal decision, there should be a clear and accessible online method for the reinstatement of content to be requested.

It follows that reinstatement of content should also be technically possible. When intermediaries (who are subject to intermediary liability) are building new products, they should build the capability to remove content into the product with a high degree of specificity so as to allow for narrowly tailored content removals when a removal is legally required. Relatedly, all online intermediaries should build

[84] UNESCO report, p. 86.

[85] WIPO report, p.72.

the capability to reinstate content into their products, to the extent that they can legally do so while maintaining compliance with other applicable laws.

Intermediaries should also have policies and procedures in place to handle reinstatement requests. Between the front end (online mechanism to request reinstatement of content) and the back end (technical ability to reinstate content) is the necessary middle layer, which consists of the intermediary's internal policies and processes that allow for valid reinstatement requests to be assessed and acted upon. In line with the corporate "responsibility to respect" human rights, and considered along with the human rights principle of "access to remedy," intermediaries should have a system in place from the time that an online product launches to ensure that reinstatement requests can be made and will be processed quickly and appropriately. Any notice and takedown system is subject to abuse, and any company policy that results in the removal of content is subject to mistaken or inaccurate takedowns, both of which are substantial problems that can only be remedied by the ability for users to let the intermediary know when the intermediary improperly removed a specific piece of content and the technical and procedural ability of the intermediary to put the content back.

Indeed, intermediaries should endeavor to ensure that their processes for dealing with content restriction requests and content orders are fair. In this regard self regulatory frameworks can guide best practices for intermediaries in relation to removal requests and in particular this will require them to give clear notice to users of such orders and requests, and to provide them with access to remedy in cases where content is wrongly restricted.

V.e. An intermediary should not disclose personally identifiable information about a user without an order by a judicial authority. An intermediary liability regime must not require an intermediary to disclose any personally identifiable user information without an order by a judicial authority.

Intermediary liability, the freedom of expression, and privacy are issues that intersect in a number of ways. When governments impose restrictions on an individual's ability to express themselves anonymously, the intermediary must implement such a policy. This can be seen in South Korea's real ID policy that was implemented from 2007 - 2012.[86] One reason that governments are quick to place liability on intermediaries is a result of the access that intermediaries have. Not only

[86] Caragliano, David A., "*Real Names and Responsible Speech: The cases of South Korea, China and Facebook*," The Right to Information & Transparency in the Digital Age, Stanford University, March 11-12, 2013, accessed March 16, 2015. <https://www.ndi.org/files/Caragliano_Stanford_Paper_Apr_5_2013.pdf>

do intermediaries have access to content on their platforms and networks, but they also have access to user data including IP address, user names, and log history. Thus, removal requests by law enforcement are often coupled with user data requests. Similarly, liability regimes can include requirements that the intermediary monitor and report on a proactive basis specified activities or types of content. Such requirements infringe on the rights to freedom of expression and the privacy of users.

Although the Manila Principles do not seek to exhaustively address privacy issues, they do provide that governments should not legally require intermediaries to disclose personally identifiable information as part of an intermediary liability regime without a judicial order. To this extent, Governments should not hold an intermediary liable for failing to disclose personal data of users without such an order. Given the impact of revealing personally identifiable information on the privacy and freedom of users and the potential for misuse,[87] policies determining disclosure requirements of intermediaries must ensure safeguards for protection of users.

Such a standard is critical in protecting the privacy of users, a right affirmed under the resolution adopted at the United Nations General Assembly that calls upon States to review procedures, practices and legislation around communication surveillance including the collection of personal data.[88] This does not detract from the fact that in some cases, the disclosure of user information by intermediaries will be necessary to uphold the rights of victims.[89] But such cases should be judicially assessed.

