



Annual Report

July 2019 – April 2020

```
348 font-size: 13px;
349 }
350
351
352 /* =Menu
353 -----
354
355 #access {
356   display: inline-block;
357   height: 69px;
358   float: right;
359   margin: 11px 28px;
360   max-width: 800px;
361 }
362
363 #access ul {
364   font-size: 13px;
365   list-style: none;
366   margin: 0 0 0 -0.5em;
367   padding-left: 0;
368   z-index: 99999;
369   text-align: right;
370 }
371
372 .access li {
373   display: inline-block;
374   text-align: left;
375 }
```



*Platforms, Experts, Tools:
Specialised Cyber-Activists Network*

Annual Report

July 2019 – April 2020



*Project funded by the European Union's Rights,
Equality and Citizenship Programme (2014-2020)*

About the project

The EU-funded project sCAN – Platforms, Experts, Tools: Specialised Cyber-Activists Network (2018-2020), coordinated by Licra (International League Against Racism and Antisemitism), aims at gathering expertise, tools, methodology and knowledge on cyber hate and developing transnational comprehensive practices for identifying, analysing, reporting and counter-acting online hate speech. This project draws on the results of successful European projects already realised, for example the project “Research, Report, Remove: Countering Cyber-Hate phenomena” and “Facing Facts”, and strives to continue, emphasize and strengthen the initiatives developed by civil society for counteracting hate speech.

Through cross-European cooperation, the project partners are enhancing and (further) intensifying their fruitful collaboration. The sCAN project partners are contributing to selecting and providing relevant automated monitoring tools to improve the detection of hateful content. Another key aspect of sCAN is the strengthening of the monitoring actions (e.g. the monitoring exercises) set up by the European Commission. The project partners are also jointly gathering knowledge and findings to better identify, explain and understand trends of cyber hate at a transnational level. Furthermore, this project aims to develop cross-European capacity by providing e-learning courses for cyber-activists, moderators and tutors through the Facing Facts Online platform.

sCAN is implemented by ten different European partners, namely ZARA – Zivilcourage und Anti-Rassismus-Arbeit from Austria, CEJI – A Jewish contribution to an inclusive Europe from Belgium, Human Rights House Zagreb from Croatia, Romea from Czech Republic, Licra – International League Against Racism and Antisemitism from France, Respect Zone from France, jugendschutz.net from Germany, CESIE from Italy, Latvian Centre For Human Rights from Latvia and the University of Ljubljana, Faculty of Social Sciences from Slovenia.

The sCAN project is funded by the European Commission Directorate – General for Justice and Consumers, within the framework of the Rights, Equality and Citizenship (REC) Programme of the European Union.

Legal Disclaimer

This Annual Report was funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). The content of the Annual Report represents the views of the authors only and is the sole responsibility of the sCAN project consortium. The European Commission does not accept any responsibility for use that may be made of the information it contains.

Content

About the project	3
Introduction	6
Tools and Resources	8
<i>sCAN Testing exercise</i>	9
<i>An integration of artificial intelligence</i>	10
<i>User Guide for using automated technologies in monitoring hate speech content</i>	11
Research	12
<i>Analytical Paper “Hot Spots of Hate”</i>	13
<i>Analytical Paper “Intersectional Hate Speech Online”</i>	14
<i>Analytical Paper “Hate speech and pandemic in the age of Internet”</i>	16
The sCAN Monitoring Exercises	17
Education	22
<i>Online course on Hate Speech</i>	22
<i>Online Course for Moderators</i>	24
<i>Advanced Monitoring Training</i>	24
Outlook and Policy Recommendations	26
Resources and further reading	29
<i>sCAN Project resources</i>	29
<i>References</i>	30

Introduction

The internet is an integral part of everyday communication worldwide. While it is most often used in a peaceful manner to communicate with friends or freely express ones' opinion on a diverse range of topics, some users spread hatred and incite to violence against disadvantaged minorities.

In recent years, several European projects countering hate speech have already been successfully implemented. To strengthen European networking and to harness synergies between different projects' results, the sCAN project closely cooperates with the International Network Against Cyber Hate (INACH) and the Facing Facts! project.

The project partners have agreed on INACH's definition of hate speech:

"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" ¹

During the second year of the project implementation (June 2019 – April 2020), the partners continued their multidimensional approach of pursuing technological solutions, research, monitoring and education. The project conducted testing exercises to evaluate the effectiveness of selected online crawlers and artificial intelligence to facilitate monitoring of online hate speech.

¹ International Network Against Cyber Hate (2018). What is cyber hate. Available at <http://www.inach.net/wp-content/uploads/WHAT-IS-CYBER-HATE-update.pdf> (last accessed 07.04.2020).

Furthermore, the partners conducted joint research projects and published analytical reports on their findings. Firstly, the partners analysed hate speech disseminated or facilitated by public figures, such as politicians, journalists and social media influencers. In the second research project the partners analysed intersectional hate speech online in the different project countries. A third research project will analyse the impact of the Current Covid-19 pandemic on online hate speech.

