WIKIMEDIA

F O U N D A T I O N

Wikimedia Foundation submission on UNESCO model framework for platform regulation

02 December 2022

The Wikimedia Foundation welcomes the opportunity to participate in the consultation on UNESCO's draft model regulatory framework for digital content platforms to secure information as a public good.

The Foundation is a nonprofit organization that hosts and promotes Wikipedia and other similar free knowledge projects.[1] Our vision is to ensure that anyone, anywhere in the world, can contribute to the sum of all human knowledge. To make strides towards achieving this, we manage the servers for these projects and support a global movement of volunteers who collaboratively build the knowledge available in these projects.

Below we provide a series of comments to help ensure that the proposed framework supports a public interest, community-governed internet, and strengthens a broader digital ecosystem that enables the effective realization of human rights.

Section One: The Goal of Regulation

1. **Comprehensive definitions of the scope and obligated subjects in any regulation of digital platforms are needed.**

The proposal needs to provide a more precise framework for the scope of any digital content platform regulation. The definition of the problem to be tackled, as well as what is and is not expected from a regulatory system, must be improved. In this sense, it is essential to offer some elements that will help shape the regulated entities and their exceptions and consider establishing modular obligations according to operating models, activities, market size, power, resources, and others. This model should recognize that there is not just the operating model of large, for-profit websites; otherwise, it could be proposing a framework with devastating consequences for nonprofit, public interest-focused websites such as Wikipedia.

In addition, if the model framework aims to guide States in developing regulations to support information as a public good, this is an ideal opportunity to protect the community-governed, public-interest internet, that is, one that is run and maintained by volunteers, and to which everyone online can contribute and use freely. For example, and in this sense, information on Wikipedia is produced, maintained, and discussed according to fully transparent rules developed by volunteers. This volunteer-driven content moderation model has helped make

---

[1] Wiktionary, Wikiquote, Wikibook, Wikisource, Wikinews, Wikiversity, Wikispecies, MediaWiki, Wikidata, Wikimedia Commons, Wikivoyage, and Meta-Wiki.

Wikipedia a global source of high-quality, neutral, and verifiable information available in more than 300 languages. To protect Wikipedia and other community-governed internet platforms, regulatory systems should clearly distinguish between the rules that a host sets—i.e., terms of service—and those established by the communities who use a service or platform. Hosts should neither be required to enforce rules that were created by communities of volunteers who understand relevant language and context nor obliged to engage in top-down content moderation proactively.

The Wikimedia model puts people at the center. Volunteer editors are not only the ones who contribute content but also the ones who develop and implement platform policies. This proposal should promote the role of people in creating and maintaining digital public spaces of public interest, such as Wikipedia. This would be a way to reinforce the user empowerment mechanisms principle outlined in Section 1 of the proposed framework.

   2. **The goals of the regulation should be the same for all users, regardless of age; the specific approaches may account for age-specific needs and challenges.**

A blanket approach to regulation will not be effective. The model regulation should, therefore, avoid overgeneralization and embrace the diversity of platforms and models that exist by ensuring that strict regulation does not unduly restrict freedom of information—which includes both access to information as well as the ability to produce and share information. Depriving people of their rights to seek and access information, education, free expression, and participation in society further disempowers them. Moreover, allowing users to access the "provenance of information" on a platform—a basic requirement for all Wikipedia articles—is one of the five key guiding principles outlined in Section I of the proposed framework.

The document makes repeated references to adult users and the regulatory goals that concern those users. However, the challenges that it identifies are experienced by users of all ages and with varying levels of media literacy. For example, the section on content moderation policies identifies a goal for "adult users to understand the nature and origin of questionable content and accounts and allow them to make their own judgment." (Section 2, Point II)

Given that teen and child users are faced with the same questionable content and accounts and are required to make the same judgments, why is this goal limited to adult users?

The traditional approach—i.e., treating youth as passive subjects of adult authority—severely underestimates their abilities to understand and exercise their rights and to make decisions in their best interests. To do so leads to disregarding their right to access reliable information and to express their opinions freely, to be heard and participate in society on matters that affect them, and to inform decision-making in accordance with their age and maturity, without discrimination.

