Content Regulation in the Digital Age

Submission to the consultation launched by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression - OBSERVACOM[1]

Introduction

More than two decades after the opening up of the world wide web for the general public and the beginning of its commercial services, the Internet is no longer has the freshness and air of novelty that marked its first fifteen years of wide use. The initial profusion of services that competed to provide better and more innovative solutions to internet users, has been replaced by the stabilization and consolidation of a handful of platforms and applications that dominate a significant part of the traffic of information and content in the network.

Evidence shows a trend towards greater concentration in the hands of a few transnational corporations as a result of the dynamics of the current Internet business model. This accumulation of power is not only a result of the success of services and goods provided to users, but also the characteristics of a "network economy": the global scale of the business, the ability to raise capital for the necessary investments, and the mergers or purchase of other competing or complementary companies, among others. The dispute over the radio spectrum and the Internet of Things (IoT), and especially the ability to monetize the resulting *big data*, lead to processes that are deepening the current level of concentration.

Concern over concentration in the area of OTT services is justified, and beyond aspects of economic competition, given that several of the business corporations that have significant market power and a dominant position on the Internet are owners of platforms that enable the free flow of information and other relevant content such as social networks, search engines, communication applications and video sharing platforms. In this concentrated environment, the potential risks to access, diversity and pluralism of ideas and information that have already been mentioned become exacerbated.

[1] The Latin American Observatory of Regulation, Media and Convergence (OBSERVACOM) is a non-profit, professional and independent regional think tank, based in Uruguay, with regional operation, composed of experts and communication researchers committed to the protection and promotion of democracy, cultural diversity, human rights and freedom of expression. OBSERVACOM addresses public policies and regulations on audiovisual communication services, the Internet and other information and communication services in a digital and convergent environment, focusing on aspects related to access, diversity and pluralism.

Since the first half of the 20th century most advanced democracies have embraced the notion that regulation in the communications sector acts as a fundamental guarantee of democracy due to the central role that a pluralistic and diverse public sphere has for its proper functioning. The quality of democracy and a vigorous civic debate depend largely on the variety of information and views competing in the public space and available to citizens.

In a scenario centralized by the traditional media, it was clear that the market on its own did not guarantee the fundamental diversity, pluralism and freedom of expression needed by democracy. With the emergence of the Internet, it seemed that part of the rationality that gave meaning and foundation to democratic regulation might have been lost. In fact, some important players in the digital ecosystem claim that regulation of the Internet is not only dangerous but should not exist, as it is no longer necessary nor possible.

However, after the initial phase of more decentralized and open network operation, new bottlenecks formed and the Internet embarked on a growing centralization among just a few actors of the digital ecosystem that has affected its potential to serve all of humanity: this was underlined by the creator of the World Wide Web, Tim Berners Lee. The trend towards concentration and threats to freedom of expression on the Internet show that diversity and pluralism - and even the notion of an open and free Internet - need regulatory guarantees so that they can be maintained as values and paradigms of modern digital communications.

This may not lead, however, to weakening the role of intermediaries. Without intermediaries, it would be humanly impossible to enjoy the enormous potential available in the network of networks. Companies that provide platforms and applications on the Internet play a key role in terms of access to an open and free Internet, given the task they perform as intermediaries between users and the content available on the network.

However, and paradoxically, this new and vital role makes them a potential risk for freedom of expression and the free flow of information on the Internet. Such intermediaries no longer provide just technical support and "transit routes", but often affect the content that circulates through such routes. Not only are they able to monitor all content produced by third parties, but they can also intervene in them, ordering and prioritizing their access and, therefore, determining what content and sources of information a user may or may not view. They can also block, eliminate or de-index content – such as speeches protected by the right to freedom of expression - as well as users' accounts or profiles. These actions can

be imposed by external pressures from government authorities or other private actors, but also by the decisions taken by the intermediaries themselves.

