



Protecting Shared Societies in the Digital Space: how to build inclusive online discourse?

1. Club de Madrid (CdM)

<u>Club de Madrid (CdM)</u> is the world's largest forum of democratic former Presidents and Prime Ministers, who leverage their individual and collective leadership experience and global reach to strengthen inclusive democratic practice and improve the well-being of people around the world.

As a non-partisan and international non-profit organisation, it counts on the hands-on governance experience of more than 100 Members from over 70 countries, along with a global network of advisers and partners across all sectors of society.

This unique alliance stimulates dialogue, builds bridges and engages in advocacy efforts to strengthen public policies and effective leadership through recommendations that tackle challenges such as inclusion, sustainable development and peace, at the national and international levels.

2. The Coalition for Religious Equality and Inclusive Development (CREID)

The <u>Coalition for Religious Equality and Inclusive Development (CREID)</u> is a consortium, convened by the Institute of Development Studies (IDS). It looks at how poverty reduction efforts can actively redress religious inequalities, support inclusive, religiously diverse communities and promote the benefits of interdependence among people beyond religious and non-religious lines.

CREID, which is funded by the UK Government's <u>Aid Connect</u> programme is governed through a collaboration between the Institute of Development Studies <u>al Khoei Foundation</u>, <u>Minority Rights Group</u> and <u>Refcemi</u>, working in partnership with dozens of NGOs, social movements, academics, media and practitioners worldwide.

The consortium draws together a formidable wealth of expertise and experience in understanding the way in which individuals and groups' right to belief (or non-belief) is upheld or violated, in poverty reduction and pathways of empowerment and in informing policy and challenging practice.

3. Rationale

One important element in the recent evolution of the multilateral system's governance and of democratic societies has been the growth and role of digital transformation and Artificial Intelligence (AI) in almost every facet of our lives. Digital technologies have

brought unprecedented opportunities and challenges and we must find a way to maximize its benefit and minimize harm. The COVID-19 outbreak and ensuing global health crisis clearly accelerated the widespread deployment of digital technologies and Artificial Intelligence, underlining our structural interdependence and the need to strengthen international cooperation and the multilateral system as we build a common digital governance mechanism capable of ensuring an open, free and secure digital future for all.

Of the many challenges to be addressed in the field of digitalization and Artificial Intelligence, we need to continue pursuing an inclusive, accountable, fair, transparent, sustainable and ethical system centred around people, where rights and a rules-based order are respected, and is of fundamental importance. Such a system would contribute to Shared Societies¹, as it is built on the inclusion of all peoples and truly leaving no one behind. Digital technologies have been elevated into fundamental instruments of resilience and inclusion, but they also have increasingly become instruments of disinformation, exclusion, hate speech, and incitement to violence. Their prominence, in all areas of society is not declining and the pandemic is receding, so, it is still important to tackle new dynamics of exclusion and discrimination arising from the digital transformation.

The negative impacts of digital transformation are disinformation, exclusion, hate speech, and incitement to (and organizing of) violence against individuals, groups, and political opponents, among others. Hate speech and verbal abuse online continues to be frequent and, although there has been progress in content regulation and the protection of fundamental rights online, people are still suffering attacks related to racism, discrimination based on sexual orientation and religious beliefs.

In this context, we proposed to co-convene a consultation process built on CdM's work on the role of political leaders in tackling divisive rhetoric and promoting inclusive speech, and on the study on migration narratives in Latin America, as well CREID's track record in monitoring and countering hate speech in the Middle East and Asia. Such a meeting aimed to turn our attention to the particular dynamics of online discourse -- how divisive and exclusionary rhetoric is spreading over the internet, the dehumanizing and potentially violent consequences of that rhetoric, and what innovative practices different power holders can take to generate counter-narratives, pro-actively shape this important space as well as foster a more inclusive and egalitarian discourse.

¹ Club de Madrid defines a <u>Shared Society</u> as one in which all individuals and constituent groups hold status as equally contributing participants, free to express their differences while integrating their voices within the broader population. It respects everyone's dignity and human rights while providing every individual with equal opportunity.

4. Background note on Club de Madrid's contribution to Our Common Agenda report

On November 27th, 2020 Club de Madrid President, Danilo Türk, received a letter from UNSG inviting "the distinguished members of the Club de Madrid, drawing on their extensive experience as world leaders, to share ideas with me in the coming months on how we can put in place the inclusive, networked and effective multilateralism that is necessary to address the world's biggest global challenges both today and in the future."

To respond to the UNSG's request, the CdM Secretariat, under the leadership of President Türk, launched an internal consultation process which included a series of online, regional conversations with CdM Members and Advisors, and a written procedure to ensure that all were able to contribute to this joint reflection.