EFF has suggested additional best practices for a third party wishing to take action against an individual for posting allegedly defamatory or otherwise illegal content:[90]

- Make reasonable efforts to notify the person whose identity is sought;

---

[87] See Noble, Graham, "*YouTube Copyright Hoax Used By Terrorists to Gain Personal Information*," Liberty Voice, November 6, 2014, accessed March 16, 2015. <http://guardianlv.com/2014/11/youtube-copyright-hoax-used-by-terrorists-to-gain-personal-information/#G5HvIArrkCVpUZwb.99>

[88] See General Assembly, United Nations, *"The right to privacy in the digital age,"* Resolution adopted on December 18, 2013, accessed March 16, 2015. <http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/68/167>

[89] European Court of Human Rights, *"K.U. v. Finland,"* December 2, 2008, accessed March 16, 2015. <http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-89964 - {"itemid":%5B"001-89964"%5D}>

[90] EFF, "*Freedom of Expression, Privacy and Anonymity on the Internet*", January 2011, accessed March 16, 2015. <https://www.eff.org/files/filenode/unspecialrapporteurfoe2011-final_3.pdf>

- If possible, agree to a timetable for disclosure of the information to the party seeking it that provides a reasonable opportunity for the Internet user to file an objection with a court before disclosure;
- Forward the exact statements and material provided by the person seeking the identity, including information about the cause of action alleged in the lawsuit and the evidence provided by the identity-seeker to the court where provided to the service provider.
- Users should be provided with a reasonable amount of time to respond before the service provider produces the requested information. This will give the user an opportunity to object to disclosure of his or her identity.

V.f. When drafting and enforcing their content restriction policies, intermediaries should respect human rights. Likewise, governments have an obligation to ensure that intermediaries' content restriction policies respect human rights.

Intermediaries should observe due process and the application of their policies must not give rise to human rights infringements. As the OAS report puts it, "Companies must seek to ensure that any restriction derived from the application of the terms of service does not unlawfully or disproportionately restrict the right to freedom of expression." [91]

The Internet has space for a wide range of platforms and applications directed to different communities, with different needs and desires. A social networking site directed at children, for example, may reasonably want to have policies that are much more restrictive than a political discussion board. The webmaster of a music review website may wish to restrict comments to those about music, and restrict those about any other topic.

Within the scope of the law and observing human rights standards, intermediaries retain control over their own policies as long as they are transparent about what those policies are, what type of content the intermediary removes, and why they removed certain pieces of content.

This distinguishes the case of a private intermediary with a public authority, which can only limit public freedom of expression towards achieving urgent objectives such as national security or protecting the rights of others.[92]

Having said that, it also remains that intermediaries are responsible for creating important public fora for deliberation and discussion, including on political and

[91] OAS report, p. 51.

[92] OAS report, p. 27.

social issues. Some, such as Facebook, are so ubiquitous that the policies that they adopt can have a significant effect on the range of permissible interaction within entire online communities. This places greater responsibility on such intermediaries to ensure that their policies respect human rights standards.

Principle VI. Transparency and accountability must be built into laws and content restriction policies and practices

Transparency and accountability are integral facets of democratic government. But the Manila Principles extends these standards to intermediaries also, in view of their role in facilitating the speech of their users. Although the Manila Principles do not prohibit intermediaries who are content hosts from restricting content for terms of service violations, this is on the basis that those terms of service are transparent, and that the intermediary is accountable for their implementation.

The UNESCO report states:

> *Transparency of laws, policies, practices, decisions, rationales, and outcomes related to privacy and restrictions on freedom of expression allow users to make informed choices about their own actions and speech online. Transparency is therefore important to internet users' ability to exercise their rights to privacy and freedom of expression.*[93]

VI.a. Governments must publish all legislation, policy, decisions and other forms of regulation relevant to intermediary liability online in a timely fashion and in accessible formats.

Governments should ensure that the statutory intermediary liability regime is explained in plain language, perhaps on the same website as its transparency report described at VI.d below. This should include any co-regulatory arrangements reached by governments and industry that are not directly the subject of legislation. Examples include the Internet Watch Foundation (IWF) in the UK,[94] safernet in Brazil,[95] the new UK "adult content" filtering scheme,[96] Project Sunblock,[97] etc.

VI.b. Governments must not use extra-judicial measures to restrict content. This includes collateral pressures to force changes in terms of

[93] UNESCO report, p. 86.

