Rights, Equality and Citizenship (REC)

Programme of the European Commission

(2014-2020)

Monitoring and Detecting Online Hate Speech

FINAL v1.0

D4.2: Best Practice Guide for responding to Online Hate Speech for internet industry

Abstract

This document focuses on the role of internet industry and develops the key principles for best practice guidelines for responding to online illegal hate speech. There is a short review of the current landscape of current initiatives and a description of current best practice guides in this area.

This is deliverable 4.2 for the Mandola project. This deliverable is dependent on the work conducted by Workstream 2 on Legal issues and especially the outcomes of Deliverable 2.1 which provides a broad, comprehensive review of the legal issues relevant to assessment of hate speech.

Contractual Date of Delivery	30 Sep 2016
Actual Date of Delivery	6-Mar-2017
Deliverable Security Class	Public
Editor	Cormac Callanan <mandola@aconite.com>
Contributors	All *MANDOLA* partners
Quality Assurance	Marios Dikaiakos <mdd@cs.ucy.ac.cy>

This project is implemented by the Mandola Consortium and funded by the Rights, Equality and Citizenship (REC) Programme of the European Commission.

The *MANDOLA* consortium consists of:

FORTH	Coordinator	Greece
ACONITE	Principal Contractor	Ireland
ICITA	Principal Contractor	Bulgaria
INTHEMIS	Principal Contractor	France
UAM	Principal Contractor	Spain
UCY	Principal Contractor	Cyprus
UM	Principal Contractor	France

Document Revisions & Quality Assurance

Internal Reviewers

1. Marios Dikaiakos mdd@cs.ucy.ac.cy
2. Evangelos Markatos markatos@ics.forth.gr

Revisions

Version	Date	By	Overview
1.0	5-Mar-17	Editor, Aconite Cormac Callanan	Completed Edits and updates
0.8	4-Mar-17	UCY, CY Marios Dikaiakos FORTH , Vangelis Markatos	Quality Assurance Review
0.9	14-Feb-17	Editor, Aconite Cormac Callanan	Edits, updates and corrections
0.8	5 Feb 16	Estelle De Marco, Inthemis	Edition of paragraphs and sections relating to legal and ethical issues.
0.8	5 Feb 16	Nikos Frydas, FORTH	Edits and updates
0.8	16-Jan-17	Editor, Aconite Cormac Callanan	Edits and updates
0.7	9 Nov 16	Estelle De Marco, Inthemis	Modifications and contribution in most sections, (paragraphs and sections relating to legal and ethical issues.)
0.6	18-Oct-16	Editor, Aconite Cormac Callanan Karmen Kegl	Major update
0.5	26-Sep-16	UM, FR Professor Adel Jomni Christian Xavier Castane Caroline Greco	Background research, range of selected case studies, and relevant links
0.5	26-Sep-16/	UCY, CY Maria Poveda George Pallis	Background research case studies and relevant links
0.5	26-Sep-16	ICITA, Bulgaria Veselin Boyadzhiev Albena Spasova	Background Research, and relevant links
0.5	26-Sep-16/	UAM, ES Alvaro Ortigosa Paloma Diaz Carlotta Urruela	Background research and relevant links
0.4	26-Sep-16	Editor, Aconite Cormac Callanan Karmen Kegl	First draft

Table of Contents

1 Introduction

This document focuses on the role of internet industry and develops the key principles to support the development of best practice guidelines for responding to online illegal hate speech. There is a short review of the current landscape of current initiatives and a description of current best practice guides in this area. This is deliverable 4.2 for the Mandola project[1].

Hate speech is a very complex issue with many different definitions and understandings of what is considered to be hate speech. In the Manual on hate speech[2] by Anne Weber it is noted that "*no universally accepted definition of the term "hate speech" exists, despite its frequent usage*". This manual highlights the issues around incitement of racial hatred when anger and hatred is directed against specific persons or groups in society specifically on the basis of being a member of a specific race. It also covers such behaviours based on religious grounds.

The definition used for hate speech in this document is what is understood as Hate Speech from the sociologically perspective based on the legal work implemented in work-stream 2 of the Mandola project but also based on the work undertaken by work-stream 3. The EC Code of Conduct[3] is very relevant to these guidelines in addition to the contributions and outputs from the Mandola Advisory Board meeting (Brussels, October 2016) and from the Mandola workshop (Brussels, December 2016) with participation from relevant sectors.

As regards legal instruments, an important international instrument is the additional protocol to the Council of Europe Cybercrime Convention[4], concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems[5]. Another instrument is Recommendation R(97) 20 of the Committee of Ministers to member States on "Hate Speech", which considers that the term "hate speech" might include all forms of expression which spread, incite, promote or even justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance. This also includes intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.[6] However, the additional protocol is only binding at a given national level in the extent that the country is party to the convention and implemented its content into domestic law. Recommendation R(97)20 is not binding on Member States.

Even within the framework of applicable legal instruments, the identification of illegal hate speech is a complex challenge since legislations cover a comprehensive range of issues which protect freedom speech but also place limits and responsibilities on these freedoms thereby restricting what is permitted, and these limits might be subtle, in addition to being different depending on the national applicable law.

---

[1] This deliverable is dependent on the work conducted by Workstream 2 on Legal issues and especially the outcomes of Deliverable 2.1 which provides a broad, comprehensive review of the legal issues relevant to assessment of hate speech. This document depends on the ongoing work of Mandola on Legal and Ethical Framework which is developing a document on relevant legal definitions in this space.

[2] http://www.coe.int/t/dghl/standardsetting/hrpolicy/Publications/Hate_Speech_EN.pdf

[3] European Commission "Code of Conduct on the countering illegal hate speech online" http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf (accessed 12-Jan-2017)

[4] http://www.coe.int/en/web/cybercrime/the-budapest-convention (last accessed 20-Oct-2016)

[5] http://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/189

[6] http://www.coe.int/t/dghl/standardsetting/hrpolicy/Other_Committees/DH-LGBT_docs/CM_Rec(97)20_en.pdf

Therefore, the question of what actions the internet industry should or should not take, where online content is alleged to be online hate speech, is a very complex issue relating to the protection of freedom of expression, the presumption of innocence and even of freedom of trade (in case of censorship of a content). Indeed, all online hate speech might not be illegal and this illegality is in many cases difficult to determine. In addition, the actions taken by industry might occur before any decision of an independent judge is issued in relation to the alleged illegality of content Moreover, States and ISPs have a special duty to ensure the appropriate protection of the freedom[7], of the freedom of expression and of the freedom of assembly[8] on the Internet.

In 2016, the European Commission launched a Code of Conduct document aimed at guiding Facebook, Microsoft, Twitter ("the IT companies") and YouTube activities as well as sharing best practices with other internet companies, platforms and social media operators. Facebook, Microsoft, Twitter and YouTube– also involved in the EU Internet Forum – share, together with other platforms and social media companies, a collective responsibility and pride in promoting and facilitating freedom of expression throughout the online world.

The current document does not provide detailed legal descriptions for companies but identifies best practice in responding to illegal hate speech while ensuring the preservation of other human rights. In addition, detailed legal and ethical issues are comprehensively covered in the work and deliverables of Mandola on Legal and Ethical Framework. Further research will be required by readers of this document to fully understand and cope with the complex nuances when interpreting hate speech.

The current document provides a short introduction to the Mandola project and the partners. This is followed by a short description of the current landscape of current initiatives and the key strategies which identify a comprehensive, effective approach to responding to online hate speech. The role of legislation, policy and regulation is explored and the important aspects of best practices are identified. The document provides a brief focus on the potential role of industry and then offers some practical guidelines for internet users and potential victims of online hate speech. The European Hotlines accepting reports about Hate speech are profiled as an example of good practice combating illegal hate speech.

The document then provides sources for additional reading on this complex subject.

This document has been developed by the partners to the Mandola project based on research on current best practices guide. It consists in a comparative study of international, European and national best practices guides, but also encompasses legal, ethical and IT expert views. Through this Best Practices Guide, Mandola aims to empower a harmonized and increased involvement of Internet Industry in the countering of online hate speech.

[7] Council of Europe, Recommendation CM/Rec(2016)5 of the Committee of Ministers to member States on Internet freedom, https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016806415fa

[8] Council of Europe, Declaration of the Committee of Ministers on the protection of freedom of expression and freedom of assembly and association with regard to privately operated internet platforms and online service providers, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805cb844; Council of Europe, Recommendation CM/Rec(2015)6 of the Committee of Ministers to member States on the free, transboundary flow of information on the Internet, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805c3f20; Council of Europe, Recommendation CM/Rec(2016)1 of the Committee of Ministers to member States on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutrality, https://wcd.coe.int/ViewDoc.jsp?p=&Ref=CM/Rec(2016)1&Language=lanEnglish&Ver=original&BackCol=&direct=true.

1.1 Target Audience

This document is designed as a discussion document based on extensive research by the Mandola partners into the illegal hate speech on the Internet. The purpose is to identify the best practice based on the experiences of many organisations active in this space on the internet, with a consideration of legal and ethical aspects protecting Human rights. The document is designed primarily for service providers on the Internet who offer or enable sharing of content on the internet including social networking websites, hosting organisations, and chat rooms.

It is not intended to be prescriptive in nature but to offer guidelines about effective and ethical responses to online hate speech. Neither is it intended as a binding document but one to support and encourage action against illegal hate speech that would be respectful of other rights such as presumption of innocence, freedom of expression, freedom of assembly, freedom of trade, and even the right to privacy and to personal data protection. Sometimes such action might require censorship of illegal content or perhaps stimulating counter narrative.

2 Mandola Project Overview

MANDOLA (Monitoring ANd Detecting OnLine hAte speech) is a 24-months project co-funded by the Rights, Equality and Citizenship (REC) Programme of the European Commission. The project is fully operational from October 2015 until September 2017 with reduced support after those dates.

The project is led by the project coordinator FORTH (Foundation for Research and Technology – Hellas) in a consortium with Aconite Internet Solutions (Ireland), the International Cyber Investigation Training Academy (Bulgaria), Inthemis (France), the Autonomous University of Madrid (Spain), the University of Cyprus (Cyprus) and the University of Montpellier (France). Further, and up-to date-details are available on the project website on http://www.mandola-project.eu.

2.1 Mandola Objectives

MANDOLA aims at improving the public understanding of how on-line hate speech prevails and spreads. The project also aims at empowering ordinary citizens to monitor and report hate speech. MANDOLA's objectives are:

- to monitor the spread and penetration of on-line hate-related speech in EU member states using a big-data approach, while investigating the possibility to distinguish between the potentially illegal hate-related speech and non-illegal hate-related speech;
- to provide policy makers with information that can be used to promote policies for mitigating the spread of on-line hate speech;
- to provide ordinary citizens with useful tools that can help them deal with on-line hate speech irrespective of whether they are bystanders or victims;
- to transfer best practices among EU Member States.

The MANDOLA project addresses the two major difficulties in dealing with on-line hate speech: lack of reliable data and poor awareness on how to deal with the issue. Although in general on-line hate speech seems to be on the rise, it is not clear which member states seem to be suffering most. It is not even clear which kind of on-line hate speech (e.g. homophobia, Xenophobia, etc) is on the rise. Moreover, the available data generally do not easily distinguish between illegal hate content and harmful (but not illegal) hate content. The different legal systems in various member states make it difficult for ordinary people to make such a distinction. It is even more difficult for citizens to know how to deal with illegal hate content and to know how to behave when facing harmful but not illegal hate content. Without reliable data it is very difficult to make reliable decisions and push policies to the appropriate level.

2.2 Mandola Innovations

The project has two main innovative aspects. The first is the extensive use of IT and big data to study and report on-line hate, and the second is the research on the possibility to make clear

distinction between legal and not illegal content taking into account the variations between EU member states legislations.

MANDOLA is serving: (i) policy makers - who will have up-to-date on-line hate speech-related information that can be used to create enlightened policy in the field; (ii) ordinary citizens - who will have a better understanding of what on-line hate speech is and how it evolves, will be provided with information for recognizing legal and (illegal) on-line hate-speech and will know what to do when they encounter (illegal) on-line hate; and (iii) witnesses of on-line hate speech incidents - who will have the possibility to report hate speech anonymously.

2.3 Mandola Activities

In order to achieve the set up objectives the project envisages several activities:

- An analysis of the legislation of illegal hate-speech at national, European, and international and national level is being conducted.
- The legal and ethical framework on privacy, personal data and protection of other fundamental rights is being identified and analysed in order to implement adequate safeguards during research and in the system to be developed.
- A monitoring dashboard is being developed. It will identify and visualize statistics of on-line hate-related speech via social media (such as Twitter) and the Web (such as Google).
- A multi-lingual corpus of hate-related speech will be created based on the collected data. It will be used to define queries in order to identify Web pages that may contain hate-related speech and to filter the tweets during the pre-processing phase. The vocabulary will be developed with the support of social scientists and enhanced by the Hatebase (http://www.hatebase.org/).
- A reporting portal will be developed. It will allow Internet users to report potentially illegal hate-related speech material and criminal activities they have noticed on the Internet.
- A smart-phone application will be developed. It will allow anonymous reporting of potentially hate-related speech materials noticed on the Web and in social media.
- A Frequently Asked Questions document has been created and has been disseminated. The FAQ document will answer questions like: What is on-line hate speech? Which forms are legal and which - potentially illegal? What are Internet Service Providers doing? What can users do if they encounter a hateful video, blog, group in Facebook or similar networking site, receive a hate e-mail or come across a hate-related web site? What can they do if they become target of hate-related comments on-line? How to protect themselves and their children in social networks? The FAQ document will be disseminated via the project portal and the smart-phone app.
- A network of National Liaison Officers (NLOs) of the participating member states will be created. They will act as contact persons for their country and will exchange best practices and information. They will also support the project and its activities with legal and technical expertise when needed.

- Landscape and gap analysis. Some countries still do not have sufficient methods or structures to handle complaints or reports about hate speech. That is why a landscape of current responses to hate speech across Europe will be developed and Best Practices Guide for responding to on-line hate speech for Internet industry in Europe will be created and disseminated. A comprehensive survey among key stakeholders - major Internet Service Providers and Law Enforcement will be conducted. They will identify the key challenges and best practices in responding to hate speech transnationally.