As a result of this process, CdM prepared a report with concrete proposals aligned with the <u>12 commitments</u> of the <u>Declaration on the Commemoration of the Seventy-fifth Anniversary of the United Nations</u>, aimed to put in place an inclusive, networked and effective multilateralism able to address the world's biggest global challenges.

Once *Our Common Agenda* report was released, and following CdM's strategic pillars for 2020-2022, our organization continued working on distilling the elements crucial to building a human-centred governance of Artificial Intelligence and digital societies. Our intention was to feed into ongoing regional and global discussions and serve as a platform for dialogue and an instrument for building bridges between different communities of practice in the design and construction of an internationally agreed normative framework to govern and regulate the era of digital transformation and Artificial Intelligence.

In this context, CdM had the opportunity to participate in the **follow-up to the UNSG Our Common Agenda Report informal thematic consultations** convened by the President of the General Assembly in February 2022. Member Hanna Suchocka participated in the 3rd Cluster, "Frameworks for a peaceful world – promoting peace, international law, and digital cooperation," focusing on CdM's support for all three issues and highlighting the need to strengthen international cooperation and the multilateral system as we build a common digital governance capable of ensuring an open, free and secure digital future for all where Human Rights and rules-based order are respected and of fundamental importance.

5. Purpose of the meeting

CdM in partnership with CREID generated a reflection process to understand online hate speech, how to create and spread content which improves attitudes or contributes to the prevention of violent action; and engage in advocacy for policies addressing religious minority discrimination, to ensure that all individuals are able to exercise their right to freedom of expression online, while also protecting them from harm and ensuring a safe and inclusive online environment.

6. Agenda

<u>Date:</u> 28 April, 2023

Time: 8:00 – 9:30 EST / 14:00 – 15:30 CEST

Zoom Link:

https://clubmadrid-

org.zoom.us/j/81902881503?pwd=Z1o1cmY1NFNjcWNBdUEra05XbWFkZz09

- 1. Welcome words by CdM and CREID
- 2. Introduction of the meeting's objectives
- 3. CdM Members' interventions
 - Zlatko Lagumdžija Prime Minister of Bosnia and Herzegovina (2001-2002)
 - Rexhep Meidani President of Albania (1997-2002)

4. Open discussion to all participants

- Mariz Tadros, Professor of Politics and Development at the Institute of Development Studies (IDS), University of Sussex and Director of CREID
- Haroon Baloch, Programme Coordinator, Bytes for All
- Salam Omer, Executive Director, Kirkuk Now
- Jaffer Mir, Al-Khoei Foundation
- Maryam Kanwer, HR Advocate and Researcher HR Advocate and Researcher, Al-Khoei Foundation
- Ali Abbas, Executive Director, HIVE Foundation
- Wendy Via, Co-founder//CEO and President, Global Project Against Hate and Extremism
- Heidi Beirich, Co-Founder //Global Project Against Hate and Extremism
- Azin Mojarherin, Human Rights Lead, MIAAN Group
- Htet Swe, Executive Director, Peace Point-Myanmar (PPM)
- May Homira Rezai, President, Hazara Committee in the UK
- James Chalm, Stakeholder Engagement Manager, Content Policy at Meta
- Peter Stern, Director, Content Policy Stakeholder Engagement at Meta
- Esther M. McIntosh, Senior Advisor, Shared Societies Project, CdM
- Rubén Campos, Programmes Coordinator, CdM
- Diego Gitano, Programme Officer, CdM
- María Fernanda Robayo, Programme Officer, CdM

5. Wrap up & conclusions

Facilitator: Claire Thomas, Co-Executive Director, Minority Rights Group International (MRG)

7. Recommendations

Apply human rights online

a) Core Principles

- To respect existing international norms and standards on hate speech, freedom of speech, discrimination and inclusion and develop novel ways of responding to this dynamic space.
- To recognise that religious (as well as ethnic, religious, racialised, linguistic and descent) minorities are particularly vulnerable to hate speech amplified by digital divide and the relationship between online-offline mobilization of speech hate.
- To promote the development of people-led approaches to counter hate speech through the localisation agenda.
- To incorporate digital awareness as a development and human rights in its social development and poverty alleviation programmes.
- To share and circulate best practice in countering hate speech and encourage knowledge exchange among communities of practice.

b) Key Commitment/ Pledges/ Actions

UN/Intergovernmental

- Make the UN Forum on Business and Human Rights accountable for human rights commitments and compliance of businesses operational in the digital sector, including funding marginalised- and hate speech impacted- communities to attend the Forum for an initial period.
- To encourage multilateral agreements that recognise that businesses in the digital sector have a special responsibility to be held accountable for operations and impacts beyond the national borders of states in which they are established.
- To include measures for building an accountable, inclusive digital sector in the UN Business and Human Rights Treaty
- To incorporate indicators and targets in the post 2030 SDG agenda pertaining to freedom from hate speech and follow up mechanisms of implementation
- To ensure that UN Senior staff including UN Resident Coordinators are fully aware of hate speech and its risks and create action plans in consultation with affected groups.