Moreover, the growing incidence of intermediaries as a gateway to the information content available on the Internet has also generated a change in the flow of resources within the digital economy that would seem to be indirectly affecting diversity and pluralism, while negatively impacting economic benefits that producers of traditional information content receive, especially those that have high fixed costs, such as those that carry out investigative journalism and hard news.

Based on these facts, OBSERVACOM presents its contribution to the debate on content regulation in the digital age.

A) General Principles

A.1 – Big intermediaries must be subject to public obligations

The real possibilities of access, diversity and pluralism in a free and open Internet are concentrated among a few intermediaries or private corporations, whose platforms and services -for example, social networks- occupy the role of new public spaces. All this takes place, however, in the absence of accountability. To what extent is it possible to impose public obligations on private stakeholders?

A central argument for such obligations is that human rights must have horizontal effectiveness. Member States have to respect and promote human rights in their vertical relations with citizens. But also companies, in their horizontal relationships with users, are obliged to respect such rights. The United Nations Human Rights Council (UNHCR) expressly acknowledged this obligation in 2011 by approving the "Guiding Principles on Business and Human Rights".

This perspective is also justified when it is confirmed that the main platforms have a significant market power, and offer services that can be considered essential. Their market share and impact on essential services such as searches, social networks and audiovisual platforms endows them with an undeniable public dimension and requires, at the very least, regulations that guarantee diversity and pluralism.

A.2 – Intermediaries should not be liable for third-party content when no intervention has occurred

States should promote and protect the exercise of freedom of expression by adopting legislation, policies and administrative practices that provide an adequate regulatory environment for OTT service providers, in order to deal with threats and unlawful pressures of content removal, filtering or blocking by State authorities and other private actors.

Regulation should incorporate the notion that "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')" as set down in the Joint Declaration on Freedom of Expression and Internet 2011.

This does not mean that intermediaries do not have "any responsibility" for the exchange of content through their platforms. Unlike conduits they are not mere technical services and do undertake interventions when prioritizing or amplifying certain contents of third parties, for example, without being pressured by the State to do so.

A.3 - The platforms neutrality should also be a basic principle of the Internet

Inter-American standards include the principle of net neutrality as an indispensable condition for freedom of expression on the Internet. The objective is, as mentioned above, to ensure that "freedom of access and choice of users to use, send, receive or offer any content, application or legal service through the Internet is not conditioned, directed or restricted by means of blocking, filtration, or interference".

The same principle should be extended to other intermediaries - that is to say not just ISPs – and with the same purpose of ensuring diversity, pluralism and access to a free and open Internet. This is important because many of these platforms - and the algorithms they use - are increasingly responsible for fundamental decisions about the content that people access.

The level of potential or real interference with Internet content places a huge responsibility on intermediaries who -and if no democratic regulation is in place- in fact become a form of

private regulators never witnessed before. This situation is aggravated by the weakness of democratic states to regulate phenomena that transcend their administrative boundaries.

The concept of "neutrality" also holds true for these actors of the digital ecosystem, as OTT service corporations have the potential to affect freedom of expression "by conditioning, directing or restricting" content "through blocking, filtering, or interfering" if they do not act in a neutral way with respect to the information and opinions that circulate through their platforms and applications.

That this ability to be a *gatekeeper* lies in the control of a physical or virtual layer of access, should not affect the principle that gave rise to the notion of net neutrality and placed it as a key issue in the agenda for freedom of expression of the Internet. In fact, there is no indication of systematic and widespread evidence of a violation of freedom of expression based on political or ideological reasons on the part of ISPs to identify a serious problem for this fundamental right, and to conclude that it was a basic principle which should be regulated through the adoption of national laws.

A.4 - The content intervention by intermediaries is only justifiable in flagrantly discriminatory or illegal cases and following strict criteria.