Furthermore, the sCAN partners participated in two monitoring exercises, one with the European Commission and one with the International Network Against Cyber Hate (INACH) and the project Open Code for Hate-Free Communication (OpCode). The goal of the monitoring exercises was to evaluate the adherence of the IT companies Facebook, Twitter, YouTube and Instagram to the Code of Conduct on countering illegal hate speech online, developed in 2016 by the European Commission. The sCAN partners have already been participating in previous monitoring exercises organised by the European Commission and INACH.

Research and monitoring alone are, however, not sufficient to combat cyber hate. Therefore, the partners developed and conducted online courses and offline training workshops to build capacities of Civil Society Organisations (CSOs) and individual activists to counter hate speech in a diverse manner, through monitoring, engaging in counter speech or moderating online discussions.

All project outputs can be accessed on the project's website www.scan-project.eu. Direct links to the respective documents will be provided at the end of this report.



Tools and Resources

One of the main objectives of the sCAN project is to help and to provide new monitoring tools in order to facilitate research efforts to combat online hate speech.

This detection of online hate speech could be provided by several types of tools such as web spidering, crawlers, software and artificial intelligence. However, most of those tools are not easily accessible for CSOs and cyber-activists. Indeed, some conditions must be taken into account by organisations wishing to use automated monitoring tools: human resources, materials,

resources and language challenges and the existence of an IT desk or coding skills. In addition to these elements, it is crucial to understand that not all online platforms provide the same options regarding the detection of hate speech: their level of privacy has a direct impact on possibilities for using automated technologies. Another important observation is that automated monitoring tools should not be considered as the only efficient way to combat online hate speech. Human expertise in terms of knowledge, adaptation capacity and analysis skills are still crucial in monitoring hate speech.

During the first year of this project, the sCAN partners contributed to a dataset of keywords and key-expressions in all project languages, including additional information on the context in which those words are used in the respective national speeches. This research provided important insights into the nature of hate speech in the analysed countries. To complement these findings, a mapping study has been produced, identifying some of the available software solutions and tools to automatically monitor cyber hate. For CSOs, it is important to explore the use of automated software tools in monitoring hate speech. Nonetheless, some conditions need to be met: crawlers, software or artificial intelligence need to be supplied with a specific set of keywords taking into account the national context and patterns of hateful speech in each country. Other criteria are also crucial to ensure the integration of automated tools in monitoring hate speech by CSOs. Those include, i.a., the costs of using a certain tool or the technical skills required to operate it.

During the second year of the project, the testing exercise launched in September 2018 has been continued. Furthermore, the project has developed a partnership with Factmata,

a company specialised in artificial intelligence. Two sessions of testing campaigns have been scheduled along the project: the first one dedicated to crawlers and the second one to Artificial Intelligence.

sCAN Testing exercise

During the sCAN project, jugendschutz.net and Licra developed a common methodology for testing a selection of automated tools. The main objective was to provide an evaluation of the accuracy and relevance of the selected tools in order to integrate them in the SCAN consortium monitoring task. This testing exercise was organised in two different campaigns.

The first testing campaign was dedicated to crawlers at the beginning of the project, in September and October 2018. This two months long campaign was focused on testing several crawlers on websites, blogs and social media platforms by using relevant keywords selected for the hate ontologies report. The partners concentrated on the social media platforms that had signed the Code of Conduct on countering illegal hate speech online with the EU and were also included in the project's monitoring exercises.

During this first campaign, the following tools have been tested: TAGS v 6.1 on Twitter, HTTracks for web 1.0, SociScraper on Instagram and YouTube and CrowdTangle on Facebook. Each selected tool has been tested according to a common methodology including the use of a keywords list. The keywords used in the testing campaign were selected based on the hate ontology published during the first year of the project. In addition, a non-exhaustive list of criteria was agreed on to evaluate the selected tools: the price of the crawler/software, training, required skills, manual assistance, results regarding hate speech categories, time, bugs and issues, pros and cons, linguistic settings.

An integration of artificial intelligence

Artificial intelligence (AI) gains a foothold in detection of online hate speech. Therefore, the second testing campaign was focused on algorithms and AI. It was implemented from December 2018 until the end of the sSCAN project's implementation. As the development of AI requires resources not easily affordable for CSOs and Human rights cyber-activists,

a partnership was formed with the company Factmata.

Factmata works on a technology combining artificial intelligence, algorithm and expert knowledge to deal with hate speech and fake news. Based in London, Factmata proposes an anti-fake-news artificial intelligence platform (API) and services by providing a scoring system for the content across the web. Regarding hate speech content, the Factmata API scores content according following criteria: "insult", "obscenity", "toxicity", "stereotyping", and "threat", "identity hate" as well as "sexism" and against "any particular gender". To refine their algorithm, the start-up requires human back-stopping in order to improve hate speech detection and results. jugendschutz.net and Licra have contributed as part of the community of users to fact-check content for quality with the help of their artificial intelligence tools.

jugendschutz.net, Licra and Factmata have launched their partnership in November 2018. During several testing sessions of the AI, jugendschutz.net and Licra have participated in annotation sessions regarding criteria of hate speech, threats, insults and obscenity. The tool has been trained in a North American context. As a consequence,

both European organisations provided a European experience regarding how to define hate speech: they contributed to integrate country-specific and European transnational criteria of hate speech trends and as consequence to improve the model for detecting hateful content. Due to the very quick evolution of hate speech vocabulary, this methodology strives to address language and evolution issues.