3 Landscape of Current Initiatives

Several forms of hate speech are illegal in the European Union (EU), but all Member States do not punish the exact same behaviours. Other forms of speech are strongly protected by the freedom of expression, which is widely ensured in Europe, as well as in many other jurisdictions around the world. The debate lies in the question to know what is or not legally permitted, in a given society, which is a very complex issue. There are complex issues relating to democracy, legality and issues around ethics and morality which place limits on freedom of speech in order to prevent illegal hate speech. The Mandola project was created to identify internet tools to support responses against online hate speech

3.1 Background

The Internet can be characterized as the most effective global means of communication that enables freedom of expression and speech to a unrivalled extent.

Often, internet users can retain their anonymity or pseudo-anonymity and can access the internet at any time from a variety of smart mobile electronic devices. Such (perceived) anonymity is a critical feature that facilitates the commitment of cyber crime including online illegal hate speech. Hate speech is recognized by international law as a type of behaviour that should be criminalised, or at least appropriately sanctioned, and many national laws make it a penal offence[9] (even if some civil or administrative torts also exist).

The current legal and regulatory situation is extensively described in the Mandola intermediate report on the definition of illegal hatred and implications[10].. It is important to read this current document in the context of the comprehensive review completed in that document.

When incidents of online illegal hate speech occur they need to be recognised, scrutinized, and countered.

3.2 Landscape of current initiatives

Many campaigns such as Nohatespeech movement[11] have been launched across Europe and some European projects have been funded, such as INACH[12] and BRICKS[13]). However, there are a few best practice guidelines that have been published for confronting online hate speech.

- In June 2016, the European Commission developed a code of conduct against hate speech online[14].
- In November 2012, the organisation known as A Jewish Contribution to an Inclusive Europe (CEJI) under the Facing Facts project[15] produced the publication “Guidelines for Monitoring of Hate Crimes and Hate Motivated Incidents”[16].

[9] The term "penal offence" is used rather than the word “crime” since in many jurisdictions the word crime refers to the most serious type of offences. This is not always the situation for hate speech which is often treated as a misdemeanour.
[10] http://www.mandola-project.eu/m/filer_public/7b/8f/7b8f3f88-2270-47ed-8791-8fbfb320b755/mandola-d21.pdf
[11] https://www.nohatespeechmovement.org/ (last accessed 20_oct-2016)
[12] http://www.inach.net/index.php (last accessed 3-Oct-2016)
[13] http://www.bricks-project.eu (last accessed 3-Oct-2016)
[14] http://ec.europa.eu/justice/fundamental-rights/files/code_of_conduct_hate_speech_en.pdf (last accessed 20-oct-2016)

- In 2015, UNESCO (United Nations, Educational, Scientific and Cultural Organization) published the "Countering Online hate speech" guide[17].
- In 2014, in the context of No Hate speech movement, the "Starting points for Combating Hate Speech online"[18] was published and in 2016 a new revised "Manual for combating hate speech online through human rights in education"[19] was produced.
- ADL (Anti-Defamation League)[20] organisation has published "best practices online for responding to cyber hate"[21]. According to the ADL organisation, it would be helpful if ISPs facilitate and support witnesses and victims of online illegal hate speech by providing them with the opportunity to submit reports related to suspected illegal content. In such cases, they should provide explanations regarding the process they follow for responding to these reports and, if necessary, respond to the users reports as fast as possible. (ADL company, "Best practices for responding to cyberhate").

3.3 Fundamental Principles

The already existing guidelines of the above sources include the following principles that are essential for responding to online hate speech.

a. Awareness raising

It is very important for internet users to be fully aware of the offence of hate speech. The legal definition of a hate offence (often called hate crime) usually confuses the majority of internet users who are sometimes unaware that hate speech is prohibited by legislation.

Although democracy supports and encourages open debate and freedom of speech, there are limits to freedom of expression which need to be highlighted and explained to users in order to avoid being accused or prosecuted as perpetrators of hate speech.

Potential victims need to understand and demand respect for their fundamental rights. They should know the steps to be followed in order to report illegal incidents and demand restoration (includes deletion of offending content and taking steps to repair the damage done) and sometimes restitution (includes compensation).

b. Reporting mechanisms

Cooperation between national Civil Society Organizations (CSOs), local authorities, law enforcement and other governmental bodies is crucial in order to respond to illegal online hate speech. The expertise and the services that different stakeholders from different sectors can provide enable the reporting mechanism to collect the necessary expertise together to prevent and to respond to illegal online hate speech. Since the internet is not constrained by physical geographical or political boundaries, the

[15] http://www.facingfacts.eu (last accessed 3-Oct-2016)
[16] http://www.ceji.org/media/Guidelines-for-monitoring-of-hate-crimes-and-hate-motivated-incidents-PROTECTED.pdf (last accessed 20-oct)
[17] http://unesdoc.unesco.org/images/0023/002332/233231e.pdf (last accessed 20-oct 2016)
[18] https://www.coe.int/t/dg4/youth/Source/Resources/Publications/2014_Starting_Points_for_Combating_Hate_Speech_Online.pdf (las accessed 20-oct-2016)
[19] http://www.coe.int/t/dg4/youth/Source/Resources/Publications/BOOKMARKS.pdf (last accessed 20-oct-2016)
[20] http://www.adl.org/ (last accessed 3-Oct-2016)
[21] http://www.adl.org/assets/pdf/combating-hate/2016-ADL-Responding-to-Cyberhate-Progress-and-Trends-Report.pdf (last accessed 20-oct-2016)

cooperation and exchange of best practices among national LEAs is a significant challenge.[22]

Internet hotlines operate in the majority of European countries and offer to users reporting mechanisms for illegal internet content. Despite the fact that the priority of the hotlines is the elimination of online child sexual abuse material (csam), many hotlines accept reports on racism and xenophobia. After analysis by trained content analysts, verified reports which are considered to be potentially illegal are forwarded to the national law enforcement for further investigation and prosecution.

According to the recent EC Code of conduct on the countering illegal hate speech online", it would be helpful if IT companies could also encourage users to report incidents they encounter relating to online illegal hate speech.[23]

c. Role of IT Companies

The category of IT companies is used from the EC Code of Conduct as a title to describe the collective of four specific companies called Facebook, Microsoft (Microsoft-hosted consumer services, as relevant), Twitter and YouTube involved in the EU Internet Forum. Such service providers are described as social media companies and other platforms and offer a complex range of IT services including hosting internet content written and published by third parties and social networking websites.

It is undoubtedly valuable that IT Companies, where relevant, provide reporting tools in order to report potential hate speech on their websites. When they process the reports themselves, they need to have trained staff processing these notifications and empowered to take the appropriate measures to respond to illegal hate speech, in compliance with law. This might involve notifying the author or requesting the removal of the illegal content, or in extreme measures disabling access to the content, where legally permitted and compliant with the rule of law and the principle of the preservation of fundamental rights. A response given in a short period of elapsed time reduces the impact, and the EC Code of Practice suggests a period of less than 24 hours after the receipt of the notifications. Where legal provisions are not sufficient to enable the IT company to contribute to the combat against potentially illegal content, it is advised to adopt terms and conditions of service (TOS), which provide detailed information to the users to notify them about content that is forbidden by the IT companies (where the legal framework authorises the IT company to forbid certain kinds of content[24]) or illegal

[22] DEJI under the Facing Facts project, November 2012 "Guidelines for Monitoring of Hate Crime and Hate Motivated Incidents" http://www.ceji.org/media/Guidelines-for-monitoring-of-hate-crimes-and-hate-motivated-incidents-PROTECTED.pdf (last accessed 20-Oct-2016)

[23] European Commission "Code of Conduct on the countering illegal hate speech online" http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf (accessed 12-Jan-2017)

[24] Globally and outside any specific law, the European Convention of Human Rights principles require that any limitation of the freedom of Internet users to access Internet services must be necessary and proportionate (and primarily justified by the service to be delivered), in addition to be legally based and to pursue a legitimate purpose. This might prevent a large part of ISPs to prohibit certain kinds of contents. See inter alia Council of Europe, Recommendation CM/Rec(2012)4 of the Committee of Ministers to member States on the protection of human rights with regard to social networking services, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa9b, especially §3 ("*Threats (to Human Rights) may, in particular, arise from lack of legal, and procedural, safeguards surrounding processes that can lead to the exclusion of users; "), §5 ("Freedom of expression includes the freedom to impart and receive information which may be shocking, disturbing and offensive. Content that is unsuitable for particular age groups may well also be protected under Article 10 of the European Convention on Human Rights, albeit subject to conditions as to its distribution") and §10 ("develop editorial policies so that relevant content or behaviour can be defined as "inappropriate" in the terms and conditions of use of the social networking service, while ensuring that this approach does not restrict the right to freedom of expression and information in the terms guaranteed by the European Convention on Human Rights").* See also Council of Europe, Recommendation CM/Rec(2011)8 of the Committee of Ministers to

according to the law. The form and content of the notification to the IT company might moreover be regulated by law (such as in France).

d. Role of Internet Service Providers (ISPs)

Whereas the term "IT companies" described in the last section refers to a specified list of companies who have adopted the EC Code of practice, the definition of Internet service providers includes those that offer a wider range of internet services including internet access services, email and hosting services and technical support.

The same comments made in relation with IT companies fully apply to ISPs, keeping in mind that these stakeholders are subject to a reduced liability regime[25] (particularly access providers), and they cannot be subject to a general obligation to monitor the content they broadcast or store[26], since their neutrality towards content is a condition of the exercise of several freedoms on the Internet[27] and of the development of the digital economy[28].

In conclusion, to effectively respond to illegal online hate speech all the four aspects listed in a)-d) previously need to be properly considered. Ideally, stakeholders from academia, Internet Service Providers, including IT Companies, law enforcement, governmental authorities and other CSOs should exchange experience – both good and bad - and share best practices. Main objectives are to raise awareness among users to better understand the issues and to promote safer internet use and to encourage them to report illegal hate speech incidents and provide useful and reliable reporting systems at national and international level.

3.4 Known Challenges

There are several issues encountered in the response against online hate speech. These issues can be categorized in two groups.

Firstly, there is the issue raised by the lack of a clear concise accessible definition which makes it difficult to fully understand and recognize what is illegal online hate speech and which behaviours have to be considered illegal rather than simply (non-illegal) provocative, offensive or insulting, and therefore cannot benefit from the protection of the freedom of expression.

Secondly, the issue related to the intrinsic characteristics of the Internet which does not recognise geographical or political boundaries, combined with the lack of international law and

member states on the protection and promotion of the universality, integrity and openness of the Internet, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805cc2f8; Council of Europe, Declaration of the Committee of Ministers on the protection of freedom of expression and freedom of assembly and association with regard to privately operated Internet platforms and online service providers, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805cb844; Council of Europe, Recommendation CM/Rec(2015)6 of the Committee of Ministers to member States on the free, transboundary flow of information on the Internet, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805c3f20; Council of Europe, Recommendation CM/Rec(2016)1 of the Committee of Ministers to member States on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutrality, https://wcd.coe.int/ViewDoc.jsp?p=&Ref=CM/Rec(2016)1&Language=lanEnglish&Ver=original&BackCol=&direct=true.

[25] Articles 12 to 14 of the EU Directive 2000/31/EC.

[26] Article 15 of the EU Directive 2000/31/EC.

[27] See footnote 24. For a national example, regarding hosting providers, see the French decision CA Versailles, 14ème ch., 12 décembre 2007, http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=2118.

[28] See the preamble of Directive 2000/31/EC. For a national example, regarding hosting providers, see the French decision CA Paris, 6 mai 2009, S.A. Dailymotion c/ M. C., Société Nord-Ouest Production et S.A. UGC Images, http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=2634.

the disparity between the national legislations undermines the response against illegal online hate speech.

3.4.1 What is illegal online hate speech?

All current legal definitions of hate speech[29] are quite problematic, since there is no universally shared single definition. Illegal hate speech refers to broad and contested concepts which have yet to be consensually, globally defined. Dr. McGonagle believes that the "*term is a convenient shorthand way of referring to a broad spectrum of extremely negative discourse stretching from hatred and incitement-to-hatred; to abusive expression and vilification; and arguably also to extreme forms of prejudice and bias*"[30]. This broad spectrum includes those cases where people are simply venting frustrations or anger with unfortunate, inconsiderate thoughtless phrases exposing their prejudices. However, contents belonging to this category are illegal only where they are precisely prohibited by a national law. In addition, most of the time, they are illegal only if committed in particular circumstances (for example if the motivation of the perpetrator is of a particular nature, if the content is publicly accessible, or if specific results are obtained, such as a public disorder). Consequently, identifying whether specific material is potentially illegal implies that the applicable national law needs to be identified, and to further confirm if the precise terms of the identified national legislation are applicable. Contents that do not correspond to these terms are permitted and strongly protected by the freedom of expression, even where information or ideas "*offend, shock or disturb the State or any sector of the population*"[31]. Within such a context, the differentiation between illegal behaviours from permissible ones might be particularly difficult, and any broad scope definition is not of any help, unless it is used to identify globally the contents that MIGHT be potentially illegal from those that are incontestably not illegal, in the area of hate speech. But in this case it must be kept in mind that this broad definition will encompass both some legal and illegal speech.

For this reason, European and International instruments and competent institutions are working to enhance the harmonisation of legislations. This task is very sensitive since penal law is a matter of national sovereignty for each Member State, as well as the definition of morals.

In parallel, the European Court of Human Rights (ECtHR) has identified some "forms of expression which are to be considered offensive and contrary"[32] to the European Convention on Human Rights, and a number of parameters that enables characterisation of such content as unacceptable "hate speech"[33]. This does not mean that these forms of expression are illegal, but it means that Parties to the Convention, under certain circumstances, are entitled to prohibit them.

These forms of expression, which "it may be considered necessary in certain democratic societies to sanction or even (to) prevent", are those that "spread, incite, promote or justify

[29] See Deliverable D2.1 of the MANDOLA project.
[30] Dr. Tarlach McGonagle,"The *Council of Europe against online hate speech: Conundrums and challenges*, (Council of Europe Conference on Freedom of Expression, 7-8 November 2013), p. 5. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016800c170f (on 09/26/2016); *and*, James B. Jacobs and Kimberly Potter, *Hate Crimes: Criminal Law and Identity Politics* (New York, Oxford University Press, 1998), p. 11; *See also*, I. Gagliardone et al., *Countering Online Hate Speech*, UNESCO Series on Internet Freedom, UNESCO Publishing, 2015.
[31] *Handyside v. the United Kingdom*, Judgment of 7 December 1976, para. 49.
[32] European Court of Human Rights, Factsheet - Hate speech, February 2012, p. 1.
[33] Ibid..

hatred based on intolerance"[34].For example, among speech examined by the Court and considered incompatible with the values proclaimed and guaranteed by the Convention, lie statements denying the Holocaust, justifying pro-Nazi policy, linking all Muslins with a grave act of terrorism, or portraying the Jews as the source of evil in Russia[35]. However, the determination of the behaviours that must be punished still rely on States, and without the identification of the applicable law and of its provisions, it remains impossible to ascertain that a behaviour based on intolerance and spreading, inciting, promoting or justifying hatred is or not illegal. It must be a case-by-case analysis, taking also into account the precise circumstances of the potentially illegal behaviour, as previously explained.