The treatment of content internally by intermediaries is justifiable in flagrantly discriminatory or illegal cases, because of the volume of content exchanged and the speed and dynamics of exchange, which may, on the one hand, create indelible impact and, on the other hand, make it difficult for conflicts between parties to be treated only under the judicial power of each country. However, this content treatment is only be valid if it:

- Is based on public criteria defined by co-regulation (State and private sector, with multisector participation), aligned with international standards expressed in United Nations and regional instruments;
- Is limited to what is necessary to deal with cases that involve high risk due to their volume and speed of dissemination;
- does not substitute the public mediation of conflicts established by the due legal process of each country;
- has transparent and accountable rules and criteria for handling content, both for general rules and for individual cases;
- involves mechanisms of dispute and grievance, defined by public instruments.

B) Responses to the specific issues raised by the SR on the call for submissions

1. Company compliance with State laws:

a. What processes have companies developed to deal with content regulation laws and measures imposed by governments, particularly those concerning:

i. Terrorism - related and extremist content ;

ii. False news, disinformation and propaganda ; and /or

iii. The "right to be forgotten" framework?

[No specific contribution – see General Principles]

b. How should companies respond to State content regulation laws and measures that may be inconsistent with international human rights standards?

Companies should make use of the tools available to all natural or legal persons. They should appeal to the legal system in the country and, if ineffective, present the cases to the appropriate international bodies, such as the OAS Human Rights Commission and Court, in the case of the American continent.

2. Other State Requests: Do companies handle State requests for content removals under their terms of service differently from those made by non - State actors? Do companies receive any kind of content - related requests from States other than those based on law or the company's terms of service (for ex ample, requests for collaboration with counter speech measures)?

[No specific contribution – see General Principles]

3. Global removals: How do / should companies deal with demands in one jurisdiction to take down content so that it is inaccessible in other jurisdictions (e.g., globally)?

The regulatory challenges posed by OTT services include the difficulty of applying regulatory measures - and the questioning of the role of national governments – given that their activities take place in one or more countries, that they maintain global operations outside

the locations where services are provided or consumed, and that their reliance on international transactions. These difficulties cannot provide justification for OTT service providers to operate outside of the legal national or supranational framework that each State decides to adopt.

This means that companies must follow national rules and must act to guarantee the maximum right of access of the public combined with the right of protection to creators and artists. It is not acceptable, therefore, to impose national rules from one country to another.

4. Individuals at risk: Do company standards adequately reflect the interests of users who face particular risks on the basis of religious, racial, ethnic, national, gender, sexual orientation or other forms discrimination?

Discrimination is one of the most sensitive issues because it primarily affects marginalized and oppressed groups. As already said in the general principles, the content intervention by intermediaries is only justifiable in flagrantly discriminatory or illegal cases and following strict criteria. The treatment of content internally by intermediaries is justifiable in flagrantly discriminatory or illegal cases, because of the volume of content exchanged and the speed and dynamics of exchange, which may, on the one hand, create indelible impact and, on the other hand, make it difficult for conflicts between parties to be treated only under the judicial power of each country. However, this content treatment must only be valid if:

- based on public criteria defined by co-regulation (State and private sector, with multisector participation), aligned with international standards expressed in United Nations and regional instruments;
- limited to what is necessary to deal with cases that involve high risk due to their volume and speed of dissemination;
- do not substitute the public mediation of conflicts established by the due legal process of each country;
- have transparent and accountable rules and criteria for handling content, both for general rules and for individual cases;
- involves mechanisms of dispute and grievance, defined by public instruments.

5. Content regulation processes: What processes are employed by companies in their implementation of content restrictions and takedowns, or suspension of accounts? In particular, what processes are employed to:

a. Moderate content before it is published;

No “moderation” should be allowed before publication, as it implies a form of private prior censorship, forbidden by international standards. Moderation is censorship. In the American Convention on Human Rights, prior censorship is only acknowledged as legal for “the sole purpose of regulating access to [public entertainment] for the moral protection of childhood and adolescence”.

b. Assess content that has been published for restriction or take down after it has been flagged for moderation; and/or

Contribution shown together to item c – see below.

c. Actively assess what content on their platforms should be subject to removal?