For the second testing campaign, the test was conducted with English content only, as the algorithm was not trained on other languages. Nonetheless, thanks to the partnership developed with Factmata, it could be possible to train the tool in other languages.

User Guide for using automated technologies in monitoring hate speech content

The results of these testing campaigns have been analysed in order to produce a comprehensive user guide on automated monitoring tools. In this guide, the sCAN consortium aims to explain how to use available tools to improve hate speech monitoring and removal.

The User Guide strives to provide guidelines on available and inexpensive tools with a comprehensive interface in order to improve the monitoring process and collection of data. For each tool, guidelines are presented following the same reasoning: presentation of the tool; conditions for using the tool; illustrated steps for using the tool; and advantages and disadvantages of the tool.

Additionally, two webinars sessions on the content of the user guide were provided for the consortium in order to facilitate the use and the integration of the selected automated tools. Furthermore, an internal online tutorial on how to use the INACH database during the 2nd monitoring exercise was organised in May 2019.



Research

Throughout the sCAN project partners dedicated themselves to drawing up analytical papers with a focus on current as well as controversial topics of utmost importance. Those topics were chosen based on the consortium's expertise and experiences and produced with the aim to share knowledge and provide an overview on important trends and developments of the phenomenon of online hate speech. In the period between

July 2019 to April 2020 the sCAN consortium published two analytical papers.

Since the scope of the sCAN project did not allow for the implementation of extensive qualitative and/or quantitative analyses, the consortium decided to focus on exemplary case studies to provide a thorough overview and understanding of the phenomena in discussion.

Analytical Paper

"Hot Spots of Hate"

The third of overall four analytical papers throughout the project was dedicated to the topic "Hotspots of Hate – the online responsibility of public figures". This paper was based on the experience awareness of the project partners that public figures such as politicians, journalists and online influencers can have a strong influence on their followers on social media and in the online sphere. When taking a closer look at the case studies, the consortium found that such "influencers" often communicate continuously with their followers via social media and therefore have the possibility to considerably shape their perceptions. With their enormous reach and the quantity of content, social media are the perfect tool to influence public opinion. Their wide reach is the reason social media bear special responsibility when it comes to spreading disinformation or implicit (or sometimes explicit) incitement to hatred. In several European countries, leading politicians and other public figures use their online presence to incite hatred or to encourage hate speech by posting biased and populist comments to their social media profiles. Even if the posts

themselves do not constitute illegal hate speech, they incite hatred and stimulate hate speech in the comment sections. It can be challenging to counter such instances of triggering hate speech while upholding freedom of expression. If the original post remains online, it is prone to attracting further hateful comments.

By analysing case studies from all countries involved in the project, it has become quite clear that social media companies need to examine these instances closely and start exploring ways for addressing it on their massive scales. Incitement works like a single match causing an entire forest to catch fire, therefore hotspots of online hate need special attention from social media companies: they cannot be ignored, nor underestimated. Furthermore, politicians and news outlets need to take their responsibility to moderate comments on their profiles and channels more seriously. Additionally, users can challenge hateful expressions with counter speech, by deconstructing hateful stereotypes and debunking fake news and manipulation. They can also opt for expressing solidarity and support for targeted individuals and communities.

Analytical Paper

“Intersectional Hate Speech Online”

The fourth analytical paper was dedicated to the topic of “Intersectional Hate Speech Online”. The concept of intersectional discrimination has its origin in the movement of black feminism and the term “intersectionality” was coined by Kimberlé Crenshaw. All the organisations involved in the sCAN project perceive the phenomenon of intersectionality as a constant challenge in the analogous as well as online sphere. The sCAN organisations took a closer look at intersectionality and followed the suggestion of the Fundacion Secretariado Gitano² to analyse how a specific incident of discrimination would have been different if one of the intersecting characteristics had been absent.

By analysing a number of case studies, it was established that intersectional hate speech is common in all countries involved in the project (and probably

beyond). Overall, perceived women*, perceived LGBTIQ+ persons and/or persons affiliated or belonging to an ethnic and/or religious minority – based on a combination of their (legally protected) identity categories were identified as the most frequent target groups. Furthermore, people with visible characteristics as well as those in public positions were shown to be particularly affected by intersectional hate speech.

Government departments should enshrine the principle of intersectionality in all equality policies, to avoid overlooking the experiences of the groups most affected by intersectional discrimination. Furthermore, governments should put in place strong consultation mechanisms with a broad range of diverse stakeholders. Politicians and high-ranking officials of public authorities should firmly condemn hate speech and promote counter-speech with a specific focus on multiple discrimination and the phenomenon of intersectional hate online and its impact on those directly affected. Furthermore, all political parties should condemn discriminatory, inflammable or hateful speech, with a specific focus on multiple discrimination, and call on their members and followers to abstain from using hateful speech during election campaigns.