3.4.2 Borderless Internet

The internet's geographic and political uncertainty and the disparity between different national legislation foster a range of jurisdictional and extraterritorial issues that makes it difficult to enforce national laws against illegal online hate speech. An example of these conflicts is highlighted in the case of Yahoo, Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme, et al[36]. Two French student organisations lodged a complaint against Yahoo! which had displayed content relating to Nazi memorabilia which is forbidden by French penal law. This activity, originated from Yahoo in the USA where it was permitted by the First amendment. In France, Yahoo! was found liable, but the Court in the USA ruled conversely that such a decision was in breach of its fundamental right for freedom of speech[37]. The result of such situations, where contested behaviours are not globally prohibited, is the difficulty to respond in a simple and efficient manner to illegal online hate speech. This makes it even more difficult, for citizens, to understand legal responsibilities surrounding hate speech since such understanding would require they are aware of a wide range of different national legislations regarding illegal online hate speech and their appropriate interpretations. Without this knowledge Internet users might especially, unintentionally be in breach of non-domestic legislation without any intent or even awareness to commit a penal offence.

The differences between penal and constitutional laws are unavoidable, since the determination of the appropriate borders between acceptable and unacceptable speech is both a question of choice of society and a question of means to combat a particular phenomenon (e.g. prohibition vs authorisation accompanied with disclaimers - which is seen by certain authors as a choice between how much citizens are treated like children ("infantilized") or of empowering citizens[38]). The challenges caused by differences between legislations may be minimised by international approaches aiming at harmonising these legislations to the most possible extent. In this sense, such international approaches are crucial, and have led to the adoption of, inter alia, the additional protocol to the Convention on cybercrime and, at the EU level, of the Council

[34] *Erbakan v. Turkey*, Judgment of 6 July 2006, para. 56; see also, Recommendation No. R (97) 20 of the Committee of Ministers to Member States on "*Hate Speech*", 30 October 1997, p. 107.
[35] *Delfi AS V. Estonia*, Grand chamber Judgment of the European Court of Human Rights of 16 June 2015, para. 136.
[36] http://www.tjsl.edu/slomansonb/5.2%20Yahoo%20US.pdf (last accessed 20-oct-2016) and Yahoo! Inc. V. La Ligue Contre Le Racisme Et L'antisemitisme and L'union Des Etudiants Juifs De France, No. 01-17424 ,United States Court of Appeals, Ninth Circuit 2006.
[37] *Yahoo, Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme et al*, 24 September 2000, http://www.internetlibrary.com/cases/lib_case17.cfm (on 09/26/2016).
[38] See for example Raymond Aron (referring to *Democracy in America*, by Alexis de Tocqueville), *Essai sur les libertés (Essay on freedoms)*, ed. Hachette, coll. Pluriel, 1976, pp. 132-133. See also Estelle De Marco, *L'anonymat sur Internet et le droit* (*Anonymity on the Internet and the Law*, thesis, Montpellier 1, 2005, ANRT (ISBN : 978-2-7295-6899-3 ; Ref. : 05MON10067), n° 84 (« La protection de l'anonymat sur Internet et la démocratie »).

Framework decision 2008/913/JHA[39]. However, these legal instruments have not been transposed into domestic laws in the most complete manner, and they enable States to restrict certain advised prohibitions, and to create or maintain additional ones[40]. As a result, a high disparity is noticed between the legislations of the EU Member States.[41]

[39] European Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV:l33178 (last accessed 20-Oct-2016)
[40] Mandola project D2.1 on the legal definition of hate speech
[41] ibid

4 Outline of best Practice Guide

4.1 Legislation, Policy and Regulation

The MANDOLA project has created this document as a practical guide for the Internet industry developing strategies to tackle illegal online hate speech.

This guide primarily addresses the legal responsibilities in addition to the ethical and social responsibilities of business offering internet services. It is not targeted towards state agencies or nations which also have responsibilities to address illegal online hate speech, nor is it intended to identify the social impact of hate speech online behaviour. The main purpose of this best practice guide is to ensure that every organisation can gain a deeper understanding of the options that exist in relation to tackling illegal online hate speech. It provides valuable experiences and understandings that are needed in order to make reasoned judgements, so that the organisations are able to make informed decisions and predict consequences of these choices[42]. The Internet industry, especially Internet Service Providers and Technology design and manufacturing companies, who have an important role as internet stakeholders and have technical and legal responsibilities and proscribed influence over content and publications including restrictions caused by the neutrality principle and according to their specific liability regime are the target readership for this best practice guide. They have an important role as intermediaries, both in relation to the protection of the freedom of expression and of the freedom of assembly and in relation to their practical and technical capabilities to contribute to the safety of Internet users. In this sense, these organisations might help governments to enforce the law in such a non-traditional, decentralised place as the Internet, as well as in raising awareness in society in general and users particularly about online hate speech and the importance of tackling this phenomenon.

This Best Practice Guide has been developed by researching and analysing the articles and tools which have been and are currently available when dealing with online hate speech; and by identifying important issues which can contribute to either the spread of the online hate speech phenomenon, its evolution or its eradication. Finally, it also proposes additional readings, information and resources should any reader feels the need to widen their knowledge and attention regarding online hate speech.

[42] Cohen-Almagor, R. (2011). Fighting Hate and Bigotry on the Internet. Policy & Internet, 3(3), 1-26

5 Best Practices identified

5.1 Background

Public regulation and self- and co-regulated measures have been identified in managing illegal Internet content. The former consist of laws and directives at national and supra-national levels and the latter refer to mechanisms such as hotlines, codes of conduct and TOS, filtering software and rating systems[43]

There is a comprehensive complex variety of internet industry stakeholders and each has different levels of technical and legal visibility on the challenge of online illegal hate speech. Some organisations develop hardware and software for laptops, mobile devices and tablets and they can design devise with enhanced security and privacy options to mitigate risks for the end-user. Software companies can configure their operating systems and applications to collect, store and share data and should develop such systems aware that trust and confidence can be achieved with good both security and privacy designs. Service providers include telecommunications companies which have limited visibility on the data or the content of the data packets that are communicated across the fixed and mobile telecommunications network. These providers are often required to provide lawful interception capabilities for authorized national agencies based on clear and transparent national procedures and oversight. Internet Service providers range from those which offer internet access services such as fixed, mobile and satellite broadband services to those which offer complex social networking services or information distribution services.

Internet business models continue to evolve as new ideas emerge and new technologies offer new social and business opportunities and sometimes include hidden, unexpected, unpredicted risks. It is important to understand the strengths and weaknesses of all internet organisations in order to determine the unique knowledge and capabilities each organisation can contribute to both open and free democratic speech and to illegal online behaviours. Of greater importance is for society to debate and investigate what legal, technical and moral roles and responsibilities these individual industry sectors have in responding to online illegal activities or abuse of their services for criminal purposes. This is a complex undertaking and requires open engagement from both civil society and internet industry to achieve the appropriate balance.

The most important regulation relating to the combat against online content might be Directive 2000/31/EC, which organises specific liability regimes for access providers and hosting providers, who, basically, are not responsible for the content they broadcast or store where they stay neutral toward this information, and, in relation with hosting providers only (since neutrality of access providers is particularly crucial[44]), if they act expeditiously to remove or to

[43] Akdeniz, Y. (2001). Controlling Illegal and Harmful Content on the Internet. In D. (. Wall, Crime and the Internet (pp. 113-140). London: Routledge

[44] See for example Council of Europe, Recommendation CM/Rec(2016)1 of the Committee of Ministers to member States on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutrality, 13 January 2016, https://wcd.coe.int/ViewDoc.jsp?p=&Ref=CM/Rec(2016)1&Language=lanEnglish&Ver=original&Site=COE&BackColorInternet=C3C3C3&BackColorIntranet=EDB021&BackColorLogged=F5D383&direct=true.

disable access to the information where they obtain such knowledge or awareness[45]. Several court cases have in addition specified the actions that belong to a neutral action and actions that make the hosting provider going beyond his purely technical (and therefore hosting) role[46].

Moreover, these provisions do not affect the possibility for a court or administrative authority, in accordance with Member States' legal systems, of requiring the service provider to terminate or prevent an infringement[47].

These provisions enable to rapid action to be taken when illegal content is noticed - unless the hosting provider is located outside the EU and refuses to cooperate - to the benefit of victims, but can also raise several concerns. The most important one appears to be that the removal or blocking of the content, performed before any judgement by an independent court, is a kind of private justice (or administrative one, depending on the context, with in certain cases a lack of guarantees in terms of other rights protection[48]), susceptible to hurt in an excessive manner some other rights at stake (freedom of trade, freedom of expression...), with very practical consequences (closure of a business realising online most of its turnover; impossibility of expressing one important view in a closing debate...), in case the removed or blocked content is not declared illegal, at the end. Reported errors of this kind, whether human or automatic, are numerous[49]. This is all the more problematic than the access to the Internet is considered as protected by the right to freedom of expression, even when it is not currently considered as a fundamental right in itself[50], since the Internet is "*a vast platform for cultural expression, access*

[45] Articles 12 and 14 of Directive 2000/31/EC. For further details see Deliverable D2.1 final of the MANDOLA project.

[46] At the EU level, see CJUE, 23 March 2010, Google France / LVM, Viaticum, Luteciel, CNRRH and others; CJUE 11 September 2014, Sotiris P. / O Filelefttheros Dimosia Etaireia Ltd, Takis K., Giorgos S..
At the French level, see C. cass., 1ère ch. Civ., 17 February 2011, Sté Nord-Ouest et a. c/ Sté Dailymotion, http://www.courdecassation.fr/jurisprudence_2/premiere_chambre_civile_568/165_17_19033.html;
C. cass., 1ère ch. Civ., 17 février 2011, M.O. X. c/ Bloobox.net : http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=3103;
TGI Troyes, 4 juin 2008, Sté Hermes international c/ C. F. et a/, http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=2320
TGI Créteil, 14 décembre 2010, INA c/ YouTube http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=3052 ; CA Paris, Pôle 5, Ch. 1, 2 décembre 2014, TF1 et autres / Dailymotion, http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=4401.

[47] Articles 12 and 14 of Directive 2000/31/EC.

[48] See for example the French debate relating to the possibility, for the French administration, to require the blocking of child pornography and terrorist content (for ex. Guillaume Champeau, 10/07/2014, "Loi anti-terroriste : ce que dit le texte des sites web", numerama, http://www.numerama.com/magazine/29975-loi-anti-terrorisme-ce-que-dit-le-texte-sur-le-blocage-des-sites-web.html; Marc Rees, "Le blocage administratif des sites terroristes est sur la rampe", 09/01/15, Next Inpact, http://www.nextinpact.com/news/91674-le-blocage-administratif-sites-terroristes-est-sur-rampe.htm; Marc Rees, "Terrorisme, pédopornographie : où en est le décret sur le blocage administratif?, Next Inpact, http://www.nextinpact.com/news/91634-terrorisme-pedopornographie-ou-en-est-decret-sur-blocage-administratif.htm; Marc Rees, "Le projet de loi sur le terrorisme déjà menacé d'une QPC", 29/10/2014, Next Inpact, http://www.nextinpact.com/news/90674-le-projet-loi-sur-terrorisme-deja-menace-dune-qpc.htm), and its weaknesses (for ex. Marc Rees, "la France active le blocage des sites sans juge. Une première", 6/02/2015, Next Inpact, http://www.nextinpact.com/news/92978-la-france-active-blocage-sites-sans-juge-une-premiere.htm; Le Monde, "Une erreur bloque l'accès à Google pour les clients d'Orange", 17/10/2016, http://www.lemonde.fr/pixels/article/2016/10/17/une-erreur-bloque-l-acces-a-google-pour-les-clients-d-orange_5014900_4408996.html).

[49] For a recent case, see Le Monde, ibid; "Google.fr bloqué pour apologie du terrorisme suite à une erreur humaine d'Orange" (Google.fr blocked for terrorism apology following a human errors of Orange), Marc Rees, 17/10/2016, NEXTInpact, http://www.nextinpact.com/news/101786-google-fr-bloque-pour-apologie-terrorisme-orange-invoque-erreur-humaine.htm (last accessed on 9/11/16). On blocking consequences, see also the Opinion of avdocate general Mr. Pedro Cruz Villalón, delivered on 14 April 2011, in relation to the EUCJ court case C-70/10, Scarlet Extended v Société belge des auteurs compositeurs et éditeurs (ABAM), § 105 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:62010CC0070:FR:HTML; see also Cormac Callanan, Marco Gercke, Estelle De Marco and Hein Dries-Ziekenheiner, Internet blocking - balancing cybercrime responses in democratic societies, October 2009, available at http://www.aconite.com/blocking/study (French version available at http://juriscom.net/2010/05/rapport-filtrage-dinternet-equilibrer-les-reponses-a-la-cybercriminalite-dans-une-societe-democratique-2/).

[50] See for example the European Parliament resolution of 10 April 2008 on cultural industries in Europe, 2007/2153(INI), § 23, accessible on http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2008-0123+0+DOC+XML+V0//EN; see also French constitutional Council, Decision n°2009-580 DC of 10 June 2009, JO of 13 June 2009,

to knowledge, and democratic participation in European creativity, bringing generations together through the information society"[51]. As a result, the access to the internet cannot be limited by measures which would not be legally based, which would not pursue a legitimate aim, and which would not be necessary and proportionate[52].