[94] See Internet Watch Foundation, UK, accessed March 16, 2015. <https://www.iwf.org.uk/>

[95] See SaferNet, accessed March 16, 2015. <http://www.safernet.org.br/>

[96] See House of Lords UK, *"Notes on the Online Safety Bill as introduced in the House of Lords on 10th June 2014,"* accessed March 16, 2015. <http://www.publications.parliament.uk/pa/bills/lbill/2014-2015/0016/en/15016en.htm>

[97] See Project Sunblock, accessed March 16, 2015. <http://www.projectsunblock.com/>

service, to promote or enforce so-called "voluntary" practices and to secure agreements in restraint of trade or in restraint of public dissemination of content.

Outside of legal regimes of intermediary liability, there is also a grey zone within which intermediaries are "encouraged" by regulators or third parties to take "voluntary" action to police content on their networks. This is apparent, for instance, in Article 16 of the ECD which encourages the development of codes of conduct by intermediaries to deal with third-party content.[98] Also late in the negotiation of the NETmundial Multi-stakeholder Statement in April 2014, the following language advocated for by rights-holder representatives was inserted into the final text:

> *Intermediary liability limitations should be implemented in a way that respects and promotes economic growth, innovation, creativity and free flow of information. In this regard, cooperation among all stakeholders should be encouraged to address and deter illegal activity, consistent with fair process.*[99]

In the October 2014 leaked text of the Trans-Pacific Partnership, a proposal which follows other free trade agreements in force that also aim at such cooperation, we see a similar text proposal. The proposal although requiring a legal obligation of intermediaries, does not directly link this with safe harbor protection (square bracketed text removed):

> *Each Party shall provide legal incentives for online service providers to cooperate with copyright owners or {help} / {take action} to deter the unauthorized storage and transmission of copyrighted materials.*[100]

Initiatives of this type have motivated some jurisdictions to implement graduated response schemes or "three strikes systems" aiming to reduce copyright infringement online, by requiring an intermediary (generally an ISP, in this case) to send notifications to their customers warning them they are alleged to have infringed copyright law. Some such schemes can encourage ISPs to take technical measures such as displaying an intrusive pop-up warning to the user that an infringement notice has been sent, reducing the user's bandwidth, blocking

---

[98] See European Parliament and the Council of the European Union, *European E-Commerce Directive 2000/31 (ECD),* article 16.

[99] See NETmundial, *"NETmundial Multistakeholder Statement,"* April 24, 2014, accessed March 16, 2015. <http://netmundial.br/wp-content/uploads/2014/04/NETmundial-Multistakeholder-Document.pdf>

[100] See Wikileaks, *"Updated Secret Trans-Pacific Partnership Agreement (TPP) - IP Chapter (second publication),"* October, 16, 2014, accessed March 16, 2015. <https://www.wikileaks.org/tpp-ip2/>

protocols or, in the worst scenario, temporarily suspending the user's account for alleged repeated infringement.[101]

Even CDA 230,[102] through its so-called "Good Samaritan" provision, opens the door to pressure for extra-legal content takedown, by exempting intermediaries from liability for any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.

Such "soft" obligations and incentives pose a further risk of chilling speech through intermediaries that is less easily quantified than a hard obligation on intermediaries to take down content following a well-defined, legally sanctioned process. Such practices, which are neither accountable nor transparent, are proscribed here.

VI.c. Intermediaries should publish their content restriction policies online, in clear language and accessible formats, and keep them updated as they evolve, and notify users of changes when applicable.

Governments should ensure that intermediaries maintain the ability to adapt their terms of service to what they feel is appropriate and needed for the services offered. In turn, intermediaries should ensure that their terms of service are clear and transparent and provide users avenues for remedy.

