Quarterly Report on Cyber Hate (May, June and July 2016)

By Tamás Berecz and Charlotte Devinat

Project Research - Report - Remove: Countering Cyber Hate Phenomena

INACH

Executive Foreword

This publication was written within the framework of the *Research – Report – Remove: Countering Cyber Hate Phenomena* project of the International Network Against Cyber Hate (INACH); funded by the European Commission Directorate-General for Justice and Consumers. The duration of the project is 2016-2017, and its aim is to study, document and report on online hate speech in a comparative and comprehensive way; and to establish structures for a transnational complaints system for instances of cyber hate.

Hate speech is intentional or unintentional public discriminatory and/or defamatory statements; intentional incitement to hatred and/or violence and/or segregation based on a person's or a group's real or perceived race, ethnicity, language, nationality, skin colour, religious beliefs or lack thereof, gender, gender identity, sex, sexual orientation, political beliefs, social status, property, birth, age, mental health, disability, disease.

This report was completed with the participation of the different members of the Network and partners in the project, namely the Zivilcourage und Anti-Rassismus-Arbeit (ZARA) from **Austria**, the Movimiento contra la Intolerancia (MCI) from **Spain**, jugendschutz.net from **Germany**, the Ligue Internationale Contre le Racisme et l'Antisémitisme (Licra) from **France**, the Inter-Federal Centre For Equal Opportunities and Opposition to Racism from **Belgium** (now called Unia), and the Magenta Foundation from the **Netherlands** (MDI); who provided most of the data this report is based upon.

Legal Disclaimer

Table of content

I. Introduction

As a fundamental part of the Research - Report - Remove: Countering Cyber Hate Phenomena project, INACH collects data from all project members from multiple countries (Austria, Belgium, France, Germany, the Netherlands and Spain) on a monthly basis. We collect and merge these pieces of data in order to synthesise a comprehensive and extensive picture of cyber hate in Europe in the 21st century. In this report, the data collected between May and July 2016 will be explored and discussed. Furthermore, INACH also - with the help of the project partners - collects information on drivers, trends and tools that lie behind online hate speech.

As it will be denoted later on with the data collection regarding hate types, antisemitism, racism, anti-Muslim hate and anti-refugee hate were the highest on the list. Regarding the possible justification and explanation as to why those hate types were so high within those three months, the exploration of those drivers, trends and tools that were reported by each of the project partners will be observed in the first place to enable a better understanding of the phenomena.

II. Drivers, trends and tools

1. Drivers

Regarding the **drivers** that emerged within those three months, in Germany, jugendschutz.net reported the following; firstly, they detected that the European Championship led to many discussions within the far-right movement. This was the case especially regarding the multiculturalism of the team, with for example Mesut Özil visiting Mecca, and the fact that the players did not sing along to the national anthem which seemed problematic for many. The concept of a "national" team was – in accordance with the general discussions about the decline of the "real" German people – considered as a failure. The Nice attacks on the 14th of July were also used by the right-wing spectrum. Therefore, topics such as migration, open borders and multiculturalism were used as scapegoats and those who engaged in a migrant-friendly culture were defamed as "do-gooders". Moreover, July was marked by further violent and deadly events that took place in Germany such as the one of the teenage refugee who attacked people with an axe on a train in Würzburg, a German-Iranian teenager who killed and harmed several people in a shooting in Munich, a Syrian refugee who killed a woman with a machete and injured several others in Reutlingen, and another Syrian refugee who committed a suicide bombing at the entrance of a music festival in Ansbach. In social media, these attacks were interpreted as the arrival of Islamic terrorism in Germany.