It is particularly problematic to approach policy questions related to online content moderation and information literacy in this traditional manner for two reasons:

Imagine a world in which every single human being can freely share in the sum of all knowledge.
wikimediafoundation.org · 1 Montgomery St, Suite 1600, San Francisco CA 94104

2

1. Statistical and anecdotal evidence shows that adult users—i.e., members of the generations with the most legal, political, and economic power—are just as vulnerable to certain forms of online exploitation, if not more so, than younger users, including phishing attacks, romance scams, and political disinformation.

2. Globally, one in three internet users is a child. In the global South, this rate is likely to be even higher. At least three generations of digital natives—i.e., Millennials, Generation Z, and Generation Alpha—have grown up under legal regimes that are simultaneously hostile towards their exercising rights and freedoms, and unduly alarmist concerning their vulnerability to exploitation by private actors. As a result, critically important youth perspectives, experiences, needs, and ideas have been excluded from policy discussions.

The final proposal may decide not to cover this age group, in which case it should state so explicitly. But if it does so, it needs to consider and address the issues discussed above.

Section Two: Fulfilling the Goal

1. **Requirements on transparency should also address targeted advertising business models and the algorithms that drive the benefits of ad placement.**

The proposed recommendation can go further by requiring greater transparency and accountability concerning the business models and algorithmic systems influencing the rapid spread of and user engagement with digital content. A better understanding of these aspects would provide better information with which to empower users and help build evidence-based measures to secure information as a common good.

2. **Requirements on content management policies should encourage decentralized, community-driven content moderation practices.**

As stated above, this is an opportunity for UNESCO to promote community-based content moderation models, rather than forcing the implementation of automated models incapable of understanding context and intent. It is also essential that the requirements for content management policies should not be overly prescriptive, so as to allow for the possibility that bottom-up content moderation models can continue to flourish.

In addition, recommending platforms to have systems capable of identifying "synthetic or manipulated media designed to mislead or create a false impression (unless clearly identified as such for artistic or creative purposes)" poses a risk to freedom of expression, especially as it relates to parody. As it is generally known, parody imitates the targeted speech, using irony and exaggeration to show the target's lack of logic or absurdity. Labeling parodic content would defeat its very purpose, highlighting this recommendation's weakness by suggesting that platforms should implement systems that can hardly capture the nuances of speech diversity. Systems that also, incidentally, operate within a business model that benefits from the accelerated dissemination of legitimate content but that, due to their reach, can become potentially harmful.

Imagine a world in which every single human being can freely share in the sum of all knowledge.
wikimediafoundation.org · 1 Montgomery St, Suite 1600, San Francisco CA 94104

3

In the case of Wikipedia, it is volunteer editors themselves who take measures to protect the integrity of the knowledge on the website.[2] They do this through a hybrid content moderation system that relies on human review as well as technical tools to determine whether participating users are contributing high-quality, neutral, and reliably-sourced knowledge. It is important to note that volunteers are also who decide whether any particular content remains in Wikipedia, and not an automated review or moderation system. In addition, anyone can review the editing and decisions made on each article, since everything is publicly visible in the "Talk" and "edit history" logs of these articles.

3. **Fostering an enabling environment requires moving away from paternalistic measures that ignore that people have agency to determine their actions and decisions.**

Imagine what policy conversations about youth, women, LGBTQ+, and users from other marginalized identities online would sound like—beyond the need to protect them from harm—if the global community recognized these people as users and contributors who are agents of their own lives and full rights-bearing members of society.

The measures recommended in point 3 of section II are paternalistic—that is to say, they restrict the actions and decisions of others allegedly thinking of their best interests. We are aware that the groups mentioned in the recommendation, all too often, experience higher levels of violence, which often leads them to self-censorship. However, requiring a platform to provide a higher level of protection for source-based speeches, especially if it is felt that such speeches should be reviewed or removed, is misguided.

We recommend that the focus of this recommendation be on better supporting policymakers, regulators, technology companies, and other stakeholders in their advocacy and accountability efforts to incorporate gender as well as intersectional approaches into their work.

4. **Regulators should not be in charge of deciding whether platforms are "properly" enforcing their content standards.**

It is undesirable to delegate to regulators the task of deciding whether a platform is appropriately enforcing its content standards, as recommended in Section II, point 5. The recommendation makes us wonder under what circumstances and how often regulators can exercise such control powers, what indicators or analysis criteria they will use to make such determinations, what sanctions they could impose, and what consequences those sanctions would have on free speech. Not having clarity on the scope of this power could lead platforms to over moderate content and to the silencing of political enemies, dissidents, journalists, and others.