² Fundacion Secretariado Gitano (2019). Guide on intersectional discrimination – The case of Roma women, p. 6. Available at https://gitanos.org/upload/22/65/GUIDE_ON_INTERSECTIONAL_DISCRIMINATION_-_ROMA_WOMEN_-_FSG_33444_.pdf (last accessed 08.04.2020).

Using hate speech to create an atmosphere of intolerance and undesirability for one group in our society can trigger violent incidents. Therefore, law enforcement authorities should not only ensure adequate investigation of hate speech and other discriminatory incidents, but also be aware of and take into account aggravating factors such as e.g. skin colour, sexual orientation, sexual identity, gender identity, disabilities, age and religion. Law enforcement authorities should strengthen their cooperation with various (socially constructed) groups and communities in order to better understand how certain groups and communities are affected by hate speech, specifically intersectional hate speech.

Intersectional hate speech is even harder to classify and fight than hate speech targeting one actual or perceived characteristic. We tend to use counter-hate tools designed for one specific form of hate speech. When many forms of hate speech intersect, some expressions of hate speech might be underreported, simplified or even ignored. CSOs should therefore strengthen their efforts to report and counter intersectional hate speech.

Analytical Paper

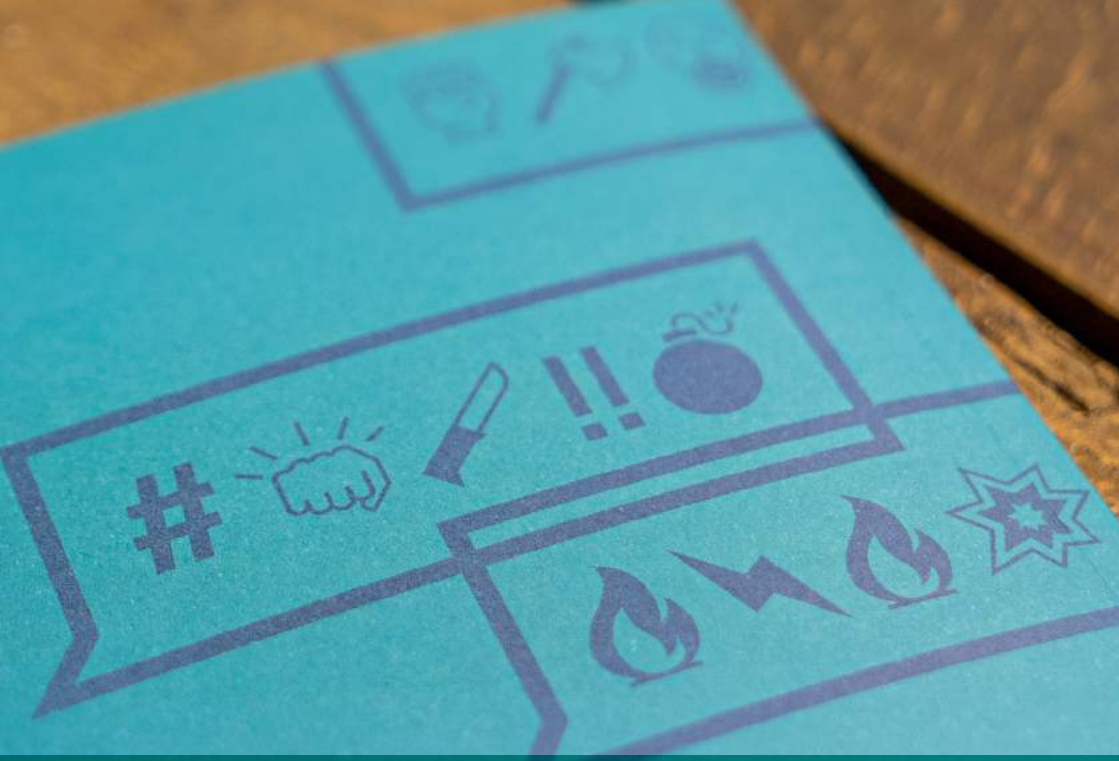
“Hate speech and pandemic in the age of Internet”

An additional analytical paper was dedicated to the topic “Hate speech and pandemic in the age of Internet”. The Covid-19 pandemic has disrupted human social, economic and political life conditions in all the EU. As the pandemic expands, online phenomena of conspiracy theories, rumours, fake news and hateful contents connected to this global disease are growing. Because of the lockdown procedures, people are likely to spend more time online, and discuss and interact through social media.

Online hateful speech found in the pandemic follows both traditional and possibly new patterns. The mechanisms of scapegoating and the spreading of rumours lead to a wide dissemination. This well-known social and psychological phenomenon has already been observed during previous dangerous global pandemic episodes, as for example for the Black Death during the Middle Ages. Almost every social, economic and health crisis may

cause the rise of conspiracy theories including hateful beliefs. Discovered plots grant access to “rationality” and “explainable phenomena”. Crises like pandemics may cause or deepen a divide in societies by the spread of rumours, conspiracy theories, accusations and as consequence violent acts against the “Other”.