In France, Licra denoted that a social conflict emerged through the presence of many demonstrations, violence between demonstrators and police, affirmation of counter-government

movement, with the example of the "Nuit Debout", and strikes in public transports and refineries. These social and economic tensions created a general background conducive to extremism and hate-speech on the internet. Another social driver was the context of the pre-presidential elections of April and May 2017. There was a polarized political debate on the ideas presented by right wing groups about issues such as immigration, integration, liberalism, "laïcité" (secularism), and the welfare system. This debate led to an increase in extremist and radical political statements, and Twitter was used for that purpose extensively. This phenomenon was linked to another long-term situation being the consequences of the terrorist attacks in Paris. The prolongation of the state of emergency and the multiplication of the police presence created an atmosphere of tension. Governmental interventions did not reassure the public as the European Championships started in June, even though security was at the heart of the public policies. Licra also noted that, similarly to the case of Germany, the Nice attack made matters much worse, as it happened only 9 months after the attacks of November, and a little over a year after the Charlie Hebdo and Hyper Casher attacks, conclusively weakening the already unstable French Republic. This situation was then exploited by the Front National, a far-right populist political party.

According to MCI, in Spain, cyber hate related to the refugee crisis attenuated. However, after the Orlando attack some drivers emerged based on the idea that Islam was a homophobic religion that could not coexist in a diverse democracy. Moreover, only weeks after the attack there was an increase in homophobic content on Twitter after an Argentinian teenager started publishing extremist opinions on that social media platform. The impact of the Nice attack was, again, observed by MCI as well. Lastly, Brexit had an impact on Spanish politics as it was used by political parties in the midst of the general elections to instigate heated debates, in turn leading to online hate.

In Belgium, Unia observed that the refugee crisis and terrorist attacks had been drivers in the past but had slowed down in those months. The attacks in Orlando led to an increase in online hate speech against the gay community and a small rise in the number of complaints. Unia also received a number of complaints concerning (offline) hate speech about the putsch in Turkey.

In Austria, ZARA also examined that new drivers such as the Nice attack, attacks in Germany, and the attempted coup in Turkey (which led to several demonstrations and extended news coverage on anti-democratic development in Turkey), strengthened already existing anti-Muslim, anti-Turkish and generally racist resentments, therefore triggering an increase in online hate speech. The media also reported on real and fictitious offenses, especially sexual offenses, committed by asylum seekers in Austria. Right-wing politicians spread propaganda about the ongoing migrant/Muslims "invasion" of Austria which was surrounded by many hateful comments against refugees and Muslims in general.

2.Trends

Regarding the new **trends**, jugendschutz.net observed that the traditional political activities on the 1st of May were used by right-wing extremist groups to organize demonstrations, which were announced online. The success of these actions was celebrated afterwards on social media as part of self-promotion. Furthermore, the German Minister of Justice, Heiko Maas, became a prominent victim of anti-democratic resentments as right-wing protesters disturbed a public talk in Zwickau, which was also presented as another triumph against the government. Maas, one of the greatest enemies of the extremist scene, was mocked and defamed on respective platforms. Moreover, "Anonymous" (Anonymous.Kollektiv), one of the most popular Facebook profiles in terms of spreading cyber hate, fake news and conspiracy theories, called for buying weapons in order to prepare for the inevitable "bang" caused by the "import" of criminals. "Anon" posted a test report about how to acquire weapons easily. The original post was linked from VK to Facebook and in turn linked to an online shop called www.migrantenschreck.ch ("migrant shock"), currently under the URL http://www.migrantenschreck.ru. This site was registered by the supposed operator of the "Anon" profiles, Mario Rönsch. Meanwhile, weapons by the name of "Migrantenschreck" were tested on Anonymous' YouTube channel. On a video a masked person was shown aiming at pictures of German politicians. The comments showed a lot of affirmation that supported the invoked scenario of a civil war. Furthermore, a new join-in activity of the Identitarian movement in the context of their campaign against "migrant violence" attracted attention. Followers were called to paste over stumbling blocks (reminders of Holocaust victims). Although the number of participant was low, the activists achieved great response on social media. Regarding the attacks in Orlando, jugendschutz.net noticed that it was also exploited for hate speech against Muslims. Moreover, relating to the attacks in Germany, many propagators of cyber hate used "alternative" media and fake news. Respective users also blamed the political system which had to be changed radically in order to save the German people.