5. **Media and information literacy is the primary responsibility of governments, in which different stakeholders can also play an important role.**

---

[2] For instance, see: Harrison, S. (21 November 2022,). No, Wikipedia Is Not Colluding With DHS. *Slate*.

Imagine a world in which every single human being can freely share in the sum of all knowledge.
wikimediafoundation.org · 1 Montgomery St, Suite 1600, San Francisco CA 94104

4

Point 6 of Section II proposes to delegate the mandate of media and information literacy to platforms. However, it is worth remembering that literacy is a state's obligation. It would be a different matter if the proposed framework recommended collaboration efforts between the State, technology companies, and civil society groups, among others. In fact, nowadays, it is common for digital platforms to carry out media and information literacy efforts in collaboration with civil society groups and the public sector.

For instance, Wikipedia, a [useful tool for teaching media and information literacy skills](#), is leveraged to responsibly build the capacity to access, understand, and participate in or create content using digital media. Hence, the Wikimedia volunteer communities have engaged in various digital literacy initiatives targeting different social groups (e.g., education communities, senior citizens, indigenous peoples, etc.) and aiming to close the knowledge gaps within Wikimedia projects.

On the other hand, with an increasingly complex range of digital services and technologies, people need to be confident and skilled to engage more meaningfully and responsibly in digital spaces. To assume that only certain groups in society need digital literacy is to place the onus on them not to engage in online situations that may cause them harm. Hence, recommending that platforms implement specific information and media literacy measures for women, children, youth, and indigenous groups is misguided. In that sense, the recommendation should be to invite all stakeholders to implement information and media literacy initiatives that are sensitive to gender and other groups of people.

6. **Digital accessibility should be a guiding principle of the proposed framework.**

The United Nations [estimates](#) that 15% of the world's population lives with some type of disability. Digital accessibility is a human rights issue. If people with disabilities cannot access information in an accessible format, they will not be able to exercise their human rights in digital spaces.

Consequently, we believe that this framework can serve as a guide for digital platforms to incorporate better digital accessibility features, such as mandatory alt-text for images and captions for video, in their reporting mechanisms. This is also true for governments, which should make their regulatory oversight systems more accessible. In fact, it would be fair to ask that the processes put in place by policymakers, regulators, technology companies, and other stakeholders for purposes of their advocacy and accountability efforts be compatible with assistive technologies.

Section Three: The Independent Regulatory System

The proposed framework seeks to give regulators broad authority to: define platforms in scope based on risks posed to the user, determine whether platforms are enforcing their own content moderation policies, and dictate the penalties to be imposed on the platforms and/or how platforms must alter their policies and processes to comply with new regulations.

Imagine a world in which every single human being can freely share in the sum of all knowledge.
wikimediafoundation.org · 1 Montgomery St, Suite 1600, San Francisco CA 94104

5

The framework relies heavily on the idea of a "completely" independent regulatory system, made up of either existing or to-be-created regulators in each state, and on an "oversight body" to hold the independent regulator accountable.

The document describes these processes generally, failing to acknowledge or engage with two important realities:

1. It is neither possible to create a truly independent regulator nor ensure the "complete" independence of an existing regulator with an expanded scope that is reporting and solely accountable to a legislature or newly created board.
2. Notwithstanding the above, it is necessary to set clear, specific processes and procedures to ensure *any* baseline level of independence for whichever regulators and oversight bodies are created or designated.

1. **It is impossible to create a truly "independent" regulatory system, free from external economic and political pressures and influences, to govern online platforms' compliance with their own terms and conditions and content moderation policies.**

The proposal sets out an impossible standard for a "truly independent" regulatory system: it needs to be "free from economic or political pressures or any external influences" and will not seek or take "instructions from any external body." This impossibility of this standard is evidenced, in part, by the fact that there isn't a single legislative or regulatory system or body in existence, either currently or historically, that meets the standard.

The proposal also states that the regulatory system must have "sufficient funding to carry out its responsibilities effectively" but does not address where this funding will come from, what sources of funding are permissible, or how the system will be "free from economic or political pressures" without independent, non-government (including intelligence and law enforcement) sources of funding.