For all these reasons, the consortium shared its experiences, its knowledge and some key-explanations about possible mechanisms of the interaction between hate speech and the rise of a global pandemic in the age of the Internet and social networks. This analytical paper aims at analysing online hateful trends on pandemic periods including a historical perspective. The main objectives of this report are to identify some of the events which resulted in an interaction between a pandemic and a rise of hateful speech and acts, in order to better explain and tackle today’s stereotypes and theories regarding this new health crisis our world is facing.



The sCAN Monitoring Exercises

During the second year of the project implementation, the sCAN partner organisations participated in two monitoring exercises, one with the European Commission and one with the International Network Against Cyber Hate (INACH) and the project Open Code for Hate-Free Communication (OpCode). The goal of the monitoring exercises was to assess Facebook's, Twitter's, YouTube's and Instagram's

compliance with the EU Code of Conduct on countering illegal hate speech online. The sCAN partners have already been participating in previous monitoring exercises organised by the European Commission and INACH.

The overall third sCAN monitoring was conducted during the monitoring exercise organised by the European Commission from November 4th to

December 13th 2019. During this six-week period, the sCAN partners reported 635 instances of illegal hate speech to the IT companies Facebook, Instagram, Twitter, YouTube, Dailymotion and Jeuxvideo. Facebook received the most reports from the sCAN partners (280 cases), followed by Twitter with 198 cases. YouTube received 102 reports about illegal hate speech and Instagram received 37 such reports from the sCAN partners.

84 cases were escalated through the channels available only to trusted flaggers of the IT companies, after not having been removed within a week after the initial report through general

user reporting channels. Twitter received 59 trusted flagger reports, Facebook and Instagram received 10 trusted flagger reports each and YouTube received 5 reports through trusted flagger channels. No cases were escalated to Dailymotion and Jeuxvideo.

Overall, 67,56 % of the content was no longer available at the end of the monitoring in the country it was reported from (64,25 % removed, 3,31 % restricted). This number is in line with the results of previous monitoring exercises conducted by the sCAN partners. The IT companies acted on 58,74 % of the cases directly after the

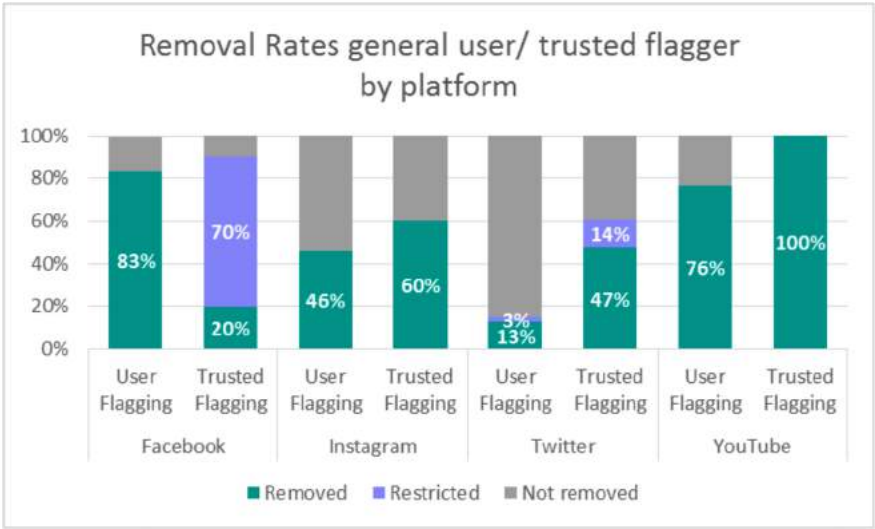


Figure 1: Removal Rates by platform; sCAN monitoring exercise 4th November – 13th December 2019

first reporting through normal user channels (57,80 % removed, 0,94 % restricted). Some partners escalated content that was not removed within a week after initial reporting by reporting it again through the channels available to trusted flaggers. The companies acted on 66,67% of the trusted flagger reports (48,81 % removed, 17,86 % restricted).

Jeuxvideo removed 100% of the cases reported to them through general user reporting channels within 24 hours. Dailymotion removed 33% of the cases reported to them within the 24 hours frame.

Facebook achieved the highest removal rate (83,21 %) for cases reported through general user reporting channels. YouTube removed 76 % of reported cases, Instagram 46 % and Twitter only took action in 16 % of cases by removing 13 % and restricting (geo-blocking) further 3 %.

All platforms performed considerably better for reports submitted through trusted flagger channels. YouTube removed 100 % of reports submitted by trusted flaggers. Facebook took action on 90 % of cases by restricting 70 % and removing 20 %. It is not clear to the project partners, why they chose to restrict such a large percentage of cases rather than removing them.

Instagram removed 60 % of cases reported by trusted flaggers. The most significant increase in action rate was seen for Twitter. The company took action on 61 % of cases (47 % removed, 14 % restricted), which is almost four times as much as the action taken on cases reported through channels available to their general user base.

The fourth sCAN monitoring took place between January 20th 2020 and February 28th 2020. It was an unannounced monitoring in cooperation with the INACH secretariat and the OpCode project. The sCAN partners reported 484 cases of illegal online hate speech to the IT companies Facebook (242 cases), Twitter (127), YouTube (66) and Instagram (49). In order to test the reaction of the IT companies to notifications by their general user base, the notifications were first sent anonymously through publicly available channels. In a second step, 94 cases that had not been removed after notification as general users were reported again through reporting channels available only for trusted flaggers.