Governments

- To take a very cautious approach when legislating against online hate speech as there
 are examples of such legislation being used against peaceful and legitimate freedom
 of speech.
- To legislate to regulate the human rights compliance of digital provision companies with extraterritorial coverage given the inherently global and transnational nature of almost all digital services and functions.
- To use soft diplomatic power to influence all UN member states to address breaches of human rights online in state treaty reports (CERD, CEDAW, ICCPR etc.) as well as during Universal Periodic Reviews and other human rights dialogues.

Private sector

- Maintain appropriate levels of human rights trained specialist staff, efficient in different languages, at all times and regularly review all company processes for human rights compliance and any negative human rights related downstream or unintended effects, take swift and effective action to remedy problems identified.
- Ensure transparency in content moderation processes while responding to government requests to remove or block legitimate content under the guise of hate speech or blasphemy.
- Attend and contribute constructively to the UN Forum on Business and Human Rights.

Civil society

- Provide independent input into draft new legislation, implementation of existing legislation, oversight board decisions and similar.
- Undertake strategic litigation to establish precedents concerning egregious, flagrant, longstanding and/or high risk breaches of human rights.

Accountability for discrimination and misleading content

a) Core Principles

Recognise that content that incites or encourages discrimination against any individual or group linked to any protected characteristic is defined as prohibited hate speech (ICCPR), and is a breach of international human rights treaty commitments and should be prohibited i.e. that threat of physical violence is not an essential defining characteristic of all hate speech.

b) Key Commitment/ Pledges/ Actions

Governments

- To include digital literacy and critical thinking, including fact checking content, in all education curricula (diversity-friendly).
- To educate adults via public information materials (multilingual and diversity inclusive in design and content) on the dangers of believing everything you can read, hear or see online.
- To seek to agree pre-election cross party agreements to avoid scapegoating vulnerable sections of society for electoral gain (potentially with support from neutral third parties).
- To establish truly independent mechanisms to investigate allegation of state resources being used to fund or support the production of hate speech online content and provide access to data and records to allow full and transparent investigations.
- To distribute government advertising transparently and giving priority to platforms that don't carry hate speech or which positively promote counter hate speech content.

Governments, UN and private sector

• Establish (arms length managed) funds to support civil society counter hate speech projects e.g. public education of high risk groups, positive messaging, anti-hate speech campaigns.

Private sector

- Increase the level of content moderation staff (especially in high hate speech risk locations).
- Invest in Artificial Intelligence to more accurately and more consistently prevent hate speech being uploaded ensuring racial, religious, linguistic biases are avoided.
- Carry out regular diversity audits of all staff teams and seek to maximise inclusion of under-represented and at risk groups in staffing at all levels and in all functions.
- Ensure full participation of hate speech impacted communities in all new and revised content policies.
- Enter into regular dialogue with representatives of minority and other hate speech impacted communities to understand their experience of using digital services and platforms.
- Fund public interest journalism platforms that promote counter hate speech with advertising and support for joint initiatives.

Civil Society

- Support and empower grassroots marginalised groups to take the lead on countering hate speech e.g. on establishing the levels, targets, perpetrators and impacts, educating journalists, influencers, and editors on the damaging impacts of hate speech, positive messaging campaigns to rebalance content biases.
- Supply independent evidence based data for advocacy to show how, where and why hate speech is still appearing in digital spaces, its impacts and constructively suggest measures that can address it.

Signatories

- AlKhoei
- Bytes for all
- Club de Madrid (CdM)
- The Coalition for Religious Equality and Inclusive Development (CREID)
- Global Project Against Hate and Extremism (GPAHE)
- Hazara Committee in UK (HCUK)
- KirkukNow Media Outlet Iraq
- Mariz Tadros
- Minority Rights Group (MRG)
- Peace Point

Annex: Good practice examples

Of monitoring hate speech

https://minorityrights.org/publications/online-hatred-pushing-minorities-to-the-periphery-an-analysis-of-pakistani-social-media-feeds/

https://minorityrights.org/publications/hatefull-places/

Of publishing counter hate speech

https://kirkuknow.com/en/myStory

 $\frac{https://creid.ac/blog/2022/09/21/mohsins-story-grassroots-activist-from-poor-christian-quarter-in-islamabad-now-writes-for-major-media-outlets-in-pakistan/$

Of Anti-Hate Speech campaign

https://www.ohchr.org/en/get-involved/stories/i-dont-forward-hate-online-campaignagainst-hate-speech-pakistan

On building the capacity of CSOs to tackle all aspects of hate speech

https://minorityrights.org/publications/hate-speech/

On cross national research to understand the phenomenon and share learning:

https://www.ids.ac.uk/events/incite-identifying-and-challenging-online-hate-speech/