Licra noticed that there was an affirmation of the right-wing power in France with the cancellation of Black M's concert for the commemoration of Verdun, a development of new extremist program (for example "Oztadroite"), the banalization of racist and extremist statements from politicians, and not only from the Front National (as for example Nadine Morano, a politician from the right political party, stated in "Les Républicains" that the "Gare du Nord", a train station in Paris, was like "Africa"). There was also much confusion between being Muslim, Arab and jihadist. Indeed, the consequences of ISIS terrorism in France were terrible for the Muslim communities. Besides, the development of anti-Israel feelings linked to antisemitism derived from the anti-Zionist movement (e.g. links between ISIS and Israel). Further, an increase in racism between French people from Sub-Saharan Africa and French people from Maghreb arose. This type of racism could also be linked to a misogynist vision, with the use of the word "beurettes", a derogatory term for young women from Maghreb. Lastly, Licra observed that since the beginning of the summer, the platform Google+ was increasingly used for the promotion of antisemitic and anti-Israeli speeches, regardless of the lack of popularity of this social media platform in France.

MCI noticed that cyber hate against Muslims was the principal trend on Twitter, Facebook, blogs and other social media platforms. There were many talks about the "glorification of terrorism". In addition, neo-Nazis launched a campaign in solidarity with Pedro Varela, the editor of a Barcelona based Europe Library, well known for distributing all kinds of hate material, regarding holocaust denial and Nazi glorification. This campaign was a response to the prosecution of Varela and the new attempt to definitely close its library. It was mainly led by Alianza Nacional, one of the most relevant organization in the neo-Nazi Spanish scene.

ZARA reported that in May 2016, as the final round of Austria's presidential election took place, hateful comments increasingly appeared in online discussions about the election and politics. In addition, death threats against the newly elected president was posted on the FPÖ leader's Facebook page.

MDI observed that the coup in Turkey also affected hate online. The political issues of July in Turkey had particular influence in the Netherlands because of the big Turkish community in the country. The hatred and discrimination between the Gülen and Erdogan followers was seen on forums, YouTube and other platforms. Those discussions saw discrimination towards Turkish people, Dutch people and criticism about Dutch policy towards Turkey.

3. Tools

Regarding the new **tools**, jugendschutz.net observed that fake news was often used by extremists in Germany to incite hatred. The story of the knife attack at the train station in Grafing (near Munich), for instance, spread fast via social media. Although it was clear shortly after the attack that there was no terrorist background, innumerable hoaxes could be found on the internet. Even after the investigative authority declared that the perpetrator was an autochthonous German who was known to the police as mentally confused, right-wing extremists carried on propagating that this was a misinformation campaign. Furthermore, Marco Delgardo, a popular former music producer, claimed on his blog that the perpetrators name was Rafik Youssef. Despite the obvious falsehood of the statement, the post was shared several thousand times and taken up from "alternative media" like KOPP-Verlag and COMPACT-Magazin.

Licra looked at the story of the "so-called" fake antisemitic Twitter account of Djamel Boumaaz, an ex-member of the Front National and a current municipal councillor in Montpellier. He had created a Twitter account where he had published antisemitic tweets and statements. In his Twitter account's description, he mentioned anti-Jewish legislation of the 30's and the 40's amd "dogs and Jews prohibited" laws. Djamel Boumaaz had already been criticized because he had removed the gay flag from the city hall of Montpellier on the 18th of May, the day against homophobia. The account was suspended and, later, Djamel Boumaaz affirmed that this account was false and filed a complaint for identity theft. There was also the case of Serge Aurier, where a French footballer