Intermediaries should have clear policies that are published online and kept up-to-date to provide its users notice of what content is and is not permitted on the company's platform. The UNESCO report states, by way of background, that:

> *While all social networks list content they prohibit, none of the companies studied has provided much public information about procedures for evaluating content. Industry sources have described internal rules and procedures for evaluating content in conversations with concerned stakeholders, held on condition of non-attribution, but such processes are generally not made public. It is usually through anecdotal evidence via news reports that the public learns about specific examples.*[103]

---

[101] See Malcolm, Jeremy, *"Australia's Proposed Copyright Alert System Allows Rightsholders to Spy on Users,"* Electronic Frontier Foundation, February 2015, accessed March 16, 2015. <https://www.eff.org/deeplinks/2015/02/australias-proposed-copyright-alert-system-allows-rightsholders-spy-users>

[102] See Section 230 of Title 47 of the United States Code (47 USC § 230).

[103] UNESCO report, p. 162.

Notice to the user about the types of content that are permitted encourages her to speak freely and helps her to understand why content that she posted was taken down if it must be taken down for violating a company policy. This should also include any self-regulatory arrangements on which a number of intermediaries have reached between themselves, or with other industry segments such as copyright owners, payment intermediaries and advertisers, to the extent that these impact the content permissible on the intermediary's platform.

There are legitimate reasons why an ISP may want to have policies that permit less content, and a narrower range of content, than is technically permitted under the law, such as maintaining a product that appeals to families. But since these policies are poorly documented by intermediaries at present, there are community initiatives that have been established to gather evidence of how intermediaries, specifically major social media platforms, are applying their own content policies. These include onlinecensorship.org and the Ranking Digital Rights project.[104]

VI.d. Governments must publish transparency reports that provide specific information about all content orders and requests issued by them to intermediaries

As part of the democratic process, the citizens of each country (as well as non-citizen residents) have a right to know how their government is applying its laws, and a right to provide feedback about the government's legal interpretations of its laws. Thus, all governments should be required to publish online transparency reports that provide specified information about content orders issued by government to intermediaries.

As such, the UN Special Rapporteur for freedom of expression has called upon States that currently block websites to provide lists of blocked websites and full details regarding the necessity and justification for blocking each individual website.[105] An explanation should also be provided on the affected websites as to why they have been blocked. Any determination on what content should be blocked must be undertaken by a competent judicial authority or a body which is independent of any political, commercial, or other unwarranted influences.

For example, this information should include aggregate numbers of orders issued, reasons for content restriction, the legal nature of the orders (that is, executive or judicial), the government branch requesting the restriction, and the numbers of cases where content was reinstated. Where possible, the aggregate data that

[104] See "Ranking Digital Rights", accessed March 16, 2015. <https://rankingdigitalrights.org/>.

[105] La Rue, Frank, 2011, Supra. Paragraph 70.

constitutes each government's transparency report should be made available online, for free, in a common file format such as .csv, so that civil society may have easy access to it for research purposes. Of course, personally identifiable information (PII) should be removed from any such data sets.

There may be merit in publishing this information centrally across all of government, providing a holistic view of the burden imposed on intermediaries, encouraging dialogue between different branches of government about how best to create and enforce internet content regulation, and between the government and its citizens about the laws and policies applicable to internet content.

Governments should also allow for intermediaries to publish transparency reports detailing all content orders requests from government agencies and courts in a periodic transparency report, accessible on the intermediary's website, that publishes information about the requests the intermediary received and what the intermediary did with them in the highest level of detail that is legally possible (see VI.e below).

Further, citizens should have the right to request copies of content orders from the government and access to information legislation could provide a legal basis for access to such information. For example in India under the Right to Information legislation citizens can seek information from the government on content restriction orders.[106]

VI.e. Intermediaries should publish transparency reports that provide specific information about all content restrictions taken by the intermediary, including actions taken on government requests, court orders, private complainant requests, and enforcement of content restriction policies

Similar transparency obligations are expected of intermediaries. Regrettably, this is an area that intermediary liability regimes overlook. The UNESCO report states:

> *The practice and scope of company and government transparency about surveillance practices, filtering and service restrictions vary across jurisdictions. In none of the countries studied are ISPs legally required to be*

[106] See Pahwa, Nikhil, "*Our Right To Information Request On India's Order To Block 245 Web Pages*," Medianama, August 21, 2012, accessed March 16, 2015. <http://www.medianama.com/2012/08/223-right-to-information-request-on-indias-order-to-block-245-web-pages/>

> *transparent about their policy or practice regarding filtering, service restrictions, or surveillance measures.*[107]

The scope of information to be disclosed is broad, though steps should be taken to prevent the disclosure of personal information in the publication of transparency reports, and it will not be necessary to provide details of the individual (and likely automated) blocking of malicious content such as spam and phishing material.