5.2 Areas of Focus

Best Practices which have been identified cover the following areas of interest

- Reporting to authorities
- User's education and awareness-raising
- Counter-speech
- Codes of Conduct
- Terms of Service
- Hotlines
- Content Moderation
- Proactive and Reactive Content Monitoring
- Filtering, rating systems and removal tools
- Reputation scoring systems
- Additional issues
    - Internet Anonymity
    - Licensing and Regulatory requirements
    - Time-lapse for content reviewing

5.2.1 Reporting to authorities

Several reporting mechanisms enable users and service providers to report suspected illegal content to authorities. In cases relating to illegal hate speech this might be supported by providing images or text, proof of the event, information about its origin, information regarding the user who is potentially responsible for posting the content, etc. However, reporters must take care of the legislation applicable to their report (which should be clarified at the reporting mechanisms level), since it might be illegal in certain countries to copy, store, print or distribute some kind of content (primarily child sexual abuse material). Identities of potential perpetrators must also be handled with care, since they are protected by the data protection legislation, and some countries enable law suits for abusive denunciation. Where information can be legally provided, this provision is encouraged since all relevant supporting evidence will help authorities that investigate online hate speech to achieve a successful investigation and protect the victim's rights and locate the perpetrator.

p. 9675, http://www.conseil-constitutionnel.fr/decision//2009/decisions-pardate/2009/2009-580-dc/decision-n-2009-580-dc-du-10-juin-2009.42666.html, recital n° 12.

[51] European Parliament resolution of 10 April 2008 on cultural industries in Europe, 2007/2153(INI), § 23, accessible on http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2008-0123+0+DOC+XML+V0//EN

[52] See Deliverable D2.2 of the MANDOLA project.

Collaboration with authorities is essential when tackling hate speech.

It is important to note that law enforcement agencies face great challenges when responding to, and prosecuting, illegal online hate speech. The main challenge is with the inherent diverse trans-national structure of the Internet and specifically with the potential anonymity that it provides and the cross-border nature of internet investigations. This makes it sometimes difficult and time consuming for local prosecutors and victims to uncover the identity of the author of the illegal content.

Where the offending party can be identified, multi-jurisdictional differences might lead to difficulties to investigate the case and bring the perpetrator before court. The global nature of the Internet results in countries with less progressive Internet standards and regulations to potentially become locations for users who wish to spread illegal online hate speech unaffected by the laws of the target country.

However, it is still advisable to report hate speech and to raise awareness of the phenomenon. Combining efforts and **approaching Internet hate speech from different perspectives** and by different agents is imperative[53], needing both international and responsible cooperation[54]. Reporting has already led to successful monitoring and tracking the scale of online illegal activity and content and the identification of victims and perpetrators described in annual reports from the hotline members of the INHOPE network including the Irish hotline[55] or in the UK the Internet Watch Foundation[56].

5.2.2 User's education and awareness-raising

Users can be perpetrators/authors of illegal hate speech, witnesses of illegal hate speech or victims of illegal hate speech.

For users who create and contribute to online content including on social networking, websites, chat room and comments on shared spaces, it is important that they understand the responsibility they have to contribute in a socially and legally respectful way. Some of these users are the authors of illegal online hate speech and although freedom-of-speech is an important aspect of modern democracy, users need to respect their local laws that limit this freedom, and must consider the impact of their words especially if they incite violence or hate. They need to be held responsible for such contributions and if appropriate it is important that they can be prosecuted after investigation by law enforcement, in each state that respects the rule of law[57].

Although an approach at the individual user level responding to hate speech might seem to have a limited impact on reducing online hate speech, the role of users in combating hate speech should not be underestimated. Promoting a culture of tolerance and of rejection of cyber-hate, as well as informing users about the consequences of it, would certainly make a difference in the

[53] Bailey, J. (2006). Strategic Alliances: The inter-related roles of citizens, industry and government in combating Internet hate. Canadian Issues, 56-59
[54] Cohen-Almagor, R. (2011). Fighting Hate and Bigotry on the Internet. Policy & Internet, 3(3), 1-26
[55] https://www.hotline.ie/library/annual-reports/2016/report-for-2016.pdf
[56] https://www.iwf.org.uk/accountability/annual-reports/2015-annual-report
[57] See Council of Europe, Recommendation CM/Rec(2016)1 of the Committee of Ministers to member States on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutralityhttps://wcd.coe.int/ViewDoc.jsp?p=&Ref=CM/Rec(2016)1&Language=lanEnglish&Ver=original&BackCol=&direct=true.

fight against hate speech online[58]. Education and awareness-raising of users in the matter of hate speech is therefore important, as it has also be recalled by Cohen-Almagor[59] "*Adopting norms of social responsibility is particularly important when dealing with anti-social speech that is still legal*"

Users who are victims need to be advised about how they can effectively respond against illegal online hate speech and how to protect themselves from the impact of such texts. Users who are witnesses also need to understand how they can contribute to ensuring a safer and more secure society and how they can effectively contribute to eliminating illegal online hate speech.

Comprehensive, concise, coherent, consistent awareness raising messages are important to increase widespread understanding on the boundaries between provocative challenging speech and illegal hate speech which incites hatred and violence.

5.2.3 Counter-speech

Counter-speech is a very effective response to online hate speech that also empowers and educates community members, enabling them to develop a sense of digital citizenship and online civic engagement[60]. Responding to hate speech with alternative more balanced speech offering a range of viewpoints addresses the phenomenon of hate speech and turns it into the centre of attention, by exposing the problem[61]. Counter-speech has been considered as a faster, more flexible and responsive way of dealing with hate speech such as extremism, retaining the principle of freedom of speech in an open space[62] .

However, sometimes counter-speech can be counter-productive. Richards & Calvert[63] note that "The effectiveness of counter-speech, for instance, may be limited by the amount of time available to refute the pernicious speech in question and "whether the counter-message comes to the attention of all the persons who were swayed by the original idea[64]". Some groups and individuals might have very limited access or resources (monetary or emotional resources) to be able to respond to the harmful and illegal content immediately or even respond at all.

It should be noted that the most effective counter-speech is the one that attracts a wider audience and especially media attention and which can reach a significantly increased audience[65]. A greater range of media is available for the average user than years ago, and it offers more diversified capabilities and different ways of accessing, communicating and participating[66]. Therefore, the immediacy and universality that characterises the Internet should not considered as place for cyber hate, but also as an opportunity for an effective response against hate speech with counter-speech - by targeting wider sections of society and raise awareness of this phenomenon.

---

[58] Banks, J. (2010). Regulating hate speech online. International Review of Law, Computers and Technology, 24(3), 233-239
[59] Cohen-Almagor, R. (2011). Fighting Hate and Bigotry on the Internet. Policy & Internet, 3(3), 1-26 Page 6
[60] See Intermediaries and hate speech: Fostering digital citizenship for our information age by Citron, D.K. & Norton, H.L.
[61] Cohen-Almagor, R. (2011). idem
[62] Bartlett, J., & Krasodomski-Jones, A. (2016). Counter-speech on Facebook UK and France. DEMOS.
[63] Richards, R. D., & Calvert, C. (2000). Counterspeech 2000: A New Look at the Old Remedy for "Bad" Speech. Brigham Young University Law Review, 553. Page 554
[64] Vincent Blasi, *Propter Honoris Respectum: Reading Holmes Through the Lens of Schauer: The Abrams Dissent, 72* NOTRE DAME L. REV. 1343, 1357 (1997).
[65] Richards, R. D., & Calvert, C. (2000). Counterspeech 2000: A New Look at the Old Remedy for "Bad" Speech. Brigham Young University Law Review, 553.
[66] McGonagle, T. (2013). The Council of Europe against online hate speech: Conundrums and challenges.

5.2.4 Codes of Conduct (CoC)

The question of whether ISPs, and primarily hosting providers, should contribute to the combat against illegal content is challenged, for reasons already evoked (primarily, the reason is that they do not have the competences of independant judges, whereas the decisions they take can violate the rights of other citizens, and even prevent the development of their own activities, which are crucial for the development of the electronic society). In this context, the EU legislation opted for the contribution of hosting providers in the extent they are made aware of the illegality of a content, and that they do not act expeditiously to remove or to disable access to the information (outside the possibility, for the judiciary, to require the termination or prevention of an infringement to any Internet stakeholder). Most of the debates therefore currently lie in the criteria to be met to qualify a content as "illegal" outside any court-at-law ruling. Another question, directly linked, is the unconfortable position of ISPs, primarily hosting providers, whose potential liability depends on their capability to correctly assess the legality of a content (since they might face liability if they remove a legal content, as well as if they do not remove an illegal content).

In this context, unclear legal situations might be compensated by Code of conducts, which present the advantage of clarifying the conditions of their action towards potentially illegal content, and foster *"standards for responsible and acceptable practices for Internet users*"[67]. One of the most common and popular practice when addressing online problems such as hate speech is the development of Codes of Conduct for the national and international internet industry.

In a report on Self-Regulation of Digital Media Converging on the Internet: Industry Codes of Conduct in Sectoral Analysis[68] by the Oxford Centre for Socio-Legal Studies Programme in Comparative Media Law & Policy, it is written that Codes of Conduct are adopted for many different reasons including as an alternative to direct statutory regulation or to prevent direct statutory regulation by the state. Sometimes they are adopted to build public trust, consumer confidence or to avoid legal or user-perceived liability or to protect children and other consumers or to exert moral pressure on those who otherwise behave in an “unprofessional” or “social irresponsible” way or as a mark of professional status or to develop a set of common standards for services and products; or to raise the public image of their industry. The report follows the five areas of review (known as the 5C’s) (namely constitution, coverage, content, communication and compliance) developed in the IAPCODE tool of analysis which was a methodology used by PCMLP in previous work carried out by Monroe E. Price and Stefaan G. Verhulst.[69] A **Code of Conduct (CoC) can** consist of a list of specific organisational behaviours that are required and a list of prohibited ones and is adopted by companies covering a sector of commercial interests. A CoC outlines the self-regulatory environment that the internet industry has committed to. It is seen as a quality mark of assurance for the internet services provided. Mandola believes that a CoC is best developed and adopted with the support and engagement of the board of directors or senior management of a company or service after consultation with industry representative bodies and relevant regulations.

[67] Cohen-Almagor, R. (2011). Fighting Hate and Bigotry on the Internet. Policy & Internet, 3(3), 1-26 Page 15
[68] http://pcmlp.socleg.ox.ac.uk/wp-content/uploads/2014/12/IAPCODEfinal.pdf
[69] Monroe E. Price & Stefaan G. Verhulst, In Search of the Self: Charting the Course of Self-Regulation on the Internet in a Global Environment, in REGULATING THE GLOBAL INFORMATION SOCIETY 59 (Christopher Marsden ed., London: Routledge 2000)

A CoC often includes a description of minimum terms-of-service (TOS) which is the contract between the customer and the service provider (also known as acceptable use policies (AUPs) which clearly detail restrictions on customer behaviours that use internet services.

In the European Union[70] the adoption of a CoC by companies and service providers can be encouraged when providing electronic services in member states

A good example of a code of conduct is the Irish Internet Service Provider Association Code of Practice.[71]

Once an ISP applies its own regulations, terms of service or CoC regarding hate speech, it must review and monitor the effectiveness of the CoC and the ToS to ensure that the rules continue to be relevant and effective.

This ISP self-regulation might empower the ISPs (without prejudicing here to the lawfulness of such contractual rules) to deny services, and it often includes measures and procedures such as removing offensive web material that breach its policies, as well as cancelation of the ISP service if users do not operate within or comply with the TOS agreements[72,73]. In that sense, ISPs must prepare their self-regulation with care, since it might in turn violate some citizens' rights and be a source of legal liability[74]. Two main sources of liability can be evoked.

- The first one belongs to consumer law. Indeed, according to the EU legislation, most national laws regulate the possibilities in which professionals can refuse to provide a service to a consumer, and code of conduct must not contradict these rules.
- The second one belongs to the protection of citizens' fundamental rights. Indeed, as already exposed, access to the Internet is protected by the freedom of expression and ISPs must ensure that their action do not violate this freedom, as well as, more generally, other (fundamental) rights. Regarding freedom of expression, freedom of assembly and privacy, which are the main freedoms at stake, this implies that the ISP's action is allowed by law, pursues a legitimate aim, and is necessary and proportionate[75], which notably implies that the restriction must be "narrowly tailored and executed with court oversight"[76]. Any restriction that would not be imposed in the respect of these principles could drive to incur the ISP liability before its national judge. Particularly, the use of certain online services might be considered as being a citizens' right, (given their valuable impact in the current exercise of private life by enabling people to establish relationships with others) or freedom of expression[77]. Services of this kind should more than the other take care to not restrict citizens' rights in a disproportionate way. As a minimum, codes of conduct must indicate the ways a decision from the ISP can be challenged by the person who have

[70] Directive 2006/123/EC of the European Parliament and of the Council of 12 December 2006 on services in the internal market OJ L 376, 27.12.2006 pp 36-68
[71] http://www.ispai.ie/code-of-practice/
[72] •Banks, J. (2011). European Regulation of Cross-Border Hate Speech in Cyberspace: The Limits of Legislation. European Journal of Crime, Criminal Law and Criminal Justice, 1-13.
[73] Cohen-Almagor, R. (2011). Fighting Hate and Bigotry on the Internet. Policy & Internet, 3(3), 1-26
[74] See footnote n°24.
[75] See Deliverable D2.2 of the MANDOLA project.
[76] Council of Europe, Guide to human rights for Internet users, Recommendation CM/Rec(2014)6, p. 4, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016804d5b31.
[77] Implicitly, see for example EUCJ, judgment of 16 February 2012, case C-360/10, Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v. Netlog NV, recital 48 & 50, http://curia.europa.eu/juris/document/document.jsf?docid=119512&doclang=EN.

seen his or her freedom limited, and public authorities must ensure that codes of conduct respect these principles[78].

In addition, any handling of users' information is submitted to the data protection legislation.

5.2.5 Terms of Service

Terms of Service (or ToS) are basically rules and obligations specified by a service provider to which customers are obliged to abide. By abiding to these contractual agreements, they are allowed to use the provided internet services. A TOS is based on the specific ethos of a company. Often they contain clear rules forbidding customers to intentionally create, host, transmit material which is unlawful / libellous / abusive / offensive / vulgar / obscene / calculated to cause unreasonable offence.

All that has been written above in relation to the Code of Conduct for the internet industry also applies to the terms of service in the contract between the customer and the internet supplier.

The ToS should be easy to understand, exhaustive, provide specific definitions of terms and useful and realistic answers intended to tackle behaviour that constitutes (in what concerns this report) hate speech. Citroen and Norton propose that an intermediary's voluntary efforts to define and proscribe hate speech should expressly turn on the harms to be targeted and prevented and list categories including speech that threatens and incites violence, speech that intentionally inflicts severe emotional distress, speech that harasses, speech that silences counter-speech and speech that exacerbates hatred or prejudice by defaming an entire group[79]. A ToS should be transparent so as to address the harms and the consequences of policy violations.