criticized the French football coach calling him "queer". Or the case of the SFR employees who promoted antisemitic and misogynist statements in a live stream video which they published on a Periscope channel named "Petit Poney" ("Little Pony") in reference to Dieudonné's shows. On their video, one of the SFR employees calls a customer a "Jewish bastard". The other employee explained "it comes from the deep of the asshole of the heart". A few seconds later, they both call a young woman walking in the street a "half-bitch, half-submissive" (in reference to the French feminist NGO "Nor-bitch, nor-submissive"). Licra's community manager saw the video on Periscope relatively quickly and immediately tweeted about this case asking SFR to react. On another matter, there were conspiracy theories related to the European Championship of 2016. Benzema and (other football players of Muslim origin) had supposedly been rejected from the French football team because the Jewish lobby refused the Muslims to be part of it. Further, there was a conspiracy theory related to the Brexit. Jews were supposedly behind Brexit in order to destroy the White European Civilization. Finally, fake photos appeared, such as the one of a presumed refugee wearing a T-shirt "Fear for your wife", shared by the "Fachosphere" with 200 retweets in one day. Actually, the photo had been taken in Australia in 2013.
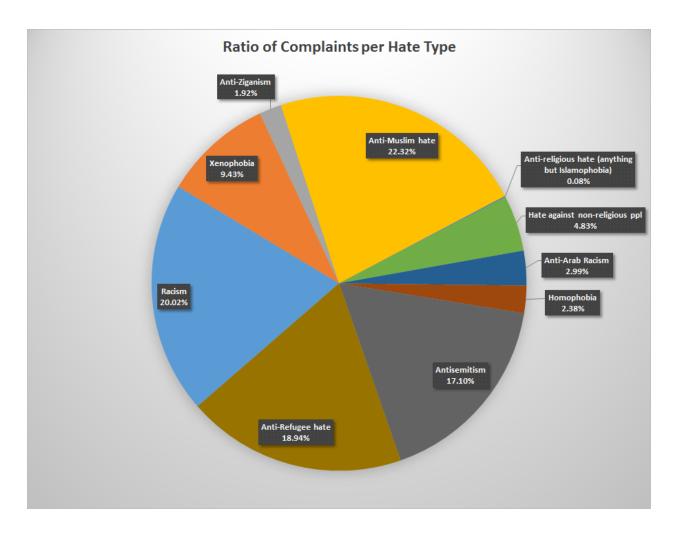
MCI also found that fake rumours and news were being spread via WhatsApp.

Finally, Unia observed a return of chain emails in June used to promulgate online hate. Chain emails were a relatively frequently used tool in the past, but had somehow disappeared.

III. Data Collection and Analysis

1.Hate Type Analysis

Now that a background information about drivers, trends and tools was outlined, it is possible to move on to the data collection and analysis part of this report, with a better understanding of the general atmosphere of the situation in Europe. During the monthly data collection INACH put particular focus on 10 different hate types, due to their prevalence and pervasiveness on the internet. These hate types are the following: racism, xenophobia, anti-Ziganism (hate against the Roma community), anti-Muslim hate (ie. Islamophobia), anti-religious hate (everything but Islamophobia), hate against non-religious people, anti-Arab racism, homophobia, antisemitism and finally anti-refugee hate.

Ratio of Complaints per Hate Type

- Anti-Ziganism 1.92%
- Anti-Muslim hate 22.32%
- Xenophobia 9.43%
- Anti-religious hate (anything but Islamophobia) 0.08%
- Hate against non-religious ppl 4.83%
- Anti-Arab Racism 2.99%
- Racism 20.02%
- Homophobia 2.38%
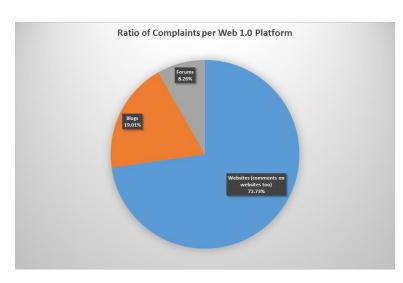- Antisemitism 17.10%
- Anti-Refugee hate 18.94%

These hate types fluctuate immensely from month to month. Some hate types are very prevalent in some countries, while they are scarce in others. Also, the differences between INACH's project partners add to this variation. Licra, in France, mainly focuses on antisemitism, for instance, therefore they always deliver a high number of cyber hate cases against the French Jewry. Other partners focus more on anti-Muslim hate or other types of racism, so their numbers tend to be higher in different hate types. The last factor affecting the numbers is the difference in size and funding amongst the project partners. Jugendschutz.net is a major organization in Germany with a lot more manpower and resources than, for instance, ZARA in Austria or MCI in Spain. Hence, the number of cases we receive from Germany tend to be a lot higher than from other countries where our project partners reside. However, altogether, the numbers received from all partners give a fairly extensive and wide insight into cyber hate in Europe.

