

Group 1

Main Session Proposal

Draft 2 – v. 16.07.2019

Co-organizers: Sylvia Cadena; Susan Chalmers and Jutta Croll

Mailing list msession_cont_gov@intgovforum.org

Title: Content Governance: Rights, Responsibilities, Responses and Risks

Brief description / objective

(200 words or less high-level pitch for participants to join this session)

The reach of content published online is amplified through social media platforms at a speed never seen before. This has promoted democratic values by empowering individuals, giving a growing voice to those who have not been heard before. Among this potential there are increasing risks associated with the proliferation of hate across social media platforms, including the amplification of violent extremism, which has affected the way people feel safe and secure both online and offline. This session will focus on content governance by Internet platforms, or the policies by which different organizations determine whether to remove, let stay, flag, or otherwise treat harmful content that is uploaded to their platforms. It will consider various regulatory and non-regulatory approaches to these important and complex Internet policy issues.

Full Description

The reach of content published online is amplified through social media platforms at a speed never seen before. This has promoted democratic values by empowering individuals, giving a growing voice to those who have not been heard before. These platforms are a great example of how Internet-powered innovation have enhanced the way people participate in society from an economic, social, political, and cultural perspective.

Among this potential there are increasing risks associated with the proliferation of hate across social media platforms, including the amplification of violent extremism, which has affected the way people feel safe and secure both online and offline. In light of events like the March 2019 Christchurch mosques attacks, there is a growing expectation for responses that quickly identify threats and bring effective action to prevent and minimize the damage that the viral amplification of harmful content can cause. In this context, it is important to discuss how harmful online content -- and hate speech -- is defined in a human rights framework, including the risks associated with such responses as different rights and freedoms come into play (such as freedom of expression, freedom of thought, conscience and religion).

In response, social media platforms have embarked on processes to develop their own community standards, incorporating the feedback from their community of users to drive upgrades, new services and tools as well as decide what is acceptable content and behavior online and what is not. Besides these efforts around self-regulation, other approaches to

these challenges include co-regulation, working with regulators, etc. However, it is not entirely clear how these processes are strengthening the rule of law, following due process, and been consultative, inclusive and open enough.

This session will focus on content governance by Internet platforms, or the policies by which different organizations determine whether to remove, let stay, flag, or otherwise treat harmful content that is uploaded to their platforms.

- **Rights** – Although international human rights law is only directly binding on states, under the UN Guiding Principles on Business and Human Rights, platforms have a corporate responsibility to respect human rights, including freedom of expression. In regard of children platform providers' responsibility can also be derived from the UN Convention on the Rights of the Child (Art. 3,1) and the EU Charter on Fundamental Rights (Art. 24,2). How is this put into practice in content governance?
- **Responsibilities** – How far does the responsibility of an Internet platform stretch with regard to account holders? How should account holders be responsible for their own behavior – i.e. not spreading abusive content or reporting it when they encounter it?
- **Responses** – Different governments have responded in different ways to this issue, from creating criminal liability for executives through legislation to joining voluntary declarations, such as the Christchurch Call. How are the various government responses creating an impact – on the root cause of the problem, on business operations, and the free flow of information online?
- **Risks** – What are the potential risks to the freedom of expression or, alternatively, the health and safety of account holders, of different approaches to content governance? How are these risks addressed?

Policy Questions

1. **Rights:** How to define hate speech and harmful content? What legal definitions do we already have based on national or regional legislation? Would a global definition make sense? Which types of harmful content should be covered? What is meant by "harmful" and to whom is that content "harmful"? How can the needs of especially vulnerable groups be addressed; how can they be protected? How do the several types of online harm affect and infringe human rights?
2. **Responsibility:** As platforms operate on a global market, local action for change has very limited impact, so when bringing about change at a global scale, could global norms help drive security and safety for all Internet users? Or will the lack of accountability inspire states to move towards a more regulatory approach with the associated risks of fragmentation? How can multi-stakeholder cooperation in that respect be reinforced to establish global norms and accountability?
3. **Responses:** What is the most appropriate policy development process? Are policies solely determined by platform operators legitimate, appropriate, effective? Should platform operators rely on their community of users for developing user policies? Should they wait for government regulations to determine how they process content? Should citizens or governments have more of a say? Are these processes consultative and inclusive enough? What are the technical and operational considerations that need to be taken into account? What mechanisms are needed to achieve this? Will forcing platform operators to take legal responsibility for illegal or harmful content uploaded by their users suppress too much speech? While not all harmful content is illegal how can the difference be addressed in regard of platform

operators' responsibility? Is it reasonable to expect platforms to identify and take down all objectionable content before it is ever aired?