There are legal limits and restrictions which can apply to the Terms of Service of a company since such contracts are protected by fair contract consumer legislation. The TOS should be read and accepted by users the first time they access the services, and it also should be available to consult at any time.

The ToS are often supported by a privacy policy indicating how personal data is used by the internet service provider and what data is shared by the company with 3rd parties. This might include terms which recall under which conditions law enforcement agencies are entitled by law to access these personal data. A privacy policy is obligatory in many countries.

It is good practice that the CoC and the ToS are posted in different languages for foreign users.

5.2.6 Hotlines

Although end-users are frequently advised to be aware of safety and security issues when using the internet and noting their concern regarding privacy has increased over the years, it is impossible to avoid encountering insulting or abusive content at some point. End-users can report illegal content to relevant authorities, assisting ISPs to respond to illegal content hosted on their servers without their knowledge.

[78] Council of Europe, Guide to human rights for Internet users, op. cit., especially pp. 4, and 7.
[79] See Citron, D., & Norton, H. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. Boston University Law Review, 91, 1435.

Organisations known as hotlines (or tiplines), enable users to (anonymously if desired) report content that they find potentially illegal. It is one of the most popular and effective strategies in tackling illegal online hate speech. Hotlines are one of the most constructive and effective methods of interaction between users and public authorities, civil society or ISPs when dealing with illegal content. By providing users with a hotline reporting facility, end-users are empowered and can participate as active agents in tackling illegal online hate speech. In Section 8, this report includes a case study of the INHOPE – International Network of Internet Hotlines responding to illegal content on the internet. It describes how reports about online illegal hate speech are received, verified by trained and specialised content analysts, traced and reported to law enforcement for investigation and internet industry for appropriate action.

Most of the time, Hotlines assess received reports in order to confirm they fit the description of illegal online hate speech, before forwarding the information to the competent authorities (where the hotline is not operated by public authorities themselves such as the PHAROS platform[80]).

5.2.7 Reactive and Proactive Content Moderation

An access provider should never block content on the basis of a report unless provided for by law or required by a court at law. Otherwise this violates the principle of the specific liability regime for the access provider together with its obligation of neutrality. The same principle applies in relation to hosting providers, but in this case law generally authorises these stakeholders or imposes to them certain types of actions in precise circumstances. As a result, initiatives involving hotline or reports directly to hosting providers do involve a reactive process by the hosting provider. The hosting provider must have skilled, trained personnel capable of responding to and acting on reports received by the internet hotline or reports to determine the appropriate legal response in a reasonable timeframe. A properly designed ToS and CoC, where legal provision are not sufficient usually empowers the hosting provider to require content determined as likely to be illegal (in a national court of law where that content has been published) to be redacted edited or removed.

Proactive monitoring is more challenging, legally and technically.

- Legally speaking, in the EU, such a requirement cannot be imposed on hosting providers (or access providers) as stated in Directive 2000/31/EC[81]. Furthermore, proactive monitoring would lead to a situation where the hosting provider chooses which contents are hosted. This might result in a legal view that such a provider is going beyond a purely technical role and would then prevent this provider from the benefit of the special hosting-providers' liability-regime organised in Directive 2000/31/EC and national laws that implement it. Therefore, such monitoring is not legally desirable for the provider, unless it is performed

[80] https://www.internet-signalement.gouv.fr/PortailWeb/planets/Accueil!input.action.

[81] Article 15 of Directive 2000/31/EC. The European court of justice detailed that, for instance, a national court cannot issue "an injunction against a hosting service provider which requires it to install a system for filtering information which is stored on its servers by its service users, which applies indiscriminately to all of those users, as a preventative measure, exclusively at its expense, and for an unlimited period, which is capable of identifying electronic files containing musical, cinematographic or audio-visual work in respect of which the applicant for the injunction claims to hold intellectual property rights, with a view to preventing those works from being made available to the public in breach of copyright". See EUCJ, judgment of 16 February 2012, case C-360/10, Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v. Netlog NV, http://curia.europa.eu/juris/document/document.jsf?docid=119512&doclang=EN.

without any subsequent obligation and judicial impact on the provider's liability[82]. As best practice (where financially sustainable for the provider[83] and after the content has started to be publicly hosted[84] and where proactive monitoring is the more effective one in terms of least impact on other Internet users' freedoms[85]) with the cooperation of victims that accept any liability caused by of the removal of the specified content[86].

- It is logistically and technically impossible for all content to be reviewed by humans before it is published on the internet. Furthermore, human review would be subject to error and might lead to the blocking of *legal* content[87]. Where child sexual exploitation material (CSAM) is concerned, there have been technologies which have been developed in collaboration between law enforcement, hotlines and industry to develop known digital signatures of known illegal images and videos. Using these signatures it is possible to automatically detect those specific files and prevent sharing and distribution on global websites[88]. Of course new images are not identified by this system. Where this system works well with known content types such as images and videos, it is more currently challenged in the area of written speech due to the variety of languages, cultural abbreviations, spoken word and colloquial phrases used to hide illegal hate speech, and, at the same time, more or less explicit speeches used to denounce illegal speeches on information website or even art including movies or theatre storyboards, which must not be censored. It is therefore a very difficult - even impossible - process to automate the identification of illegal hate speech, and specifically illegal hate speech content. However, some hosting providers do try to implement such mechanisms (for example, Twitter has developed an algorithm which aims to detect content that incites to hatred and violence[89]).

A variation of such moderation is the installation of time-lapse for content reviewing which involves a delay into the publishing cycle when content is not immediately published. Instead, a short lapse time between its writing and its publication in order for the user to review the

[82] This is important, in order to preserve both information society services and hosting providers' volunteering to perform such kind of control.

[83] The need for such a sustainability is clarified in the EUCJ court case SABAM v. Netlog, op. cit.

[84] In some countries, the a priori monitoring (i.e. the monitoring before the content is published) leads to the inapplicability of the special hosting providers' liability regime (jurisprudence which might be considered as regrettable) in relation with discussion forums. See for ex. in France TGI Paris, référé (= interlocutory proceedings), 18/02/2002, SA Telecom City, Monsieur J. M. et Monsieur N. B. c/ SA Finance.net ("Boursorama" case) http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=200; TGI Lyon, 21/07/2005, 14ème ch. du tribunal correctionnel, Groupe Mace c/ Gilbert D., http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=1589; CA Versailles, 12/12/2007, Les Arnaques.com c/ Editions Régionales de France, http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=2118; see also Forum des droits sur l'Internet, recommendation, "Quelle responsabilité pour les organisateurs de forums de discussion sur le web ?" (What kind of liability for discussion forums organisers?), 08/07/2003, http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/044000213.pdf, p. 229.

[85] It must be reminded that any ISP's action on Internet contents must be limited to the strict necessary in accordance with articles 8 and 10 of the European Convention on Human Rights, since any content blocking or removal consist of a limitation of either the freedom of expression or the freedom of exercising private life online.

[86] A good illustration is the cooperation between hosting providers and rights holders in the purpose of creating fingerprints of intellectual works which rights holder prohibit the broadcast of, and which hosting providers endeavour to prevent the distribution of on this basis. For court decisions validating such an approach in France see for example TGI Paris, 13 May 2009, L'Oréal et autres c/ eBay France et autres, http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=2639; TGI Paris, 24 June 2009, 3ème ch., 3ème section, Jean-Yves Lafesse et autres / Google et autres, http://www.legalis.net/spip.php?page=jurisprudence-decision&id_article=2682; TGI Strasbourg, 20 July 2007, SAS Atrya c/ Sté Google France et a., http://www.foruminternet.org/specialistes/veille-juridique/jurisprudence/IMG/pdf/tgi-str20070720.pdf.

[87] See for example "Orange blows up French govt website in terrorism censorship snafu", 16/10/2016, The register, http://www.theregister.co.uk/2016/10/18/orange_blows_up_french_gov_website/; "Silent email filtering makes iCloud an unreliable option", 28/02/2013, http://www.macworld.com/article/2029570/silent-email-filtering-makes-icloud-an-unreliable-option.html.

[88] https://www.interpol.int/Crime-areas/Crimes-against-children/Internet-crimes (last accessed 20-Oct-2016)

[89] https://support.twitter.com/articles/18311 (last accessed 20-Oct-2016)

content and, in some cases, to evaluate whether or not it complies with national legislation and, where appropriate, ISP policies.

It might be a huge societal evolution if internet technologies could take proactive action to prevent and delete illegal hate speech on internet servers while simultaneously protecting personal data, guarantee privacy and fundamental rights to free speech (which would be directly impacted by unfettered proactive monitoring). There are contradictory opinions that believe that crimes should be punished and not hidden[90], and that combatting hatred rather demands education and adapted public policies[91]. The MANDOLA partners are not entitled to settle such a debate, which reflects a debate between two societal choices, but to expose best practices that appear to be the most compatible with the principles of protection of human rights, as objectively as possible, in order to feed this interesting debate which should take place before Parliaments.

Beyond recommendations to policy makers that will be issued at the conclusion of an analysis of the legislation of illegal hate-speech and of an analysis of the applicable legal and ethical framework on privacy, personal data and protection of other fundamental rights, the Mandola project is developing a monitoring dashboard which will endeavour to derive and visualize large-scale statistics of the spread of potential on-line hate-related speech via social media (such as Twitter) and the Web (such as Google)"primarily in order to gain a better idea of their extent. A smart-phone application will also be developed. It will allow anonymous reporting of potentially hate-related speech materials noticed on the Web and in social media..

It is moreover to be noted that, in order to protect vulnerable adults or minors against harmful content (without distinguishing between legal and potentially illegal ones), several software applications enable the blocking of contents on each personal computer, and to adjust the setup to the user's specific needs.

5.2.8 Filtering, rating systems and removal tools

Where legally possible[92], filtering techniques might be applied in order to prevent access to illegal content (and this can also be applied to harmful material which can be considered illegal when affecting vulnerable persons including minors and vulnerable adults). Effective filtering techniques attempt to pursue the objective of early detection of illegal content, comments or images. The primary objective of Internet filtering is that content is prevented from reaching a personal computer, computer display or mobile device by a software or hardware product which reviews all Internet communications and determines whether to prevent the receipt and/or display of specifically targeted content.

---

[90] See European Digital Rights, "Internet blocking - crimes should be punished and not hidden", https://edri.org/wp-content/uploads/2013/12/blocking_booklet.pdf.

[91] See for example Iginio Gagliardone, Danit Gal, Thiago Alves, Gabriela Martinez, Countering online hate speech, UNESCO, 2015, especially pp. 46 and s., http://unesdoc.unesco.org/images/0023/002332/233231e.pdf.

[92] See our discussion under Section 5.2.7. Access providers do also benefit from a specific liability regime in so far they do not "modify the information contained in the transmission" (art. 12 of Directive 2000/31/EC), in addition to the fact that they are submitted to a principle of neutrality toward the contents they transmit. According to Article 15 of the Directive, they cannot be subjected to a general obligation of monitoring (blocking implying monitoring all contents). On this discussion, see Cormac Callanan, Marco Gercke, Estelle De Marco and Hein Dries-Ziekenheiner, Internet blocking - balancing cybercrime responses in democratic societies, October 2009, available at http://www.aconite.com/blocking/study (French version available at http://juriscom.net/2010/05/rapport-filtrage-dinternet-equilibrer-les-reponses-a-la-cybercriminalite-dans-une-societe-democratique-2/), especially from Chapter 6.

For example, in practice (without taking position on the legality of such actions, which will depend on the circumstances and primarily on who decides to implement and to perform the filtering) an email might be filtered (or classified/redirected into a specific folder) because it is suspected to be spam, a website might be blocked because it is suspected of containing malware or a peer-to-peer session might be disrupted because it is strongly suspected of communicating child pornographic content. In practice the filtering measure can be implemented at several levels, including at the computer level by the end-user directly.

Such filtering activities implicitly includes the automated monitoring of content and might also imply the use of content rating systems, depending on the exact location that the filtering is implemented (network, hosting platform, or personal computer).

Filtering content in a network is generally adopted when the content cannot be easily removed from the internet – perhaps because it is hosted in a location where the content is not considered to be illegal. As already analysed, such action might not be authorised in the EU except when demanded by a court or administrative authority, protecting the measure with sufficient guaranties that ensure the rule of law. In addition, blocking attempts against illegal content does not remove the content from its server, and does not prevent the same content from being duplicated at a different webpage or site. As a result, there are many ways to circumvent internet filtering[93] and blocking depending on the technique used to implement the filtering[94].

Filtering emails without delivering them (and without notifying the sender of the non-transmission) might not be authorised at all regarding email providers, since correspondence is highly protected in EU member States, and correspondence interception is most of the time a penal offense.

Filtering content on a hosting platform is also sensitive from a legal point of view, since -as previously analysed in this document - it might lead to a lack of application or validity of the liability regime specific to hosting providers. It might therefore have the undesired effect of activating the actual liability of the provider in several areas. This can also be the case if the hosting provider is not strictly compliant with the obligation of hosting providers to remove properly notified obviously illegal content, or if not required by a court or administrative authority which can protect the filtering initiative with sufficient protections that safeguard the rule of law. Indeed, most hosting platforms are a "*tool for expression and communication between individuals*" and "*for direct mass communication or mass communication in aggregate*"; they also "*offer great possibilities for enhancing the potential for the participation of individuals in political, social and cultural life*"; for these reasons, a "*lack of legal, and procedural, safeguards surrounding processes that can lead to the exclusion of users*" or the removal of a content might threaten the "*right to freedom of expression and information, as well as the right to private life and human dignity*".[95] Moreover, removing illegal content does not prevent the same content from being duplicated at a different webpage or site, including hosted in more permissive countries.

[93] https://freedomhouse.org/sites/default/files/inline_images/Censorship.pdf (last accessed 14-Oct-2016)
[94] http://www.aconite.com/blocking/study (last accessed 14-Ocr-2016)
[95] All the quotations of this paragraph are coming from Council of Europe, Recommendation CM/Rec(2012)4 of the Committee of Ministers to member States on the protection of human rights with regard to social networking services, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa9b.

Filtering content on a computer system at the end-user initiative, or on platforms or networks that enable a differentiated filtering depending on each specific end-user request, might not create legal issues since the access restriction is decided by the end-user itself. Such solutions are generally based on content rating systems. An example is the automatic rating of emails that may that way be classified in a specific folder if the end-user configures its software in this sense.