Overall, only 58 % of the reported cases were no longer available at the end of the monitoring. This is a major drop compared to the 3rd sCAN monitoring exercise conducted only a

month earlier. It highlights the importance of a consistent case handling by the platforms, irrespective of official monitoring exercises organised by the European Commission.

51 % of the cases were already removed after the initial notifications as general users (normal user flagging). Instagram achieved the highest removal rate with 75,51 % of cases removed after notification through general user channels. Facebook removed 71,49 % of cases after initial reporting. YouTube and Twitter performed considerably poorer. YouTube removed 25,76 % of cases

after user notification, while Twitter only removed 9,45 % and restricted 4,72% of those cases.

94 cases were escalated through trusted flagger channels after not being removed by the companies when reported through general user notification channels. Out of those, 39 % were removed by the IT companies. Instagram removed all of the cases reported to them a second time through trusted flagger channels. Facebook removed 68,75 % of the cases reported by trusted flaggers. Twitter removed a considerably higher

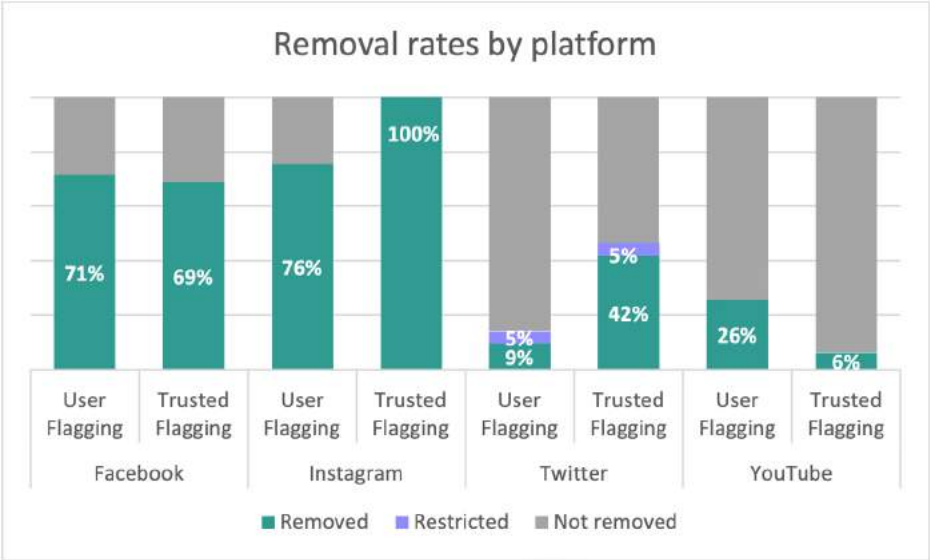


Figure 2: Removal Rates per platform; SCAN monitoring exercise 20th January – 28th February 2020

ratio of cases when they were reported through trusted flagger channels (41,86%) and restricted a further 4,65%, while YouTube removed less cases (6,45 %) than when they were reported by general users.

During the monitoring period, partners noticed several accounts posting large amounts of illegal hate speech comments and posts. Some of these pages or accounts have posted a significant number of racist, misogynist and extremely violent comments on a daily basis. Therefore, we recommend that the IT companies monitor those accounts more closely and take decisive action against every instance of illegal hate speech posted on them.

The results of these monitoring exercises highlight the need for a more consistent performance of IT companies in removing illegal hate speech online. The overall removal rate of 58% in the fourth monitoring during the implementation of the sCAN project is almost 10 % lower than the overall removal rate in the previous monitoring exercises. This includes the third sCAN monitoring exercise in November and December 2019, only one month prior. Companies must at all times ensure that they respond in a timely manner and remove illegal online hate speech.

Most companies provide more feedback to trusted flaggers than to their general user base. This can be problematic, as CSOs recognized as trusted flaggers cannot monitor and report all illegal hate speech by themselves. In the case of Instagram, the device used for reporting also seemed to have an impact on whether or not feedback was received. While partners reporting through the mobile app reported receiving feedback from the platform, partners reporting to Instagram using a desktop computer hardly received any feedback.

Involving all users of the platforms in reporting hate speech is crucial to combat illegal hate speech online effectively. Feedback is an important aspect to keeping users engaged and motivated to report, as well as to give them a better understanding of how the platforms moderate the content and enforce their community standards.



Education

Since monitoring and research alone are not sufficient to tackle online hate speech, the sCAN project developed educational training courses both online and offline. Those courses, already developed and implemented during the first project year, have also been conducted and constantly refined during the second year of the project.

The online courses focused on providing general knowledge about

hate speech, national, European and international hate speech legislation, monitoring of hate speech content online, counter-speech and moderation of online discussions.

Additionally, two offline advanced monitoring trainings were conducted in Vienna and Brussels. They included interactive sessions on how to recognise hate speech, the importance of monitoring and the art of documentation.