4. **Risks:** Would a failure to establish norms on this subject increase the pressure for governmental regulation of platforms, the Internet, or both? What are the risks of fighting harmful content for example online hate over human rights and in particular on the freedom of expression? Is the use of social media platforms by foreign governments to provide alternative views on sensitive international issues a threatening action that needs to be controlled, or a broadening of access to information?

Agenda

The discussion will be divided in 3 main parts:

- **Introduction:** welcome remarks from the moderator, introducing the structure of the session and the speakers. Instructions to the use of speaking queue, Q&A on site plus remote participation. Reference to outcomes (policy and research questions and table of initiatives). Request audience to use #contentgov #IGF2019 hashtags
- **Understanding the context: Rights and Responsibilities:** This part of the agenda will focus on the policy questions around rights and responsibilities. This part of the agenda could be presented by 3 speakers that can bring context and background to the discussion 10 min from a human rights perspective.

To encourage interaction, the moderator will ask a couple of questions focusing on rights and responsibilities and will take questions from the audience about this part of the session. Questions from the Policy Questions list could be used to support this section.

- **A reality check: responses and risks:** At least 3 concrete examples around content governance approaches. The first example will focus on government action. The second example will look at how the industry is taking action and coordinating a response. The third example will be looking at community action and civil society engagement.

To encourage interaction, the moderator will ask a couple of questions focusing on responses and risks and will take questions from the audience about this first part of the session. Questions from the Policy Questions list could be used to support this section.

- **Ideas for the future:** a conversation among the speakers from the previous two sections (government representatives, technical community, private sector and civil society representatives) looking into the future.

On this last block of questions, the moderator will encourage discussion among the speakers and take questions from the audience. Questions from the Policy Questions list could be used to support this section. The floor will be open for questions from the audience.

Chair(s) and/or Moderator(s)

MAG co-organizers and rapporteurs:

Sylvia Cadena, GRULAC & WEOG, Technical Community, Female
Susan Chalmers, WEOG, Government, Female
Jutta Croll, WEOG, Civil Society, Female

- **Moderator:** Jordan Carter, InternetNZ
WEOG (New Zealand). Technical Community. Male
- **Remote participation moderator:**
MAG member Lucien Castex
WEOG (France). Civil Society. Male.

Panellists/Speakers

Rights & Responsibilities:

- 3 speakers not yet defined to focus on the human rights framework and responsibilities

Responses & Risks (3 speakers, one per example):

- Government responses
- Industry responses
- Community responses: civil society campaigns

Plan for in-room participant engagement/interaction?

Comments and Q&A will follow each segment of the session making use of the facilities in the room (microphones in aisles) to enable the audience and remote participants join the conversation. The moderator will engage with the audience and encourage them to ask questions, managing the flow of the discussion.

Remote moderator/Plan for online interaction?

There will be a remote moderator that will encourage the remote participants to ask questions and make comments and assist with how the online conversation is integrated with the discussion in the room. The plan for the session will be posted prior to the event, questions and comments will be gathered to enrich the on site participation. The organizers will liaise with the IGF secretariat to engage remote hubs to gather input prior to the event, in case the real time options are too difficult to handle.

Connections with other sessions?

On the IGF 2019 program, there are sessions around hate speech approved for inclusion in the agenda as follows:

#150
#177
#218

The session will also be linked to the Main Session on Technical and Operational issues that was organized last year, which focused on Content Blocking and Filtering. The session covered the importance of definitions, due process, and technical implications (around 3 or 4 of the policy questions that we have listed here were covered on that session to a certain extend). Proposals & Report: <https://www.intgovforum.org/multilingual/content/igf-2018-technical-operational-topics>

From previous IGFs:

<http://beta.friendsoftheigf.org/sessoes/show/2017-geneve/a-net-of-rights-human-rights-impact-assessments-for-the-future-of-the-internet>

Desired results/output? Possible next steps for the work?

Not yet defined.