The level of potential or real interference with Internet content places a huge responsibility on intermediaries who. If no democratic regulation is in place, become a form of private regulators never witnessed before. This situation is aggravated by the weakness of democratic states to regulate phenomena that transcend their administrative boundaries.

The concept of "neutrality" also holds true for these actors of the digital ecosystem, as OTT service corporations have the potential to affect freedom of expression "by conditioning, directing or restricting" content "through blocking, filtering, or interfering" if they do not act in a neutral way with respect to the information and opinions that circulate through their platforms and applications.

Today, intermediaries do not comply with Manila Principles on Intermediary Liability, as they are neither clear nor transparent on their content regulation policy and practices.

6. Bias and non-discrimination: How do companies take into account cultural particularities, social norms, artistic value, and other relevant interests when evaluating compliance with terms of service? Is there variation across jurisdictions? What safeguards have companies adopted to prevent or redress the takedown of permissible content?

Today cultural particularities, social norms and artistic values are not taken into account by intermediaries, who seek to establish global rules and apply them indistinctly. This clashes with the UNESCO Convention for the Protection and Promotion of the Diversity of Cultural Expressions. There are also no safeguards to reconsider the removal of allowed content.

Here three concerns must be taken into account:
- Cultural peculiarities of each country should be observed and differentiated treatment of the content by jurisdiction should be permitted, always in line with international standards of freedom of expression;
- There should be greater protection for the freedom of expression of artistic expressions, in line with the criteria historically adopted in consolidated democracies;
- There should be mechanisms of dispute and grievance in the cases of takedown of permissible content.

7. Appeals and remedies: How should companies enable users to appeal mistaken or inappropriate restrictions, takedowns or account suspensions? What grievance mechanisms or remedies do companies provide?

As stated in the general principles (section A, above), content treatment must involve mechanisms of dispute and grievance, defined by legal instruments. These mechanisms must enable appeals and counter-arguments, even in cases in which the removal is done without a third-party request. The mechanisms must reflect the principles of 'due process'.

8. Automation and content moderation: What role does automation or algorithmic filtering play in regulating content? How should technology as well as human and other resources be employed to standardize content regulation on platforms?

Algorithms are responsible for key decisions about the contents we can access, facilitating or hindering access to the content on the Internet. Algorithm design and the use of forms of artificial intelligence that select the contents that we can view in terms of preferences with the aim of leaving a person "satisfied" and "comfortable" can have good intentions and be a successful commercial strategy to attract users, but are not necessarily compatible with diversity and pluralism, fundamental requirements for the proper functioning of a democratic society.

The algorithms of social networks and search engines are based on criteria that are not transparent and that affect diversity and pluralism, relegating and concealing certain information or opinions, as well as generating a segmentation of public debate. In addition, some platforms, such as those that offer video on demand services, use algorithms that can inhibit the display of national content, affecting cultural diversity.

As algorithms are programmed – or at least can be modified – by company decision, it is important to consider them as a form of content intervention.

(see General Principles – A3 about platform neutrality)

***9. Transparency**:*

a. Are users notified about content restrictions, takedowns, and account suspensions? Are they notified of the reasons for such action? Are they notified about the procedure they must follow to seek reversal of such action?

b. What information should companies disclose about how content regulation standards under their terms of service are interpreted and enforced? Is the transparency reporting they currently conduct sufficient?