The collected numbers in the second quarter of 2016 (from May to July) show that anti-Muslim hate was the most prevalent online and on social media sites with a whopping 22.32 per cent; i.e. more than a fifth of all recorded cases targeted the Muslim community. This was followed closely by general racism at a 20.02 per cent (due to our methodology, antisemitism, anti-Arab racism, anti-Ziganism and anti-refugee hate are all excluded from this number). This was followed by anti-

refugee hate (any kind of cyber hate that attacks people solely based on the fact that they are refugees or migrants) at 18.94 per cent. This hate type went through a large increase between May and June 2016, when it jumped from about 10 per cent to around 20 per cent and it has been around that level ever since. The fourth overall highest was antisemitism with its 17.10 per cent, followed by xenophobia at 8.15 per cent. All other hate types are between 2 and 4 per cent.

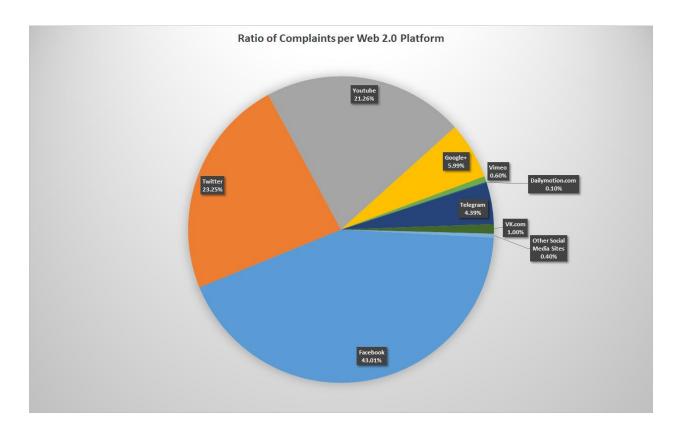## 2. The Prevalence of Cyber Hate on Different Platforms

When it comes to platforms where cyber hate is flourishing, it might not be surprising that social media trumps Web 1.0 platforms by a magnitude. However, there are still websites, blogs and forums on the internet that spew hate or provide a platform for people to post hateful messages.
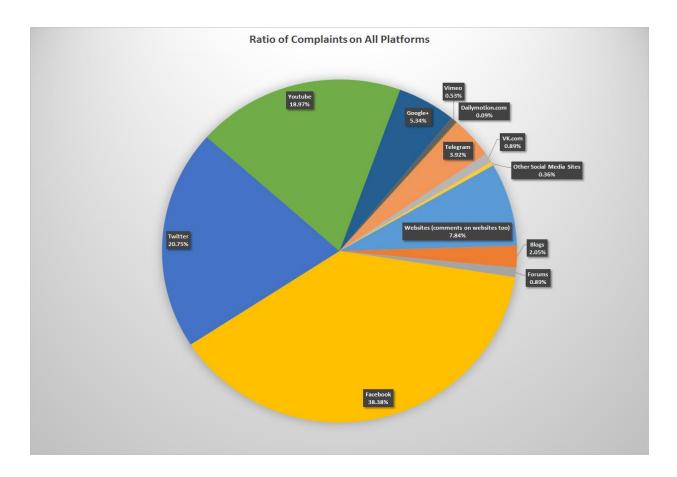
Websites are the most widely used platforms among Web 1.0 platforms to spread cyber hate, either by producing it or by providing a platform for people to post vile, violent and hateful comments. During the second quarter of 2016, 72.73 per cent of recorded cases of online hate (that appeared outside of social media) were posted on websites, followed by blogs at 19.01 per cent and forums at 8.26 per cent.

When examining social media sites, three giants emerge somewhat unsurprisingly. These are Facebook (43.01 per cent), Twitter (23.25 per cent) and Youtube (21.26 per cent); followed by Google+ at a meagre 5.99 per cent. Comparatively, all other social media platforms are not even in the ballpark, representing between basically 0 and 5 per cent of all collected cases.