2. **It is not possible to ensure the independence of an oversight body, or to ensure that the regulator will not be subjected to external influences and pressures or seek or take instructions from the designated oversight body.**

The proposal states that the regulatory system must not seek or take "instructions from any external body" but must report to "the oversight body," to which it will be accountable. The oversight body can either be a legislature or a board. The proposal makes no distinction between federal, national/regional, or state legislatures. Additionally, it fails to provide any guidance for creating transparency and accountability mechanisms for the oversight body.

The conflict-of-interest reporting rules that apply to the regulator must be as stringent as, if not more than, those that apply to the oversight body; moreover, those that apply to the oversight body must be stronger than those that typically apply to legislators.

Imagine a world in which every single human being can freely share in the sum of all knowledge.
wikimediafoundation.org · 1 Montgomery St, Suite 1600, San Francisco CA 94104

6

Legislators worldwide engage with and are lobbied by other external bodies and entities, many of which are founded and staffed by former legislators and legislative staff. The proposal does not address the "revolving door" that exists between legislatures, industry, and private lobbying groups.

It also fails to address the issue of external entities and individuals that may use individual legislators or the larger legislature—acting as an oversight body to the "independent" regulator—as a conduit to influence the regulator improperly or to influence investigations and enforcement actions taken by the legislature against the regulator.

3. **Regulators should be held to a higher standard than simply acting within the bounds of existing law, particularly where existing law conflicts with fundamental human rights principles.**

The proposal's baked-in checks on regulators' exercise of power are limited: "any regulator should act only within the law" and respect "fundamental human rights." However, if this is left unrevised, it could have the opposite intended effect: providing a blank check to legislatures and governments acting within the bounds of authoritarian laws or enforcing them, even though they are inherently incongruent with fundamental human rights.

4. **The framework must address and account for jurisdictional differences in ages of maturity, permissible and impermissible speech, and respect for the rights and freedoms of non-majority groups.**

Jurisdiction-specific and context-specific ages of maturity

Different jurisdictions and countries classify these broader categories of users and content according to their preexisting legal frameworks. Any online content-related regulations will use each respective state's legal definitions and, in particular, penal codes.

For example, countries often do so differently when classifying age ranges as "adults" and "youth" under the law. One need only do a cursory search for the age of consent, motor vehicle, voting, alcohol, firearms, tobacco, gambling, and other related laws, which set out various ages of maturity, to see that there is no globally-agreed-upon definition of "adult" or "child."

One country's permissible speech is another's "egregious" or "illegal" content; different approaches to "hate speech"

In the online safety and content moderation contexts, some countries target content that "creates disharmony" between groups or criticizes public officials or the government.[3] In other places, content that accurately describes racism and racial inequities—including the histories of

---

[3] See Singapore Online Safety (Miscellaneous Amendments) Bill; see also Protection from Online Falsehoods and Manipulation Act (POFMA).

Imagine a world in which every single human being can freely share in the sum of all knowledge.
wikimediafoundation.org · 1 Montgomery St, Suite 1600, San Francisco CA 94104

7

slavery, segregation, and apartheid—is categorized under the law as "hate speech" towards the dominant racial groups.[4]

Without carefully crafted and nuanced definitions, important and legitimate information by and about certain minority communities will be removed under the same policies that purport to protect these groups by requiring the removal of *actual* hate speech, including the use of derogatory epithets and threats of violence.

Conclusion

The community-based content governance model has allowed Wikipedia to include divergent perspectives within editorial discussions on content while being a trusted and reliable source of knowledge. This has led to Wikipedia and other Wikimedia platforms becoming respected reference work on the internet. Therefore, we believe it is a model that should be considered and even encouraged when designing any digital platform regulation. This, in turn, will also offer a way to strengthen the exercise of human rights in the broader digital ecosystem.

---

[4] For instance, the states of Florida, Georgia, Idaho, and Mississippi in the United States of America. See also: PEN America, Education Gag Orders: Legislative Restrictions on the Freedom to Read, Learn, and Teach; Education Week, "Map: Where Critical Race Theory is Under Attack"; and Movement Advancement Project, and Equality Federation, "Policy Spotlight: Curriculum Censorship & Hostile School Climate Bills."

Imagine a world in which every single human being can freely share in the sum of all knowledge.
wikimediafoundation.org · 1 Montgomery St, Suite 1600, San Francisco CA 94104

8