Subject to this, the more information that is provided about each request, the better is the understanding that the public will have about how laws that affect their rights online are being applied. Therefore, intermediaries should strive to publish the maximum amount of information about each request that they can, subject as well to the (ideally minimal) restrictions imposed by applicable law, and the economic scale. Where scale is a barrier, representative sampling may be an alternative.

Related to this, the obligation to issue transparency reports must be relative to the scale on which the intermediary operates. The public interest in transparency reporting is much higher for large intermediaries with a large number of users. For small intermediaries, such as message board and public WiFi operators, proactive transparency reporting is not expected. Community initiatives such as Chilling Effects help by providing a free central hosting repository to which intermediaries can submit content restriction requests that they have received.[108]

CDT states:

> *Disclosure by service providers of notices received and actions taken can provide an important check against abuse. In addition to providing valuable data for assessing the value and effectiveness of a N&A [notice and action] system, creating the expectation that notices will be disclosed may help deter fraudulent or otherwise unjustified notices. In contrast, without transparency, Internet users may remain unaware that content they have posted or searched for has been removed pursuant due to a notice of alleged illegality. Requiring notices to be submitted to a central publication site would provide the most benefit, enabling patterns of poor quality or abusive notices to be readily exposed.*[109]

---

[107] UNESCO report, p. 86.

[108] See *"Chilling Effects database,"* accessed March 16, 2015. <https://www.chillingeffects.org/>

[109] Center for Democracy and Technology, "*Additional Responses Regarding Notice and Action*," accessed March 16, 2015. <https://www.cdt.org/files/file/CDT N&A supplement.pdf>

A thorough transparency report published by an intermediary should include information about the following categories of requests:

- **Government requests**

  This category includes all requests to the intermediary from government agencies; from police departments, to intelligence agencies, to school boards from small towns. Surfacing information about all restriction requests from any part of the government helps to avoid corruption and/or inappropriate exercises of governmental power by reminding all government officials, regardless of their rank or seniority, that information about the requests they submit to online intermediaries is subject to public scrutiny.

  Vodafone's country by country report of government orders and demands are an example. They even have a policy on privacy, human rights and law enforcement assistance. However, they do not publish any reports on content that they are taking down pursuant to their terms of service.[110]

- **Court orders**

  This category includes all orders issued by courts and signed by a judicial officer. It can include ex-parte orders, default judgments, court orders directed at an online intermediary, or court orders directed at a third party presented to the intermediary as evidence in support of a removal request. To the extent legally possible, detailed information should be published about these court orders detailing the type of court order each request was, its constituent elements, and the actions(s) that the intermediary took in response to it. In most cases court orders are published openly as a requirement of access to justice, but there may be cases where personally identifying information should be redacted from any court orders that are published by the intermediary as part of a transparency report before publication.

  Information about court orders should be further broken down into two groups; orders against the intermediary, and orders against the party who posted the disputed content. The first category is the simplest; where court orders are directed at the online intermediary in an adversarial proceeding

[110] See Vodafone, "Country-by-country disclosure of law enforcement assistance demands," accessed March 16, 2015. <http://www.vodafone.com/content/sustainabilityreport/2014/index/operating_responsibly/privacy_and_security/law_enforcement/country_by_country.html>

to which the online intermediary was a party, either as the primary defendant or as a third party respondent.

As noted above at II.b, it will generally not be consistent for the Manila Principles for an intermediary to act upon a court order that is not directed to it specifically. Nonetheless if the user who obtains a court order approaches an online intermediary seeking removal of content with a court order directed at the poster of, say, defamatory content, and the intermediary decides to remove the content in response to the request, the online intermediary that decided to perform the takedown should publish a record of that removal.