In any case, a set of actions and protocols must be determined beforehand, including the determination of key persons that will be responsible for implementing the protocol, supervising the effectiveness of the applied filtering measures when prohibited content has been encountered, and respond to complaints in cases if content that should be authorised is blocked. The question of liability assessment (identifying the party responsible and the extent of liability) for blocking unauthorised content (i.e. legal content where the blocking takes place outside the request of a concerned end-user), or alternatively in situations when unauthorised content remains available and accessible, must also be identified. These protocols should also determine the procedure under which valid notifications will be reviewed and under which content will be removed or blocked, and the time frame under which these actions will take place. Courses and training sessions for ISP's staff might have to be periodically scheduled so as to raise awareness about the combat against illegal and -if applicable- harmful content, including hate speech prevention and counteracting, and about the legal impact of implementing such actions.

When blocking or removal is needed (especially when lawfully required by a court or an administrative authority) some Service Providers might not have the technological or financial capabilities (or even the legal support structures) to implement such blocking or deletion. In such cases support measures might be needed. For example, according to some authors, combining efforts by creating private-public partnerships could balance out the overwhelming costs of implementing technologically advanced tools in filtering and removing harmful contents[96]. Moreover, as already exposed, end-users can always employ firewalls[97] or software in order to filter out websites that include illegal content.

The most appropriate balance between the protection of victims of hate speech and democratic freedoms is a very complex issue which needs to be finally determined on a national level through extensive debate among relevant stakeholders in each country and with regard to relevant binding international instruments such as the European Convention on Human Rights., the cybercrime convention and the optional protocol.

5.2.9 Reputation scoring systems

The reputation scoring system is a significant trend which is developing over the years, and which offers the opportunity for parties to rate each other, providing an active bidirectional way of sharing content and opinion, by encouraging good behaviour[98]. This system of rating the reputation of a service or website in relation to the presence or absence of harmful content could also help the ISP (where legally acceptable) or the filtering companies that provide end-

[96] Awan & Blakemore (2012). Policing cyber hate, cyber threats and cyber terrorism. published by Routledge,; ISBN: 1317079124, 9781317079125

[97] Banks, J. (2010). Regulating hate speech online. International Review of Law, Computers and Technology, 24(3), 233-239.

[98] Josang, A., Ismail, R., & Boyd, C. A. (2007). A survey of trust and reputation systems for online service provision. Decision Support Systems, 43(2), pp. 618-644

users tools to monitor and detect this type of content more rapidly. Such reputation scoring systems empower users and provide a way for users to indirectly communicate experiences with each other.

5.2.10 Additional issues

Internet Anonymity

Internet anonymity is a complex technical, ethical and legal challenge. Whereas there are systems which attempt to protect the user's identity from disclosure, it is quite difficult and rare to achieve absolute anonymity. Using anonymising techniques and tradecraft increases the complexity for, and resources required by, the investigator to uncover the true identity of the person seeking anonymity. Even if true identification of the subject is increased, it might be difficult to identify the specific location or the specific individual at a location that is responsible for published content. This usually requires authorised state agencies to complete the investigation using standard investigation techniques.

It is difficult to achieve internet anonymity and in most cases it is possible to use open source intelligence combined with authorised data disclosure requests to internet companies to locate a range of internet electronic evidence including comprehensive contact details of the user, the location and the devices used to publish content on the Internet. However, this combined comprehensive data is more likely to be available to authorised state agencies than other internet users or internet companies. All this personal data is subject to data protection legislation in Europe and the data controller is responsible for protecting that data and for ensuring the data is accurate.

It is an ethical question for society to determine whether anonymity is a fundamental human right and what protections users are entitled to have in hostile communities or countries. It is not appropriate to assume that all anonymous users are intending to commit a crime and there can be many legitimate reasons for users to desire anonymity.

Licensing and Regulatory requirements

Many internet services do not require a state license or may not be regulated. Where such oversight might exist, it might be be the responsibility of a foreign agency with different community standards or beliefs.

Some countries do have strict regulations governing internet services and responsibilities where users and state agencies can expect specified standards of service and responsibilities in relation to internet crime.

Time-lapse for content reviewing

Time lapse content reviewing is a technique where content being published in an online space is created and reviewable by the author but where publishing of the content to the wider public is not immediate but is time-delayed. For example, it has been noted that when content is hurriedly authored and includes an emotional rebuttal of an opinion or a new allegation or accusation that the content might be overly emotional and possibly even intolerant or hateful. Delaying the publication can offer the author time to maturely reflect on the nature of the

content and whether the content accurately reflects their true position and belief on the topic under discussion. Regardless, the material is not automatically blocked or deleted and will be automatically published after a known pre-defined short period of time. However, the content can be reviewed, updated, edited, corrected or even voluntarily deleted (exclusively by the author) before that time delay has expired. The intention of this system is to minimise unintended overly emotional, aggressively harmful or illegal outbursts on public media by empowering the author to review their own content.

The disadvantages with this process are that it can be accused as supporting a chilling effect on free-speech and potentially limits the spontaneity of the discourse and candid exchange of views. The delays incurred can also disrupt the smooth flowing of the exchange of views since many messages from different participants might be simultaneously caught in a time-lapse delay and arrive out of natural sequence or at a time when the conversation has moved on to other topics.

In different environments, where permitted or possibly legally required or to protect vulnerable groups, this delay period might be used by content moderators to review content before approval for publication.

6 Guide for Internet Service Providers (ISPs)

In Europe, ISP liability and obligations towards illegal content are regulated by several EU Directives and primarily Directive 2000/31/EC, as well as by national laws that implement these directives and may add additional provisions. Outside this legal framework, that has been created in order to facilitate the combat against online illegal content without infringing other freedoms in an unnecessary or disproportionate way (such a necessity and proportionality being requirements of the European Convention on Human Rights), any private censorship might violate the applicable legal system and engage the liability of the ISP.

This best practice guide takes this legal framework into account and endeavours to identify best practices that respect this framework while complementing it in a creative way, for the benefit of responding to online illegal or potentially illegal speech. The EC Code of Conduct is relevant to this section and the contributions and outputs from the Mandola Advisory Board meeting (Brussels, October 2016) and from the Mandola workshop (Brussels, December 2016) with participation from relevant sectors contributed to the development of these best practice guidelines.

It is addressed to a wide variety of ISPs, which encompasses all Internet service providers, including those that broadcast electronic communications (such as access providers and email providers). It also includes those that store electronic communications for the purpose of the functioning of the Internet (such as caches) or on the request of an Internet user (such as hosting providers. It includes other associated stakeholders such as Internet fora, online encyclopaedias and social networks). Each type of ISP has different levels of technical and legal capabilities to respond to online hate speech.

In the first instance, potentially illegal content should be reported to the relevant authorities and if it is possible and legal according to local laws relevant evidentiary images or text, and proof of the event should be provided. This might include, information about its origin, information regarding the user who is potentially responsible for posting the content which is available to the public or that have been transmitted by the reporting user[99]. It is important to only transmit the contact details and identity of the reporting users with their consent (rules relating to the storage, use and transfers of such data must be very clear in the protocol rules available to the public and on the page that hosts or provide the reporting mechanism).

6.1 Develop and establish a code of conduct

There are initiatives that support Codes of Conduct (CoC) for ISPs in order to describe their role in the marketplace as described in Sections 5.2.4 and 5.2.5. Whereas, these CoC help coordinate the Internet industry response on a range of issues they also describe how ISPs should manage online hate and similar material located on their servers. Under certain conditions, in order to comply with the legal framework, some types of ISPs who possess the technical and legal capabilities, may voluntarily agree to prohibit users from communicating potentially-illegal

[99] In the EU, LEA access to traffic and connection data retained by ISPs requires a specific order and cannot be provided outside any specific and lawful request.

online hate speech using their services. Of course, the CoC needs to be created in compliance with legal requirements, including the EU Telecom package, the EU Directive 2000/31/EC and the CoE European Convention on Human Rights. The provisions of these terms of conduct might also need to be adapted to applicable national provisions, which might provide for additional obligations or prohibitions that are not included in the EU law.

6.2 Develop and establish clear terms of service for customers

In addition, some types of ISPs might want to establish their own Terms of Service (see section 5.2.5) regarding online hate speech. In such case, they must firstly respect their legal obligations (in terms both of contribution to the response against crime and of preservation of fundamental rights - which might require the neutrality of services provided).

Once these TOS are adopted, they need to ensure their application, which implies the need to establish application protocols and to monitor the effectiveness of such protocols. Consequences might be, *inter alia* and provided that the protocols comply with the legal framework identified above, the blocking or removal of content (knowing that the ISP might have an additional obligation to ensure the back-up of the deleted information for investigation purposes), the suspension of a user's account, or the prevention of an offending user from participating in a chat room, depending on the services provided by the ISP and technical capabilities they have available.

In any cases, the application of a code of conduct or of terms of service might have to be done in cooperation with competent public authorities. Indeed, the removal of content or the prevention of an end-user from posting comments may endanger ongoing administrative, civil or criminal investigations. Therefore, a prior agreement on the actions to be performed might need to be agreed with Law Enforcement or Prosecutors and included in the protocols described in the CoC or ToS. Such actions might include content removal (after back-up of all the necessary information) or non-removal of offending content before ad hoc agreement, etc..

Codes of conduct and terms of services must in addition be easy to understand, should be comprehensive, should provide lawful, useful and realistic answers, and should include - in the extent that is possible - the ability for the service provider to implement protocols in relation to potentially-illegal online hate speech. This document or an associated one must also provide for an operational contact point in case of complaint against a measure taken by the ISP. This document or an associated one must moreover make clear the methods whereby personal data is processed, notably in case of report of a potentially illegal content (in relation to storage, use, transfer, etc of personal data in compliance with the data protection legislation.

Furthermore, the code of conduct that is applied and the terms of service created for the internet service should be provided for in a way that users can read it the first time they access the page, and it also should be available to consult at any time. It should also be posted in different languages for foreign users.

6.3 Develop and establish an action protocol when illegal content is reported

It is important that service providers have prepared a set of actions and protocols to apply when responding to notifications about illegal content, as well as when responding to complaints against responding measures taken by the service provider.

This protocol should identify key persons that will be responsible for implementing the protocol and supervising the effectiveness of the applied measures. This protocol is described in more detail in sections 5.2.4 and 5.2.5.

6.4 Develop internal procedures and staff training in relation to recognising potentially illegal online hate speech.

Mechanisms which pursue the objective of recognition of illegal content including comments and images are needed to support the staff assigned to this area. It provides the responsible staff with the capabilities and hands-on training to review valid notifications received and to act in consequence. This might imply - in accordance with law and the protocol's rules - to remove clearly identifiable illegal hate speech or to disable access to such contents including notification to law enforcement when legally appropriate. Where relevant, courses and training sessions for staff should also be periodically scheduled so as to raise awareness about hate speech prevention and response.

6.5 Cooperate with notifications received from hotlines

It is good practice (where not legally mandatory such as in France[100]) to cooperate closely with hotlines or to support a hotline for users, so they can report content that they suspect to be illegal. The EC Code of Conduct refers these agencies as “trusted reporters”.

When accessing this reporting mechanism, it should be made very clear if reports can be given anonymously or not, and the way transmitted personal data will be processed, in compliance with data protection rules.

In addition, ISPs are encouraged (to the extent it is possible and relevant due to the nature of their exact activity) to provide advice on filtering tools and rating systems for Internet users.

6.6 Act promptly and professionally when potentially illegal online hate speech content is detected

In case of notification of the presence of a potentially illegal content on the Internet, ISPs should act promptly in accordance with law and their protocols. This might imply, as already said in relation with the action protocol, to forward the report to local LEAs (which will generally be the case for access providers), to remove or prevent access to this content (taking care, once again, of respecting the neutrality obligation impacting most categories of service providers and of non being prejudicial to ongoing or future investigations). To be noted that removing or blocking illegal content may not prevent the same content from reappearing at a different location and it won’t be effective in deterring the content from being hosted by more permissive hosting providers in the marketplace.

[100] Article 6 of law 2004-275.

In order to be able to contribute to the investigation of the event afterwards (by responding to judicial requests in this sense) the ISP should save the information it is authorised by law to save. This might also help, in certain situation, to determine the origin and the author of the online hate speech. Authorities that carry online hate speech investigations will need all the evidence they can gather in order to carry out their work, protect the victim's rights and locate the perpetrator. Collaboration with authorities - in the respect of the legal framework, and notably the production of some evidence on official lawful requests only- is essential when tackling hate speech.

6.7 Seek legal advice in order to be aware of rights and responsibilities, international and national regulation regarding online hate speech and its updates.

Service Providers that do not have legal departments should become familiar with international and national binding regulation regarding illegal online hate speech (and illegal content in general), as well as with the principles governing their liability as (mainly) access or hosting providers, their obligations and rights toward potentially illegal content, personal data and electronic communications of others. In this case it is advisable to seek legal counsel.

Mandola has developed documents as part of the project which can help everyone understand the legal definition of illegal online hate speech and understand colloquial sentences used for hate speech behaviours in different countries[101], and the civil and penal liability of service providers in relation to electronic communications that are broadcasted or stored at the initiative of the users of their services.

6.8 Support Counter Speech

The benefit of counter-speech is outlined in Section 5.2.3 and is a very effective response to online hate speech since it also empowers and educates community members. Service providers should support and encourage activities which counter online hate speech by responding to hate speech with alternative more balanced speech

6.9 Minimize the risk that content placed on a website you own and control does not fall into the category of potential hate speech.

In relation to websites established and controlled by the ISP itself (on behalf of the ISP company and not on behalf of its customers), the ISP should be aware and constantly review content and updates created and edited before publication on the ISP website. To better understand this recommendation, it should be emphasised that this is not targeting Internet spaces for which the ISP is acting as a hosting or as access provider, but content for Internet spaces (website, webpages, web banners, etc.) which is directly provided or published by the ISP under its own responsibility - and for which the ISP therefore acts as an editor or author. It important in such situations that websites owned and operated by the service provider are not subverted to promote messages of hate speech.

In cases where such web spaces enable the publication of new content by Internet users themselves, and in cases where the ISP intends to remain editor in relation to these new

[101] http://www.mandola-project.eu/m/filer_public/7b/8f/7b8f3f88-2270-47ed-8791-8fbfb320b755/mandola-d21.pdf

contents, another recommendation is to ensure that users and visitors are informed about what is allowed and what is specifically forbidden by ensuring clear and consistent Terms of Service and widely promoting this document.