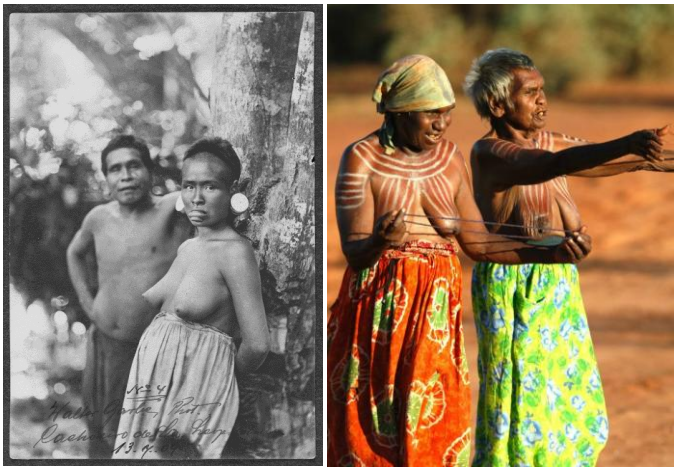
The removal of content considered "inappropriate" or "offensive" -in the opinion of the companies themselves and their "moderators"- is carried out with a lack of transparency and due process in terms of the decisions taken or right of appeal. The main companies in the sector do not even publicly report how much content they have decided to withdraw. All of which distances them from international standards on legitimate restrictions on freedom of expression, including the Manila Principles on Intermediary Liability.

10. Examples: Please share any examples of content regulation that raise freedom of expression concerns (e.g., account suspension or deactivation, post or video takedown, etc.), including as much detail as possible.

In recent years there have been cases in which the removal of content published by users was considered an illegitimate intervention and a violation of freedom of expression.

Perhaps the most visible case worldwide was Facebook's decision to censor the famous photo of a naked Vietnamese girl escaping an attack with napalm, published by the Norwegian writer Tom Engeland. In this case it even involved the temporary blocking of your profile in the social network. The censorship generated a wave of solidarity with Engeland and criticism of the company, from the Norwegian prime minister to the country's leading newspaper, Aftenposten, which published a cover with the so-called 'Dear Mark', in which it criticized the case[2].

Previously, in April 2015, Facebook removed a historical photograph from 1909 published on the page of the Ministry of Culture of Brazil in which an indigenous couple appeared. The woman appeared with bare breasts. The company alleged that the photograph did not respect its policy on decency and nudity[3]. The case was brought by the Brazilian government before the Inter-American Commission on Human Rights (IACHR) in a hearing held in April 2016. The Brazilian government alleged that the censorship violated the cultural identity of the indigenous, prohibited from appearing on Facebook in their own expression cultural. In Australia, a photo with two aborigines with naked breasts, part of a campaign to valorize feminism, was also censored[4].



Censored pictures in Brazil and Australia

[2] https://www.aftenposten.no/meninger/kommentar/Dear-Mark-I-am-writing-this-to-inform-you-that-I-shall-not-comply-with-your-requirement-to-remove-this-picture-604156b.html
[3] https://oglobo.globo.com/sociedade/midia/ministerio-da-cultura-vai-entrar-na-justica-contra-facebook-por-foto-de-india-bloqueada-1-15910229?utm_source=Facebook&utm_medium=Social&utm_campaign=O%20Globo
[4] http://www.dailylife.com.au/news-and-views/dl-opinion/facebooks-ban-of-aboriginal-activist-celeste-liddle-reveals-its-censorship-double-standards-20160314-gniycj.html

In both cases, Facebook first challenged the re-publication, and some days later it allowed the images, not as recognition of legitimacy but as a response to the strong public controversy.

For its part, Google was accused of censoring LGBT content in its restricted mode on YouTube, the platform for sharing videos produced by its users. The system is made to allow parental control, but was accused of going beyond preventing children's access to content with sexual content[5]. In addition to criticism from several bloggers, Human Rights Watch denounced how the restricted mode blocked a video about anti-LGBT censorship that discussed a Utah law[6].

This type of moral "overprotection" - generally included in the parameters established in their algorithms - also has effects in cases of domestic images published by users at times of breastfeeding, in which platforms accumulate cases of private censorship because the photos expose the mothers' breasts[7].