Ratio of Complaints per Web 2.0 Platform

Youtube
21.26%

Google+
5.99%

Vimeo
0.60%

Dailymotion.com
0.10%

Twitter
23.25%

Telegram
4.39%

VK.com
1.00%

Other Social
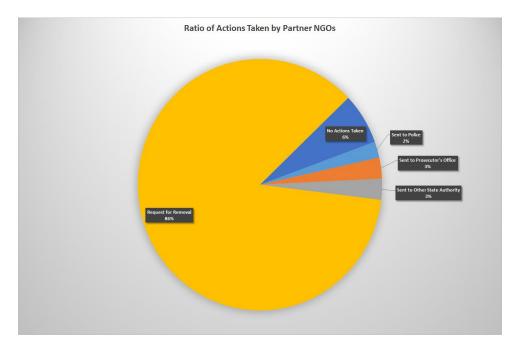Media Sites
0.40%

Facebook
43.01%

If the numbers of Web 1.0 and 2.0 platforms are merged, it becomes crystal clear that social media sites are the pivotal platforms when it comes to the spreading of cyber hate. These platforms provide a cheap or even free tool for people and extremist groups to deliver their message to a vast audience. Hence, the dominance of the three aforementioned giants remains intact in the same order previously described, but the ratio of cases recorded on websites falls to a meagre 7.84 per cent and the ratio of cases on blogs and forums falls to around or below 2 per cent.

Ratio of Complaints on All Platforms

These numbers clearly show that social media sites have completely taken over and fundamentally changed the landscape of cyber hate by letting their users spewing hateful and violent content against minority communities in the form of memes, conspiracy theories, fake news and other viral content. Even more alarmingly, these platforms made it possible for extremist groups and individuals to deliver such content to users who do not actively seek it out, paving the way for radicalization among adolescents and young adults.

3. Actions Taken by Partner Organisations Against Instances of Cyber Hate

Partner organizations that participate in the project mainly focus on getting instances of cyber hate removed from social media and other platforms. Therefore, it is not surprising that, among the reported actions that had been taken by our partners, request for removal is the unquestionable leader with 86 per cent.

Ratio of Actions Taken by Partner NGOs

This is followed by cases where no actions were taken (6 per cent). These are cases that were reported to or discovered by INACH partners, but - in the end - were deemed to fall outside of the realm of (illegal) hate speech. Finally, sometimes INACH discovers hate speech online that is so serious that it is not enough to just report it to the platform where it had been posted, but the case has to be reported to state authorities too. This can be the police, the prosecutor's office or any other law enforcement agency. Altogether, cases forwarded to these authorities counted for 8 per cent of all cases in the second quarter of 2016.

4. Removal Rate

Removal rates can be very varied and inconsistent when it comes to the three big social media platforms. INACH's project partners received most of their complaints on Facebook, followed by Twitter and YouTube. The fourth highest number of complaints was received on Google +, but that number is dwarfed by the aforementioned triumvirate.

The removal rate of Facebook was fairly high on average. Between May and July 2016, across all six partner countries, the platform's removal rate was 78.57 per cent. Twitter on the other hand was doing a lot worse, with a removal rate as low as 67.01 per cent in the same time period. Youtube's removal rate was very close to Facebook's with an exceptionally high 86.27 per cent, which is unusual even for them. Google+ is used by a lot less people than the previous three platforms, and the number of complaints on the platform is a lot lower, but these are still not sufficient excuses for the very low removal rate by the site, which was a meagre 30 per cent.