6.10 Review: Assess, evaluate and update

It is important that the best practice activities identified, the code of conduct and the terms of services are regularly assessed for their relevance and effectiveness and that they are updated to reflect current best practice, technologies or changing work practices.

7 Role of Internet users and potential victims

The primary approach to internet users is to provide assistance and advice about staying safe online and how to use the Internet safely. However, due to the scale of internet services it is impossible to avoid encountering insulting or abusive content online and sometimes such content might be considered as illegal hate speech.

Although, anyone could be a victim of hate speech, we know from experiences that there are individual persons or groups (such as minorities) more likely to be the targets of hate speech. This can be triggered due to the current socio-political context or the current existing conflicts between religious beliefs or personal morals; or due to their distinct race, colour, descent or national or ethnic origin, as well as due to their religion, gender, or sexual orientation.

The Internet provides anonymity or pseudo anonymity and access to a potential global audience, a high percentage of whom are young. Since youngsters and adolescents are digital natives and spend a lot of time on the Internet, they are more likely exposed to online hate speech than other sectors. Awareness-raising campaigns that target youth, along with counselling and intervening when small incidents occur is vital to educate those who perpetrate online hate speech or who are likely to do so. Education should stimulate greater understanding about the types of behaviours that are considered hate speech, but also address the consequences of these actions, both legal and social. Promoting tolerance, respect and human rights are key issues when tackling hate speech online.

In the first instance, potentially illegal content should be reported to the relevant authorities and if possible and legal according to local laws, provide evidentiary images or text, and proof of the event, information about its origin, information regarding the user who is potentially responsible for posting the content and that are available to the public or that have been transmitted by the user at the origin of the report[102]. The relevant authorities will treat information received with respect and will usually ensure to only transmit the contact details and identity of the reporting user having received informed consent (rules relating to the storage, use and transfers of such data should be very clear in the protocol rules available to the public and on the page that hosts or provide the reporting mechanism).

The practices selected are a compilation of best practices researched in several places and from which Mandola retains what seems to MANDOLA partners to be the most relevant. Mandola believes that the list below is a list of best practices for users when encountering hate speech, especially when they have been the subject of an attack. It also provides information so as to minimize its consequences for the target of the attack, as well as to raise awareness and tackle this phenomenon. We strongly support and promote tolerance and mutual respect which is often regards as a fundamental response to online hate speech.

[102] In the EU, LEA access to traffic and connection data retained by ISPs requires a specific order and cannot be provided outside any specific and lawful request.

7.1 Stay calm and keep a level head.

If you encounter online hate speech, try to remain calm and analyse the text or images so as to understand what upsets you and to be able to address whether or not it is potentially illegal online hate speech. Mandola has produced information about legal definitions of illegal hate speech in D2.1 on the Mandola website (www.mandol-project.eu).

7.2 Avoid responding to the attack.

Do not engage in discussions or fights with the author of the comments or images because it would fulfil their desires, provide attention and, possibly, fuel their anger.

7.3 Back up the event with documentary evidence.

In order to be able to prove the event afterwards (before a court or so as to help police with their investigation, for instance) and when it is legal to do so, create and save screen shots and any other proof of the event. This could also help determine the origin and author of the online hate speech. If possible, ask a court registered/experienced expert to perform this back-up since the evidence produced by a registered/experienced expert will have a stronger legal force.

7.4 Reconfigure your privacy settings and block the author of the comments or images.

Be aware of your privacy and avoid intruders accessing personal information on your profile. Privacy settings enable you to block strange contacts or persons who can send you personal messages or access your online contents.

An official investigation might require that the blocking is not performed in order to enable an effective investigation of the case therefore in extreme cases it is not advisable to block the person without prior agreement with investigators.

7.5 Value online privacy

Be conscious of the information you provide online (such as images, address, habits, usual locations, vacations, etc.) and ensure that strangers do not have access to this data, as well as unable them to keep records of your information.

7.6 Contact local authorities and law enforcement agencies and report the event.

Reporting to authorities should be done before reporting to the ISP if the person is a victim. The best report for a victim is a report to Law Enforcement at the national level or a letter to the national public prosecutor (under the reserve of local legal specificities of course).

7.7 Contact the website owner or the Internet Service Provider (ISP) or use the reporting mechanism/hotline provide by the website or nationally and report the event.

The most efficient reporting procedure is to report to national authorities or to a national hotline which will have close cooperation with local authorities, if the purpose is the prosecution of the perpetrator. Indeed, depending on their terms of service, a report to a websites or a social provider might lead to the destruction of the evidence that will be needed to conduct the investigation.

If the aim is only to see the removal of the content, a report may be done to the website that hosted the potentially illegal content, and if possible or inefficient to the hosting providers who might be able, depending on the service he provides and depending on local law and on his terms of service, to remove the content or at least temporarily interrupt the access to this content. . This mechanism usually enables to communicate, inter alia, with moderators and website operators in order to report and highlight online hate speech; it usually provides advice on filtering tools and rating systems for Internet users. You can find a list of social media/websites and reporting mechanisms on http://reporthate.eu/report-hate/.).

7.8 Cooperate and provide authorities with all the information regarding the event, as well as with the evidence you have been able to gather so as to help with the investigation.

Authorities that are responsible to investigate online hate speech allegations need all the evidence they can gather in order to successfully complete their work, protect the victim's rights and locate the perpetrator. Usually, the information that is provided to law enforcement agencies and authorities consists of screen shots, downloaded images and texts. Further information is obtained by carrying out forensic analyses on devices owned by the victim and the alleged perpetrator.

To be noted that in order to respect the presumption of innocence and to avoid to engage his or her own liability (for example for abusive denunciation, depending on the applicable local law), it is advised to the author of the report to be honest and to only relate facts, without accusing namely a person in case of doubt.

7.9 Seek legal counsel in order to be aware of your rights and the proceedings to take legal actions.

An Internet user has rights which should be observed and respected, as well as obligations and potential liabilities in case a report is guilty of deliberate malicious prejudicial reports against someone else. In case such rights are violated, it is advisable for internet users to seek legal counsel in order to be informed about rights and legal actions that can be taken so as to support a prosecution and sanction of illegal actions, while protecting and minimizing personal liabilities. Internet users can find the legal definition of hate speech, including a list of penalties for hate speech behaviours in different countries; and the legal actions that users can take on the Mandola website.

7.10 If needed, ask for help via contacting the victim's help phone or webpage.

Anyone can be a victim of online hate speech due to different reasons. If you have been a victim, you are not alone. Mandola wants to raise awareness regarding hate speech and to help people who have encountered it by providing them with tools and advice on how to respond to online hate speech and to better protect themselves against it. You can find a list of contacts for victims of online hate speech on the Mandola Website[103].

[103] http://www.mandola-project.eu/links/

7.11 If needed, contact victim's associations in order to ask for help or counselling.

There are also many associations and groups that support victims of online hate speech, such as the ones that you can find in this extensive list on the Mandola website[104].

7.12 Do not blame yourself for what happened and do not let the event undermine your self-esteem.

The only one to blame for hate speech is the perpetrator. Anyone can be a victim of hate speech. Often a subject justification is proffered for online hate speech but it is important to recognise that there is no objective or scientific reason or motive that justifies hate speech or the behaviour of insulting or degrading another person due to their difference, which might lie in their ideology, religion or beliefs, ethnicity, race or nation, gender, sexual orientation, family situation, illness or disability, among others. Such behaviour constitutes a denial of human rights as they are declared and protected by the European convention on human rights and the Charter of fundamental rights of the European Union (amongst other international and national texts), and are most of the time illegal, either on the basis of the prohibition of hate speech, or on the basis of the prohibition of violations of human dignity or of other personal rights of the person. It also goes against the Universal Declaration of Human Rights.

[104] ibid

8 Outstanding Issues

This document provided a short introduction to the Mandola project and the consortium. This was followed by a short description of the current landscape and the role of legislation, policy and regulation. The document provided a brief focus on the potential role of industry and then offered some practical guidelines for internet users and potential victims of online hate speech.

Further consultation with industry, law enforcement and civil society stakeholders would be helpful to gain further insights into the challenges faced by service providers and what obstacles need to be addressed in order to improve the response and effectiveness to online illegal hate speech.

8.1 Next Steps

The current document is the first version of the Best Practice Guide to Online Hate Speech document.

The results and feedback received from the WS4 workshop hosted in December 2016 will support the dissemination of the best practice guide

A survey is being planned to identify the experiences of all stakeholders in this area and report on the primary concerns.

The MANDOLA consortium welcomes community feedback and comment so that the advice provided can be improved.

9 Appendix I - Case Studies Handling Illegal Hate Speech by European Internet Hotlines

Internet Hotlines present a quick communication link for the public to report suspected illegal online content or activities under different categories. Their main objective is to fight illegal content distributed on the Internet and to forward their investigations to the relevant stakeholders in order to have this content removed promptly.

Internet Hotlines in Europe (and all over the world) follow a similar process: the hotline mechanism starts once a report has been submitted from the public -anonymously or with contact details- or from another hotline; the Hotline will ensure that the content is analysed, and if found to be potentially illegal, the information will be forwarded to the national Law Enforcement Agency (LEA) and/or the Internet Service Provider hosting the content. Generally, a prior agreement with the national LEA ensures that any content removal by the hosting provider is not prejudicial to an investigation (whether the coordination between LEA's needs in relation with the specific case and the hosting provider's action takes place directly between these stakeholders after the report has been made or through the intermediary of the hotline). When the illegal content appears to be located on foreign servers, the hotline will notify all or some of the stakeholders (depending on the prior agreements that have been made) of the relevant country, such as the hotline, the police or the concerned hosting provider. Investigating hate speech crimes is extremely complex due to the differences in national legislations around Europe.

This section comprises information offered by the Hotlines in Europe, which provide the option to report hate speech or equivalent (xenophobia, racism) among other types of illegal content and activity.

In Europe, usually only a Court of Law can determine that a criminal offence has been committed and that material relating to that offence is actually illegal. Therefore, a European hotline can only assess and determine that something is "probably illegal" with reference to the criteria given in the relevant national Law.

9.1 France - Point de Contact / PHAROS platform

Report: Includes Hate speech

Point de Contact

Point de contact is the French INHOPE member operated by the ISPs Industry from 1998.

Upon receipt of a report, Point de Contact checks that the reported content falls within its remit, and assesses whether the reported content is potentially illegal under French law. Content that is potentially illegal under French law is systematically:

- Located geographically by tracing the IP address
- Reported to the competent French authorities (Central Office for the Fight against Crime related to Information and Communication Technology, OCLCTIC)

- Notified to the hosting provider, if the content is located in France
- Forwarded to a partner of the INHOPE international network, if the content is hosted abroad, and in a country where such a partner is established.

The purpose of Point de Contact's action is to help remove illegal content from the Internet and to enable the judicial authorities, where appropriate, to quickly carry out investigations. Reports can be anonymous and such anonymity is ensured by the hotline. If the reporting person provides his or her email address, Point de Contact will automatically inform him or her of the outcome of his/her report.

Additional information

http://www.pointdecontact.net/
http://www.pointdecontact.net/about_us?language=en
http://www.pointdecontact.net/assessment_and_actions?language=en

PHAROS

This reporting mechanism is the French official portal for reporting Internet unlawful content. Reports are analysed by police officers and Gendarmes from the PHAROS platform (platform for harmonisation, analysis, cross-check and orientation of reports). This platform belongs to the central office for combatting crime linked to information and communication technologies (OCLCTIC - which in turns belongs to the central direction of the judicial police, component of the National police).

Reports are verified where possible, and if the content is potentially illegal, it is forwarded to the service which is competent for the given case, and a penal investigation might be started. Reports can be anonymous but the IP address of the reporter is saved, which might enable an identification (following strict legal procedures) in case the investigations needs it.

Additional information

https://www.internet-signalement.gouv.fr/PortailWeb/planets/Accueil!input.action

9.2 Germany - German Association for Voluntary Self-Regulation of Digital Media service providers (Hotline FSM)

Report: Includes other content harmful to young persons

The FSM examines incoming complaints about online content from a legal standpoint, assessing them with reference to the criteria laid down by German youth media protection law – the German Interstate Treaty on the Protection of Minors in the Media (JMStV). If FSM determines that there has been a breach, FSM approaches those responsible for the page or the public body with jurisdiction with the objective of having the content concerned removed.

Together with its member companies and associations, the FSM strengthens youth media protection and restricts online media content that is illegal or is harmful to young persons or impairs their development. To ensure a uniformly high youth protection standard, the FSM and its members have established voluntary commitments for various parts of the online world. The association offers regular members the option of adopting the regulated self-regulation model and calling on the FSM in the event of disputes with the KJM.

Additional information

http://www.fsm.de/en/about-us

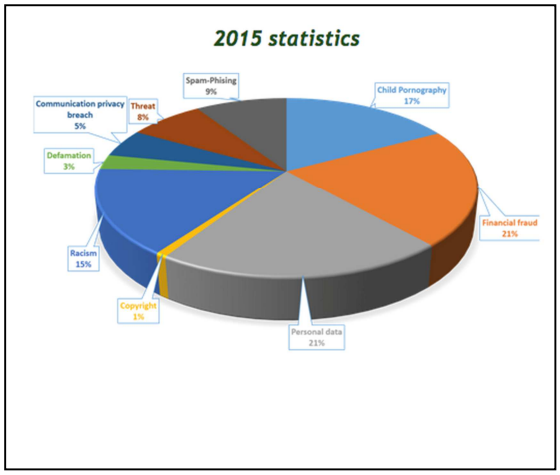
http://www.fsm.de/en/guide

Educational project entitled "Medien in die Schule" (Media to Schools) can be found here (http://www.medien-in-die-schule.de/).

9.3 Greece - Safeline www.safeline.gr

Report: Includes hate speech (racism, xenophobia, hate)

SafeLine collaborates directly with the Greek Police by forwarding reports that contain verified potentially illegal content. In Greece, there are is a specialised police unit that deal with criminal activities on the Internet: the Cyber Crime Unit in Athens (established 2004).



Additional information

http://www.safeline.gr/en/reports/how-make-report

http://www.safeline.gr/en/legislation/legal-framework

9.4 Hungary - Internet Hotline

Report: Illegal and harmful content, includes Online harassment, Racism, xenophobia

The Internet Hotline can be used to report potentially illegal content or content that may be detrimental or dangerous to minors. Categories which can be reported include paedophile content; abuse of images and videos, or other personal data; online harassment; Racism, xenophobia, portrayal of violence.