This type of court order should be broken out separately from court orders directed at the applicable online intermediary in companies' transparency reports because merely providing aggregate numbers that do not distinguish between the two types gives an inaccurate impression to users that more takedown requests are being directed at intermediaries than is actually the case. When the court made its determination of legality on the content in question, it may not have contemplated that the intermediary would remove the content. If so, the court likely did not weigh the relevant public interest and policy factors that would include the importance of freedom of expression or the precedential value of its decision.

Instead, and especially considering that these third party court order may be the basis for a number of content removals, third party court orders should be counted separately and presented with some published explanation in the company's transparency report as to what they are and why the company has decided it should removed content pursuant to its receipt of one. The intermediary should also identify in the report the legal grounds for removal (in terms of legislation violated or more generally, area of law).

- **Private party requests**

  Private party requests are requests to remove content that are not issued by a government agency or accompanied by a court order. Some examples of private party requests include copyright complaints submitted pursuant to the Digital Millennium Copyright Act or complaints based on the laws of specific countries, such as laws banning holocaust denial in Germany, and which authorize or require the ISP to act in the absence of a court order. Note that the Manila Principles does not sanction the restriction of content in response to such a request, but we acknowledge the reality that this does represent that law in many jurisdictions.

- **Policy/TOS enforcement**

  To give users a complete picture of the content that is being removed from the platforms that they use, corporate transparency reports should also provide information about the content that the intermediary removes pursuant to its own policies or terms of service, though there may not be a legal requirement to do so. All past versions of policies and any changes made to them must be included as part of the intermediaries' transparency efforts.

VI.f. Where content has been restricted on a product or service of the intermediary that allows it to display a notice when an attempt to access that content is made, the intermediary must display a clear notice that explains what content has been restricted and the reason for doing so.

If content is removed or access to it is restricted for any reason, either pursuant to a legal request or because of a violation of company's terms of service, in general a user should be able to learn that the content was restricted if they try to access it.

Requiring an on-screen message that explains that content has been restricted and why, is the post-takedown complement to the pre-takedown published online policy of the online intermediary: both work together to show the user what types of content are and are not permitted on each online platform. Explaining to users why content has been restricted in sufficient detail may also spark their curiosity as to the laws or policies that caused the content to be restricted, resulting in increased civic engagement in the Internet law and policy space, and a community of citizens that demands that the companies and governments it interacts with are more responsive to how it thinks content regulation should work in the online context.

It must be acknowledged that for conduits, as opposed to content hosts, it may not always be technically feasible to provide a notice of content which is unavailable due to filtering. However at least for website that have been filtered, it is technically simple for intermediaries to redirect the attempted access to a page which explains why the website is unavailable. There is even a proposal for web standard that would provide a standard browser error code for this purpose.[111]

Some limited exceptions to the duty to notify users of restricted content may apply, mainly for the protection of personally identifiable information. In particular, when

[111] See Internet Engineering Task Force (IETF), "*An HTTP Status Code to Report Legal Obstacles*," December 16, 2014, accessed March 16, 2015. <http://www.tbray.org/tmp/draft-ietf-tbray-http-legally-restricted-status-05.html>

personally identifiable information is removed in compliance with data protection law, to notify users of the former presence of that content could spark a "Streisand effect", whereby the restriction of access actually draws more attention to the content than when there was no restriction. This could actually obviate the purpose of the removal. It is for this reason that the European Privacy Commissioner advised Google not to notify the public of particular search results that it had removed under European data protection law on the grounds that they contained "inadequate, irrelevant or no longer relevant" information about individuals.[112] Instead, Google places a notice on every page that appears to be a search result for a personal name search, simply saying that results *may* have been removed. Whilst this is better than users receiving no notice at all, the utility of such a blanket notification is dubious.

VI.g. Governments, intermediaries and civil society should work together to develop and maintain independent, transparent, and impartial oversight mechanisms to ensure the accountability of the content restriction policies and practices.