Online course on Hate Speech

One of the sCAN project activities allowed for the Facing Facts! Online Course on online hate speech to be translated into German and French and for adjusting it to the respective national contexts. This course was developed for anyone interested in combatting hate speech online according to their possibilities and capacities. This course offers new insights and practical approaches to effectively countering hate speech online for a broad range of people such as individual activists, members of communities, representatives of CSOs or authorities.

The course works with interactive methods and provides information on the concept of hate speech and how to recognize its nature and effects. One can learn about monitoring hate speech on the internet and how monitoring can be a tool to counteract the phenomenon. Active counter-speech, counter-campaigns and counter-narratives are also an issue and the course shows which of these strategies are most suitable for a “counter-activist’s” specific goals.

The German hate speech course, titled “Hate Speech – was tun?“, addresses the German and Austrian context of hate speech online. Between July 2019 and April 2020, the online course was offered three times. Each course was offered to a stable group of participants who were involved over a period of six to eight weeks. This approach aimed at activating the course participants and gave them the opportunity to intensively interact among each other and with the two online tutors of the course. One tutor was from the Austrian partner ZARA and the other one from the German partner jugendschutz.net. They moderated the online discussion forum and provided information on details and answers to emerging questions. The courses were complemented by online webinars with invited guest experts from “Gegen Vergessen – Für Demokratie e.V.” and the No Hate Speech Committees Austria and Germany. Overall, over 200 persons participated in the German version of the Facing Facts! Online course »Hate Speech – Was tun?«.

Online Course for Moderators

Building on the general online course on hate speech, the sCAN project developed an online course on moderation of hate speech online. This course is available in English and French on the platform Facing Facts Online!. It is directed at activists, community leaders, blogger, vlogger and practitioners interested in encouraging peaceful online exchanges, but any actor of interested in the subject can follow the course.

Through interactive tools, videos, dynamic lectures, case studies, testimonies, and quizzes, it addresses the question of how to react to hateful comments in online discussions. It aims to create a better understanding of the guiding principles of online moderation and the tools to support it.

To maintain healthy conversations online, the course discusses the variety of options for interventions, from removal to counter-speech, and it also encourages participants to create their own moderation policies based on the values they learn to articulate during the course.

Advanced Monitoring Training

An offline training on advanced monitoring and countering online hatred was developed and implemented throughout the project duration. Participants had the possibility to become experts in the field of monitoring hate speech and counteraction, documenting the phenomenon, tackling underreporting, comparing results when it comes to data acquired throughout different monitoring exercises and phases, as well as applying effective human rights reporting.

The trainings included interactive sessions on how to recognise hate speech, the importance of monitoring and the art of documentation. An expert trainer from the International Network Against Cyber Hate (INACH) additionally offered one hour morning sessions on how to use the INACH database for documenting cyber hate. The courses addressed individual activists and CSOs who plan to start their own monitoring of online hate speech or seek to professionalise already existing monitoring efforts.

The two trainings on advanced monitoring and countering online hatred, implemented throughout the second year of the project (July 2019 to April 2020) were held in Vienna (October 2019) and Brussels (March 2020). The trainings were held by ZARA training experts and the training in Brussels was coorganised by CEJI – A Jewish Contribution to an Inclusive Europe. 34 participants from 10 different European countries were given the opportunity to reflect on the phenomenon of online hatred in transnational groups, to build knowledge and expertise together, to gather best practice examples, and build strong alliances and networks in order

to broadly counteract online hatred. 70 persons participated in the overall four trainings implemented throughout the entire project implementation.

Furthermore, the participants had the possibility to get involved in self-sensitisation activities to understand and differentiate various forms of online hatred, discrimination and cyber mobbing.

To guarantee the sustainability of the knowledge on monitoring and countering online hatred generated within this project, ZARA produced a training manual to enable others to conduct trainings in this field.



Outlook and Policy Recommendations

During the past two years, the sCAN partners have worked together closely to analyse and monitor hate speech online and to develop online and offline trainings. We made our insights available to the larger public and contributed to building capacities in civil society to combat hate speech together. Through our activities we gathered valuable experiences and

collected ideas for improvement. All stakeholder groups are called upon to intensify their efforts to ensure an online environment that is respectful and inclusive for all users.

The project provided policy recommendations for the institutions of the European Union, national authorities and public institutions, politicians and

public figures, social media companies, media and journalists as well as CSOs and individual internet users on how to better combat all forms of hate speech online.

To the European Union, we recommend to encourage more social media companies to join the Code of Conduct on countering illegal hate speech online and to pay attention to smaller platforms that may be considered 'safe havens' for promoting intolerance and hate speech online. Additionally, we also recommend amending the methodology of the monitoring exercises to place more emphasis on groups and individual accounts that constantly disseminate hate speech to a significant audience or act as a catalyst for illegal hate speech.

Governments should design national action plans to combat hate speech and establish or refine their national data collection systems for hate speech, in order to ensure effective records of criminal as well as misdemeanour offences.

We urge politicians and other public figures to establish a political social responsibility and to refrain from spreading or facilitating hate speech online. All political parties should condemn hate speech and call on their members and followers to abstain from spreading online hate, including during election campaigns.

Social media companies should take greater efforts to enforce their community guidelines effectively and encourage respectful online communication. Due to its large impact on society, hate speech disseminated by politicians or other public figures should be clearly labelled as such and sanctioned according to the companies' community standards.