Name of Platform	Percentages of Cases Removed	Percentages of Cases Not Removed
Websites (comments on websites too)	48.57%	51.43%
Blogs	50.00%	50.00%
Forums	50.00%	50.00%
Facebook	78.57%	21.43%
Twitter	67.01%	32.99%
Youtube	86.27%	13.73%
Google+	30.00%	70.00%
Vimeo	33.33%	66.67%
Dailymotion.com	100.00%	0.00%
VK.com	66.67%	33.33%
Telegram	20.45%	79.55%

If one takes a look at the numbers above, they might think that everything is alright when it comes to the removal of cyber hate. On average, Facebook and YouTube do have quite high removal rates, and Twitter's 66 per cent arguably falls into the tolerable category too. However, if we take a look at individual removal rates in different partner countries in different months, we can see the biggest problem NGOs that fight cyber hate have with these sites. They are outrageously inconsistent in their removal rates between countries and in cases that are very similar to each other. It is understandable that these companies' community guidelines are interpreted in relation to given countries national laws, but the guidelines are the same globally, therefore, the same infractions should be removed everywhere. However, that is most definitely not the case. Removal rates are highly influenced by the amount of complaints given social media site receives about an instance of online hate, and by who the complainer is. If it is an authority or a very well established local NGO, or other civil society organization that is a trusted reporter or flagger, it is much more likely that the hateful content will be removed; just like when a lot of people complain about a certain content. This should not be the case. Illegal content and content that violates the guidelines should be removed globally and universally, irrespectively of the number of complainers or who the flagger is.

Taking all this into account, it is very aggravating that removal rates variegate vastly between countries. For example, jugendschutz.net in Germany had a 100 per cent removal rate on Facebook
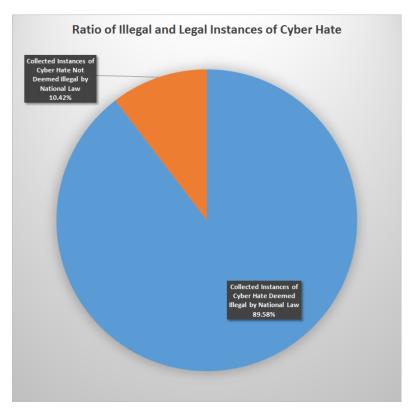
in May, whilst ZARA in Austria only had 16.67 per cent success rate in removals and Licra in France did not manage to have anything removed by Facebook in that month.

There are very similar problems with Twitter. jugendschutz.net only had a 40 per cent removal rate on Twitter in June, whilst Licra had an 85.29 per cent success rate.

In July, removal rates for Facebook were 81.3 per cent in Germany, 6.25 per cent in France and 37.5 per cent in Austria. The list could be continued, but the point is already clear. There are major differences in removal rates on a monthly basis and between countries. This insinuates that social media companies interpret their own rules and guidelines subjectively and arbitrarily. This arbitrariness makes the job of NGOs and other organizations extremely hard and frustrating, whilst it also nurtures an enabling culture online towards extremist groups and people who hold extreme ideas and ideologies. Highly illegal, violent, hateful and vile contents are left online for months without any real explanation from social media giants, whilst minor and benign infractions are removed within hours. This attitude and the companies' modus operandi must change, if we are ever to have an online community that respects the human rights of all of its members.

5. Legality of Instances of Cyber Hate

The legality aspect of the project is an interesting one which should be explored further. What is mainly noted is that, although some cases might be considered hate speech by the public or by INACH members, they might not always be considered illegal. Every month, from May to July many cases collected represented instances of cyber hate but were not deemed illegal by national law. In the second quarter of 2016, 89.58 per cent of reported or discovered instances of cyber hate were deemed illegal by the complaints officers of our partner organizations.

Ratio of Illegal and Legal Instances of Cyber Hate

Collected Instances of Cyber Hate Not Deemed Illegal by National Law 10.42%

Collected Instances of Cyber Hate Deemed Illegal by National Law 89.58%

This means that 10.42 per cent of cases assessed by our officers fell into a murky field, in which the inspected speech is highly offensive, dangerous, demeaning and/or goes against human dignity,

yet it does not fall into what given nation state considers illegal hate speech. Even though this ratio is not very high, EU member states should pay more attention to hate speech that falls through the cracks of legislation in order to be able to stand up against it even more effectively.

IV. End Remarks

Reaching extensive conclusions based on numbers collected in the first three months of our data collecting period would be fallacious and premature. Trends cannot be drawn up based on such a relatively small sample size. Therefore, we will discuss trends, shifts in the data and the conclusions that can be drawn from them in our extensive and comprehensive yearly report that we will publish in late 2017.