Governments should support independent, transparent, and impartial accountability mechanisms to verify the practices of government and companies with regards to managing content created online. The UNESCO report states:

> *It is important that companies and governments alike make commitments to implement core principles of freedom of expression and privacy. In today's globally networked digital environment, these principles must be implemented in a manner that is accountable locally as well as globally.*
>
> *Examples from the consumer privacy context include: the European Union's Binding Corporate Rules and the APEC Cross Border Privacy Rules system. Another approach to accountability for companies is through assessment and certification by independent multi-stakeholder organizations. The Global Network Initiative, a multistakeholder coalition, requires its members to undergo periodic assessments as part of an accountability mechanism for adherence to its principles and implementation guidelines focused on how companies handle government requests.*[113]

Often self regulation takes place under the "shadow of the state"; that is all sides act under the threat that the State may intervene if no compromise is found or public

---

[112] See Smith, David, *"Response to the European Google judgment,"* Information Commissioner's Office Blog, August 7, 2014, accessed March 16, 2015. <http://iconewsblog.wordpress.com/2014/08/07/update-on-our-response-to-the-european-google-judgment/>

[113] UNESCO report, p.192.

interests are seriously threatened[114] which may lead to invisible censorship with implications for human rights as these measures often do not have independent oversight mechanisms for ensuring accountability.[115] However, self regulation also takes place where there is no regulation and when done effectively and in collaboration with governments, provides the opportunity to adapt rapidly to technical progress. Relying on just one set of actors or a single approach to address content concerns may not work and restriction policies and practices, must aim at incorporating a systemic approach to self regulation by governments, industry and rights holders.[116] It is key though that any approach incorporate independent, transparent and impartial oversight mechanisms to ensure the accountability of content restriction policy and practice.

Civil society also has a role to play in encouraging comparative studies between countries and between intermediaries with regards to their content removal practices, to identify best practices. Civil society has the unique ability to look longitudinally across this issue to determine and compare how different intermediaries and governments are responding to content removal requests. Without information about how other governments and intermediaries are handling these issues, it will be difficult for each government or intermediary to learn how to improve its laws or policies. Therefore, civil society has the ability to help create increasingly better human rights outcomes for online platforms by performing and sharing ongoing, comparative research.

Civil society can also work to ensure that all relevant stakeholders have a voice in both the creation and revision of policies that affect online intermediaries. In the context of corporate policy making, civil society can use strategies from activist investing to encourage investors to make the human rights and freedom of expression policies of Internet companies' part of the calculus that investors use to decide where to place their money.

Apart from the onlinecensorship.org and Ranking Digital Rights projects already mentioned above, there is also a recently-formed Dynamic Coalition on Platform Responsibility, which emphasizes the concept of "platform responsibility" to

---

[114] See Oxford University, Centre for Socio-Legal Studies, Programme in Comparative Law and Policy (PCMLP), "*Self-Regulation of Digital Media Converging on the Internet: Industry Codes of Conduct in Sectoral Analysis*," 2004, p. 37, accessed March 16, 2015. <http://pcmlp.socleg.ox.ac.uk/sites/pcmlp.socleg.ox.ac.uk/files/IAPCODEfinal.pdf>

[115] See Prakash, Pranesh, "*Invisible Censorship: How the Government Censors Without Being Seen*," Centre for Internet and Society, December 14, 2011, accessed March 16, 2015. <http://cis-india.org/internet-governance/invisible-censorship>

[116] See Bertelsmann Foundation, "*Self-regulation of Internet Content*," 1999, accessed March 16, 2015. <https://www.cdt.org/files/speech/BertelsmannProposal.pdf>

stimulate behavior in line with the principles laid out by the UN Guiding Principles on Business and Human Rights, endorsed by the UN Human Rights Council. The Guiding Principles recognize the complementary, yet different, roles of States and companies in relation to the validity of human rights, focusing on the responsibility of private corporations to respect human rights and to grant an effective grievance mechanism. The coalition's website states:

> *The ability of users to recognize and reward this type of behaviour has the potential to generate a virtuous circle, whereby consumer demand drives the market towards human rights-compliant solutions. Accordingly, the utilisation of model contractual-provisions may prove instrumental to foster trust in online services for content production, use and dissemination, allowing platform-users to directly identify those platforms that ensure the respect of their rights in a responsible manner.*[117]

On the intermediaries side, there is also the Global Network Initiative (GNI), which has the expressed purpose of protecting and advancing freedom of expression and privacy in information and communication technologies, and seeks to hold members accountable to human-rights based standards through independent assessment.[118] The UNESCO report points out:

> *Members of the Global Network Initiative, specifically commit to "respect and protect the freedom of expression of their users" in the course of responding to government requests to remove content or hand over user data. They also commit to be held accountable to this commitment. There are two components of public accountability for GNI members: "independent assessment and evaluation" of whether the companies are upholding their commitment to the GNI principles, and also "transparency with the public." Two years after the GNI's official launch with three company members (Google, Microsoft, and Yahoo), the practice of what has come to be called "transparency reporting" began to emerge.*[119]

VI.h. Intermediary liability frameworks and legislation should require regular, systematic review of rules and guidelines to ensure that they are up to date, effective, and not overly burdensome. Such periodic review should incorporate mechanisms for collection of evidence about

[117] See Dynamic Coalition on Platform Responsibility, accessed March 16, 2015. <http://platformresponsibility.info/>

[118] See Global Network Initiative, accessed March 16, 2015. <http://www.globalnetworkinitiative.org>

[119] UNESCO report, p.123.

their implementation and impact, and also make provision for an independent review of their costs, demonstrable benefits and impact on human rights.

This principle, calling for the review of intermediary liability policies following their introduction, addresses unforeseen costs and consequences of the introduction of new intermediary liability rules. The recent experience of Canada, mentioned above, whereby rights-holders have been sending misleading notices of infringement under the notice and notice regime, provides a good example.[120] The review of such laws will ensure that such unforeseen impacts are redressed in a timely fashion.

The likelihood of the failure of intermediary liability rules can be minimized when governments ensure that all private citizens are given the right and equal opportunity to provide feedback on the balancing between their human rights and other public interests that arise in developing intermediary liability public policy. Denying Internet users a voice in the policymaking processes that determine their rights undermines government credibility and negatively influences users' ability to freely share information online. As such, it is good practice for governments to consult, online and face-to-face, on proposed laws that affect intermediaries giving users the opportunity to provide input.

As simply expressed in the NETmundial Multistakeholder Statement, "Anyone affected by an Internet governance process should be able to participate in that process."[121] This requires governments to:

1. Provide citizens with a mechanism for submitting feedback to any legislative process regarding content restriction.
2. This mechanism must be accessible (ie in the appropriate language, through an easy to use interface, widely publicized).
3. It is the responsibility of the government to effectively demonstrate that feedback submitted by citizens is equitably considered and deliberated upon.

An example of something like this in practice was the online process by which Brazil's Marco Civil was collaboratively developed, in an interactive process that incorporated feedback from stakeholders before the law was finalized.[122]

---

[120] Geist, Michael. 2015. Supra.

[121] *NETmundial Multistakeholder Statement.* Supra.

[122] See iobservatório da internet.br, "*The Internet Policy Report*, Brazil 2011," Section 2.1, p. 20-23, accessed March 16, 2015. <http://observatoriodainternet.br/wp-content/uploads/2012/11/Internet-Policy-Report-Brazil-2011.pdf>

Further, both companies and governments should embed an "outreach to at-risk communities" step into both legislative and policymaking processes to be especially sure that their voices are heard.

A relevant guiding principle for companies belonging to the Global Network Initiative states "While infringement on freedom of expression and privacy are not new concerns, the violation of these rights in the context of the growing use of ICT is new, global, complex and constantly evolving. For this reason, shared learning, public policy engagement and other multi-stakeholder collaboration will advance these Principles and the enjoyment of these rights."