Media should take care to provide unbiased reporting about disadvantaged communities. Media outlets should reinforce cooperation with CSOs working in the field of human rights protection and representatives of disadvantaged communities in order to raise awareness among journalists of the stereotypes and hate speech narratives these communities commonly face online.

While most CSOs tend to focus on specific types of hate speech such as racism, antisemitism or anti-Muslim hate speech, it is important to also include other types of hate speech (e.g., misogyny, homophobia, transphobia, interphobia, ableism, ageism) in their analyses and action campaigns. Furthermore, CSOs should strengthen their efforts to report and counter inter-sectional hate speech.

All internet users can help curb hate speech by showing solidarity with the people and communities targeted by online hate, challenging hateful expressions with counter-speech, deconstructing hateful stereotypes and debunking fake news and manipulation.

Apart from the recommendations to specific stakeholder groups, closer cooperation between CSOs, members of affected communities, the media, the internet industry and public authorities is necessary to effectively curb the spread of hate speech online. As

the internet is not limited by national borders, more transnational cooperation is needed among all stakeholder groups to find a joint approach to this problem.

The sCAN Partnership will take stock of the lessons learned and its research findings of the past two years with the aim of planning follow-up initiatives aiming at improving and increasing its contribution to the monitoring, analysis, training and advocacy efforts carried out against all forms of online hate speech.

Resources and further reading

sCAN Project resources

All project results are available on the project's blog: www.scan-project.eu

sCAN Annual Report May 2018 – June 2019:

http://scan-project.eu/wp-content/uploads/sCAN_monitoring_report_year_1.pdf

sCAN Hate Ontology:

<http://scan-project.eu/wp-content/uploads/scan-hate-ontology.pdf>

User Guide on Monitoring Software:

<http://scan-project.eu/wp-content/uploads/sCAN-project-Online-User-Guide.pdf>

Analytical Paper “Antigypsyism on the Internet”:

<http://scan-project.eu/wp-content/uploads/scan-antigypsyism.pdf>

Analytical Paper “Beyond the “Big Three” - Alternative platforms for online hate speech”:

<http://scan-project.eu/wp-content/uploads/scan-antigypsyism.pdf>

Analytical Paper “Hot Spots of Hate”:

http://scan-project.eu/wp-content/uploads/scan_analytical-paper-3_Hot-Spots_final.pdf

Analytical Paper “Intersectional Hate Speech Online”:

http://scan-project.eu/wp-content/uploads/sCAN_intersectional_hate_final.pdf

sCAN Monitoring Report 2019:

http://scan-project.eu/wp-content/uploads/sCAN_monitoring_report_year_1.pdf

sCAN Monitoring Report 2020:

http://scan-project.eu/wp-content/uploads/sCAN_monitoring_report2_final.pdf

Policy Recommendations:

http://scan-project.eu/wp-content/uploads/sCAN_recommendations_paper_final.pdf

Online Course “Understanding and countering hate speech”:

In English: <https://www.facingfacts.eu/courses/online-course-on-hate-speech/>

In German: <https://www.facingfacts.eu/courses/hate-speech-was-tun/>

In French: <https://www.facingfacts.eu/courses/combattre-les-discours-de-haine-sur-internet/>

In Italian: <https://www.facingfacts.eu/courses/discorsi-dodio-online-riconoscerli-e-contrastarli/>

Online Course “Hate Speech Moderation”:

In English and French:

<https://www.facingfacts.eu/courses/moderating-online-hate-speech/>

Advanced Monitoring Training:

<https://www.youtube.com/watch?v=8t1p5fS2N8U&feature=youtu.be>

References

Fundacion Secretariado Gitano (2019). Guide on intersectional discrimination – The case of Ro-ma women, p. 6. Available at https://gitanos.org/upload/22/65/GUIDE_ON_INTERSECTIONAL_DISCRIMINATION_-_ROMA_WOMEN_-_FSG_33444_.pdf (last accessed 08.04.2020).

International Network Against Cyber Hate (2018). What is cyber hate. Available at <http://www.inach.net/wp-content/uploads/WHAT-IS-CYBER-HATE-update.pdf> (last accessed 24.06.2019).



*Project funded by the European Union's Rights,
Equality and Citizenship Programme (2014-2020)*

Partners



LICRA – International League against Racism and Antisemitism / France
www.licra.org



Jugendschutz.net / Germany
jugendschutz.net



CEJI-A Jewish contribution to an Inclusive Europe / Belgium
ceji.org



CESIE / Italy
cesie.org



ZARA (Zivilcourage und Anti-Rassismus-Arbeit) / Austria
zara.or.at



Human Rights House Zagreb / Croatia
humanrightshouse.org



University of Ljubljana, Faculty of Social Sciences / Slovenia
fdv.uni-lj.si



Romea / Czech Republic
www.romea.cz



Latvian Center for Human Rights / Latvia
cilvektiesibas.org.lv



RespectZone / France
www.respectzone.org



Associate partner:
International Network against Cyber Hate (INACH)
www.inach.net



Project funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020)