Most of the issues have to do with nudity, reflecting a cultural sensitivity especially from the United States, but there were also cases where there are complaints of censorship motivated by political or ideological reasons, as shown in the Online censorship project developed by Electronic Frontier Foundation and Visualizing Impact.

The site monitors the censorship cases of intermediaries such as Facebook, Twitter, YouTube and Instagram and produces weekly and annual reports with the cases received. In its November 2016 report, the project highlights the following points:

• During 2016, 294 reports of content removals in the intermediaries analyzed were identified. Of those reports, 74% belong to Facebook. 47% of all removals are related to content related to nudity[8].

• Many users (44.7%) reported trying to appeal the removal of their content on all platforms. Of these, at least 28% did not receive any response;

• There was an increase in complaints of politically motivated censorship, a large part of the US elections. Although these removals have occurred against supporters of the two main

[5] http://www.cromo.com.uy/youtube-acusado-homofobia-su-modo-restringido-n1048782; https://heatst.com/tech/youtube-censors-everyone-feminists-lgbt-vloggers-pundits-and-gamers/
[6] https://www.hrw.org/news/2017/03/22/youtube-censors-hrw-video-lgbt-censorship
[7] http://www.tvshow.com.uy/farandula/griselda-siciliani-censurada-instagram-mostrar.html
[8] https://onlinecensorship.org/news-and-analysis/onlinecensorship-org-launches-second-report-censorship-in-context-pdf

political parties, most of the reports related to the critical content of Hillary Clinton, set by users who identified themselves as supporters of Bernie Sanders or supporters of Donald Trump[9].

Referring to Facebook, Online Censorship emphasizes that "currently, there is no way to assess the effects of the decisions of content moderators at scale, as takedowns linked to violations of the content guidelines are not included in its transparency report. This would be a useful first step, particularly if accompanied by information about how many content takedowns are due to posts that were reported by other users"[10].

Fake News

In April, Google announced plans to modify its search service to make it difficult for its users to access what it called "low quality" information, a concept that includes notes related to "conspiracy theories" and "false news". Social movements and organizations of the left have denounced a significant drop in the global positioning of traffic on their web pages since then.

World Socialist Web Site (a portal published by the International Committee for the Fourth Socialist International) has reported a drop of more than 70% in its visits. They denounce that Google uses an ambiguous and amorphous categorization of the concepts "conspiracy theory" and "false news", with the aim of silencing the voices of movements and alternative media that contradict the vision of the establishment media.

Terrorism and hate speech

Business Insider compiled complaints from researchers and journalists whose YouTube channels have been suspended, after their videos that spoke about Islam, Syria and related issues were deleted.

The owners of the sanctioned channels assure that none of the eliminated content promoted extremist ideologies, but that they approached them from a neutral and informative perspective.

[9] https://s3-us-west-1.amazonaws.com/onlinecensorship/posts/pdfs/000/000/088/original/Censorship_in_Context_November_2016.pdf?2016
[10] https://onlinecensorship.org/news-and-analysis/facebook-must-go-further-on-transparency

"YouTube has suspended my account because of the videos about Syria that I uploaded 2 or 3 years ago", said Eliot Higgins, a journalist specializing in this conflict.

After confrontations between white supremacist and anti-racist groups in Charlottesville, Virginia, USA, companies like Facebook, Twitter, Paypal, Uber and Airbnb began to exclude white supremacists and neo-Nazis from their platforms.

According to the New York Times, the dating site OKCupid banned a famous white extremist named Christopher Cantwell from entering the dating service. "We do not tolerate anyone who promotes racism or hatred, it's as simple as that", said Elie Seidman, executive director of OKCupid. (...)

Bumble, another dating site, announced the banning of "all forms of hate" through a collaboration with the Anti-Defamation League.

Cloudflare, a web infrastructure company, refused to service the right website, Daily Stormer. Matthew Prince, executive director, said after making that decision: "I woke up in a bad mood and decided that someone should not be on the internet". "No one should have that power".