

IGF 2017 DC COS	17-21 December 2018	Geneva, Switzerland	Report submitted by Marie-laure Lemineur, ECPAT International 17 January 2018 on behalf of the DC COS. V1.
Speakers and moderators	<ul style="list-style-type: none"> • Moderator: Marie-laure Lemineur, ECPAT International, Thailand • Remote moderator: David NG, eHelp Association, Hong Kong • John Carr, ECPAT International, UK • Larry Magid, member of the Safety Advisory Boards of Facebook, Twitter and Snapchat , USA • Mike Tunks, Public Affairs Manager, Internet Watch Foundation, UK • Marco Pancini, Director of EU Public Policy, Google • Karuna Nain, Global Safety Programs Manager, Facebook, USA 		
<p><u>Session title:</u></p> <p>Moderating illegal content online and staff welfare – Where do the responsibilities lie?</p> <p>DECEMBER 20, 2017</p> <p>DC on Child Online Safety</p> <p>4:00 – 5:00 P.M.</p> <p>ROOM XXVII</p>	<p>The session addressed the topic of moderating online illegal content. In other words, the price others pay for digital dumping – Aspects of online child protection. Huge volumes, unknowable quantities, of unambiguously illegal or profoundly harmful materials are circulating on the Internet. Child sex abuse materials (CSAM) and terrorist propaganda are the types of content most frequently mentioned in this context but there are several others.</p> <p>There has been an entirely proper focus on the supply chain companies use to manufacture or deliver their products or services. Typically, these initiatives have been designed to eliminate child labour, slavery or environmental harms. Isn't it time internet businesses and institutions were pressed to do something for those who daily have to face the unfaceable on our behalf?</p> <p>Already we are aware of at least one case that is being brought in a US court by ex-moderators who claim their former employer did not do enough to shield them from Post-Traumatic Stress Disorder. Whatever the eventual outcome of that case might be it seems a portent of other actions that could be brought. However, our aim should be to avoid the possibility of such suits by insisting on high standards.</p>		

	<p>INHOPE co-ordinates a global network of hotlines that receive reports of CSAM. To be a member of INHOPE a hotline has to sign up to several commitments in terms of staff welfare needs e.g. access to counsellors, safe spaces and secure places.</p> <p>Any company or institution/platforms employing moderators, either in-house or via third parties, should sign up to something similar to the standards that institutions like INHOPE and IWF use for their own analysts who review CSAM all day long, something similar and there should be a mechanism to reassure the public that its terms are being honoured in practice, not just in theory.</p> <p>After introductory remarks made by each presenter, the discussion was opened among panellists followed by Q&A.</p>
	<p><u>Summary of the remarks made by the presenters and participants:</u></p> <ul style="list-style-type: none"> -One study has found that there is a similar pattern between countries where old/used digital devices are dumped and where outsourced moderators are employed by internet companies to look at illegal content (visual digital outputs) and have to make a decision on whether the content is illegal or be part of the decision-making chain regarding whether or not this particular content is either illegal or contravenes the terms of services of the company that hired them to do the moderation; -The same study reveals that those outsources moderators based in the developing world look at extreme content on our collective behalf with low wages and poor working conditions. After conducting field trips, the author found that there was huge difference between moderators of those companies working at headquarters in a controlled and safe environment; -Moderators and analysts of illegal content are doing this job to make sure, all of use as internet users, can be safe online and do not have access to extreme/illegal contents, in other words, they contribute to freedom of expression by keeping the internet free from illegal content, making space for legitimate use of the internet. Analysts and reviewers are fulfilling social role. - Analysts/moderators are ordinary people doing an extraordinary

job;

-Employers of moderators have a duty of care to them;

-Human moderators should be employed combined with artificial intelligence solutions;

-Overall, we do not know the exact number of moderators employed right now. Google announced that they intend to employ 10,000 of them by the end of 2018 and Facebook announced an increase by 20,000;

-The Internet Watch Foundation (IWF) is the UK hotline tasked with removing child abuse images. In 2014, IWF was authorised by the government to do proactive searching of child abuse imagery in the UK and as a result, compared to 2013, IWF saw a 138% increase in the materials that their analysts were analysing. This had a significant impact on the welfare of the analysts. IWF has around 20 analysts. The volume of illegal content the analysts deal with is as follows: in 2016 they took down 60,000 webpages, each likely to contain hundreds to thousands of child abuse images. 50% of those images portrayed children being sexually abused in different ways under the age of ten and 10 % of those under the age of 2;

-For IWF there are two main welfare issues for moderators/analysts:

1/The psychological impact of the type of work done by these moderators/analysts. There has been not enough work done in this field to understand the implications of looking at extreme materials; and the

2/The welfare of the institution itself. In order to monitor potential peculiar behaviour of analysts, proper mechanisms are in place starting during the recruitment process which includes a psychological assessment. There is a 6-months induction process and mandatory counselling support in addition to other measures related to the working environment;

-One key issue is to acknowledge that technology can improve or add to the work done by analysts. It is key to have the analysts not see duplicates to lower the work load and also diminish the impact of viewing over and over again the same images;

-Facebook as a pro-active reporting mechanism set up for FB users. FB has teams of trained analysts who will look at those reports per subject matter. Teams are trained across the areas. Reviewers with the native language are familiar with local nuances. But moderators are not content experts. There are psychological reports and wellness programme in place. Managers are

incentivised for focusing on wellness of the team members. We need to put things into perspective. Only a very small percentage of what they see is violent or extreme. FB chooses only reputable contractors who take care of their employees and that have wellness and psychological programmes in place. Some moderators are based at FB facilities, others are outsourced;

- GOOGLE: acknowledged that repeated exposure to extreme and violent content has an impact on reviewers. There is a need for mitigation. Ideally technological solutions should be able to reviews those contents but "the human eye is needed at some point to make sure algorithms are not deciding what we find online;
- With regard to automated solutions there are progress made. For example, 98% of deflects for YouTube that Google is receiving for review of their team come from automated search on the platform. Google's policy is to avoid repeated exposure of the analysts to the same old videos, for example by extracting only frames of it. Other measures such as privacy screens, blurring, etc. are also implemented, managers monitoring any sign of stress, ...
- NCMEC precedes all companies offering counselling and screening processes and making a reality that the people who are doing this type of jobs are getting the resources they need such as working in an isolated section. They are now even considering moving to a new building;
- it is incumbent upon the tech industry to take responsibility for the conditions under which people who are working on their behalf, whether employed or contractual, are getting the same treatment they need;
- One example of initiative is the NGO Connect Safely which is trying to define standard of proof process at day 0 so when companies start to process of rolling out their products, they think about safety on all levels. We should be adding how contractors treat analysts and moderators;
- It is important to constantly reevaluate moderators' behaviors and the rotating process is also key. No one should have a long career as an analyst;
- One idea could be to create a trade association of moderating companies;
- Self-regulation has worked well in the space of child sexual abuse material particularly in the UK;
- Aspects related to handling the review of illegal content against the companies' terms of services standards and the legal standards

	<p>in a particular jurisdiction were discussed;</p> <ul style="list-style-type: none">- There are challenges in finding the right combination with respecting the terms and conditions which are global and the respect for local laws was mentioned; <p>END.</p>
Gender	42 participants: 17 women and 25 men
Link to transcripts	<ul style="list-style-type: none">• DC COS session: https://www.intgovforum.org/multilingual/content/igf-2017-day-3-room-xxvii-dc-child-online-safety-raw