Once a report is filed the content provider of the presumed illegal content or the server is requested to delete the contested content. Article 7 of Act CVIII of 2001 on Certain Issues of Electronic Commerce Services and Information Society Services provides the legal basis for the deletion, pursuant to which the service provider (content provider) shall be liable for any infringement and damage caused by making available unlawful information. The Internet Hotline may only request the removal of the contested content by citing that it infringes legislation.

The Internet Hotline notifies the content provider or, if it cannot be reached, the service provider that provides hosting services for the infringing information, thereby providing notice of its potential liability if it does not remove the information. In its calls, the Internet Hotline also generally cites the violation by the uploading party of the civil law contract concluded between the latter and the service provider, as such contracts usually prohibit the uploading of unlawful information to servers.

In case of content hosted on foreign servers, the organisation notifies the hotline of the relevant country. If the illegal content may contribute to the perpetration or preparation of a penal infringement, law enforcement are also informed.

Additional information

http://english.internethotline.hu/operation/

9.5 Iceland - Barnaheill Save the children á Íslandi.

Report: Illegal content includes Other, ie: defamation, threatening or hate speech

A report button regarding the subject of illegal and improper material is operated jointly by Save the Children Iceland, SAFT (Society, Family and Technology), Home and School, The Directorate of Health, and The National Commissioner of the Icelandic Police.

Additional information

http://www.barnaheill.is/TilkynnaologlegtefniReportillegalcontent/Reportillegalcontent/

9.6 Ireland - hotline.ie Irish Internet Hotline

Report: Includes Racism, Other

The Hotline can only deal with reports of content that appears, or is distributed as a communication, on the Internet. Descriptions given in help windows of what constitutes potentially illegal content are for guidance only and are not legal interpretations of legislation.

Under the Irish legal system, only a Court of Law can determine that a criminal offence has been committed and that material (..) relating to that offence is actually illegal. Therefore, Hotline.ie can only assess and determine that something is "probably illegal" with reference to the criteria given in the relevant Irish Law.

The establishment of the Hotline.ie Service within an industry self-regulatory framework was one of the key recommendations of the Government's Working Group on the Illegal and Harmful Use of the Internet, in 1998.

Hotline.ie is run and funded by the Internet Service Providers Association of Ireland (ISPAI) whose members are determined to take measures to counter the use of their Internet facilities for such illegal purposes. It is currently in receipt of grant support from the European Commission under the Connecting Europe Facility Telecom - Safer Internet – Programme.*

The operations of the Hotline are overseen by the Department of Justice and Equality – Office for Internet Safety.

Additional information

https://www.hotline.ie/

9.7 Latvia - Dross Internets.lv

Report: Illegal content, includes hate speech, racism

All incoming information will be collected by Safe Internet Operator, who will examine individually every report and determine the level of harmfulness. If the content is harmful (to children), but not illegal, the administration of the site is being informed and is asked to evaluate the appropriateness of the materials for children/youngsters. However, if the reported breach is an infringement of law, and appears to be hosted in Latvia then the State Police is informed. The State Police continues to examine the specific case and inflict a penalty as stated by the law. In doubtful cases there are consultations with experts State children rights protection bodies, Latvian Centre for Human rights

Annual Statistics	2009	2010	2011	2012	2013	2014	2015	2016
Hate speech/ racism	51	21	23	15	9	15	4	7

Additional information

http://www.drossinternets.lv/page/32

http://www.drossinternets.lv/page/102

9.8 Lithuania - Safer Internet Centre Lithuania: draugiskasinternetas.lt

Report: Illegal content, includes inciting racial and ethnic hatred on the web

Nationally, the SIC have a mature and well-established, multi-stakeholder network, involving the public sector, private sector and civil society, therefore with the capacity to deploy services that help make the Internet a trusted environment for children (and citizens at large) through actions that empower and protect them online.

Additional information

http://www.draugiskasinternetas.lt/en/main/hotline

9.9 Luxembourg - BEE SECURE Stopline

Report: Includes racism, revisionism and discrimination

The BEE SECURE Stopline project aims to provide a structure enabling illegal Internet content to be reported anonymously and to deal with these reports in collaboration with the relevant national and international authorities. Through its website, BEE SECURE Stopline therefore offers the public a way of taking civic action against illegal content on the Internet.

As part of the management of illegal content reports, BEE SECURE Stopline has a collaboration agreement with the Police Grand-Ducale and acts as an intermediary and expert for the receipt, analysis and transmission of reports to the relevant departments of the Police Grand-Ducale.

This collaboration agreement only concerns transmission of information relating to illegal content reported.

Additional information

https://stopline.bee-secure.lu/index.php?id=11&L=2

9.10 Portugal - Linha Alerta internet seguraøpt

Report: Includes Incitement to Racism, Incitement to Violence

The mission of Linha Alerta is to block illegal content on the Internet and prosecute their disseminators in an effective way. These objectives may be achieved by providing the Portuguese law enforcement agencies with collected information in order to facilitate the elimination of the illegal content and the identification of those responsible for these materials and by means of collaboration with national Internet Service Providers and international counterparts in the fight against illegal contents.

In order to carry out this core activity Linha Alerta has a website – linhaalerta.internetsegura.pt – where one can send complaints, which may be anonymous. This service is provided in Portuguese and English. The staff are bound by professional secrecy Linha Alerta will address the following illegal contents:

- Child abuse images;
- Incitement to violence content; and
- Incitement to racial hatred content.

The scope of Linha Alerta may be subject to further revisions according to the experience gained with this first stage of operation and the legal framework developments in Portugal.

Additional information

https://linhaalerta.internetsegura.pt/index.php?option=com_artforms&formid=4

https://linhaalerta.internetsegura.pt/index.php?option=com_content&view=article&id=15&Itemid=47&lang=en

Portuguese Legislation on Racism

Incitement to Racism includes any content that incites to hate or violence or racial or religious discrimination with the intent of encouraging it. With racist, revisionist or Neo Nazi speeches, thousands of sites, blogs and other virtual communities disseminate racial hatred and intolerance. These behaviours are punished in the Penal Code and the principle of equality is stated in Article 13 of the Constitution of the Portuguese Republic[105].

[105] How it Incitement to racism criminalized?
Article 240 - Racial, religious or sexual discrimination

9.11 Slovenia - Spletno oko

Report: Includes Hate speech

Hotline Spletno oko enables internet users to anonymously report hate speech (and child sexual abuse images while they detect online). The mission of hotline Spletno oko is in cooperation with Police, Internet Service providers, and other governmental and non-governmental organizations to reduce the volume of child sexual abuse images and hate speech online.

- Hotline operation, which enables anonymous report of illegal content on the internet
- Awareness raising about illegal online content
- Fast and effective analysis of received reports
- Cooperation with other hotlines around the world, to exchange reports and best practices

Members of Advisory Board of Safer Internet Centre are also The Office of the State Prosecutor General of the Republic of Slovenia, General Police Directorate, representatives of media and representatives other organizations, which are active on the field of child welfare.

Additional information

http://safe.si/en/spletno-oko/hotline-spletno-oko

1 – Who:
a) Establishes or constitutes an organization or develops organized propaganda activities that stir up to the discrimination, hater or violence against a person or a group of people because of their race, color, ethnic or national origin, religion, sex or sexual orientation, or that encourages it; or
b) Takes part in the organization or any other activities mentioned in the previous point or gives assistance to them, including financial support;
will be punished with imprisonment for a term from 1 year up to 8 years.
2 – Who, in a public meeting, by a writing document intended to be publicly disclosed or through the media or by any computer system intended to the disclosed:
a) Practices any violent act against a person or group of people because of their race, color, ethnic or national origin, religion, sex or sexual orientation; or
b) Defames or insults a person or group of people because of their race, color, ethnic or national origin, religion, sex or sexual orientation, namely by denying war crimes or crimes against peace and humankind: or
c) Threatens a person or group of people because of their race, color, ethnic or national origin, religion, sex or sexual orientation;
with the intent to incite racial, religious or sexual discrimination, or to encourage it, will be punished with imprisonment for a term from 6 months up to 5 years.
Law number 7/82, published on 29 April (unique article) The International Convention for the Elimination of All Forms of Racial Discrimination adopted by the United Nations on the 21st of December of 1965 witch related texts in Portuguese and in English follow this law, is approved to adhesion.

10 Appendix II - Further Reading

Neither the Mandola project team nor the Mandola partners specifically endorse any of these organisations or these views on the subject of hate speech and includes these suggested documents and web sites in order to stimulate further debate and understanding of the complex area of online hate speech.

10.1 Jurisprudence

Yahoo! Inc. V. La Ligue Contre Le Racisme Et L'antisemitisme and L'union Des Etudiants Juifs De France, No. 01-17424, United States Court of Appeals, Ninth Circuit 2006, http://caselaw.findlaw.com/us-9th-circuit/1144098.html (on 10/25/2016)

10.2 Council of Europe documentation (including court cases of the European Convention of Human Rights)

Relating to the preservation of human rights on the Internet

Council of Europe, Recommendation CM/Rec(2016)5 of the Committee of Ministers to member States on Internet freedom, http://www.coe.int/en/web/freedom-expression/committee-of-ministers-adopted-texts/-/asset_publisher/aDXmrol0vvsU/content/recommendation-cm-rec-2016-5-of-the-committee-of-ministers-to-member-states-on-internet-freedom?inheritRedirect=false&redirect=http%3A%2F%2Fwww.coe.int%2Fen%2Fweb%2Ffreedom-expression%2Fcommittee-of-ministers-adopted-texts%3Fp_p_id%3D101_INSTANCE_aDXmrol0vvsU%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-1%26p_p_col_pos%3D1%26p_p_col_count%3D3

Council of Europe, Recommendation CM/Rec(2016)1 of the Committee of Ministers to member States on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutrality, https://wcd.coe.int/ViewDoc.jsp?p=&Ref=CM/Rec(2016)1&Language=lanEnglish&Ver=original&BackCol=&direct=true.

Council of Europe, Declaration of the Committee of Ministers on the protection of freedom of expression and freedom of assembly and association with regard to privately operated Internet platforms and online service providers, adopted on 7 December 2011, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805cb844.

Recommendation CM/Rec(2011)8 of the Committee of Ministers to member states on the protection and promotion of the universality, integrity and openness of the Internet, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805cc2f8.

Council of Europe, Recommendation Rec(2001)8 of the Committee of Ministers to member states on self-regulation concerning cyber content (self-regulation and user protection against illegal or harmful content on new communications and information services), https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016804d5105

Relating to hate speech

Courts cases and analysis

Council of Europe, Anne Weber, Manual on hate speech, Sept. 2009, https://www.coe.int/t/dghl/standardsetting/hrpolicy/Publications/Hate_Speech_EN.pdf

Council of Europe, European Court of Human Rights, Internet: case-law of the European Court of Human Rights, updated June 2015, http://www.echr.coe.int/documents/research_report_internet_eng.pdf

Council of Europe, European Court of Human Rights, Factsheet on Hate speech, June 2016, http://www.echr.coe.int/Documents/FS_Hate_speech_ENG.pdf

Others

Council of Europe, Recommendation of the Committee of Ministers to member states no. R (97) 20 on Hate Speech, http://www.coe.int/t/dghl/standardsetting/hrpolicy/Other_Committees/DH-LGBT_docs/CM_Rec(97)20_en.pdf.

Council of Europe, No hate speech campaign, http://www.coe.int/fr/web/no-hate-campaign

Council of Europe, European Commission against Racism and Intolerance (ECRI), http://www.coe.int/t/dghl/monitoring/ecri/default_en.asp

10.3 Current Practices

10.3.1 Code of Conduct on Countering Illegal Hate Speech Online

http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf

A code of conduct document aimed at guiding Facebook, Microsoft, Twitter and YouTube activities as well as sharing best practices with other internet companies, platforms and social media operators.

Facebook, Microsoft, Twitter and YouTube (hereinafter "the IT Companies") – also involved in the EU Internet Forum – share, together with other platforms and social media companies, a collective responsibility and pride in promoting and facilitating freedom of expression throughout the online world;

The IT Companies also share the European Commission's and EU Member States' commitment to tackle illegal hate speech online. Illegal hate speech, as defined by the Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law and national laws transposing it, means all conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin.

The IT Companies and the European Commission also stress the need to defend the right to freedom of expression, which, as the European Court of Human Rights has stated, “is applicable not only to "information" or "ideas" that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population”.

The following public commitments were set in the document:

- The IT Companies to have in place clear and effective processes to review notifications regarding illegal hate speech on their services so they can remove or disable access to such content.
- Upon receipt of a valid removal notification, the IT Companies to review such requests against their rules and community guidelines with dedicated teams reviewing requests.
- The IT Companies to review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary.
- In addition to the above, the IT Companies to educate and raise awareness with their users about the types of content not permitted under their rules and community guidelines. The notification system could be used as a tool to do this.
- The IT companies to provide information on the procedures for submitting notices, with a view to improving the speed and effectiveness of communication between the Member State authorities and the IT Companies, in particular on notifications and on disabling access to or removal of illegal hate speech online.
- The IT Companies to encourage the provision of notices and flagging of content that promotes incitement to violence and hateful conduct at scale by experts, particularly via partnerships with CSOs, by providing clear information on individual company Rules and Community Guidelines and rules on the reporting and notification processes.
- The IT Companies rely on support from Member States and the European Commission to ensure access to a representative network of CSO partners and "trusted reporters" in all Member States to help provide high quality notices.
- The IT Companies to provide regular training to their staff on current societal developments and to exchange views on the potential for further improvement.
- The IT Companies to intensify cooperation between themselves and other platforms and social media companies to enhance best practice sharing.
- The IT Companies and the European Commission, recognising the value of independent counter speech against hateful rhetoric and prejudice, aim to continue their work in identifying and promoting independent counter-narratives, new ideas and initiatives and supporting educational programs that encourage critical thinking.
- The IT Companies to intensify their work with CSOs to deliver best practice training on countering hateful rhetoric and prejudice and increase the scale of their proactive outreach to CSOs to help them deliver effective counter speech campaigns. The

European Commission, in cooperation with Member States, to contribute to this endeavour by taking steps to map CSOs' specific needs and demands in this respect.

- The European Commission in coordination with Member States to promote the adherence to the commitments set out in this code of conduct also to other relevant platforms and social media companies.

10.3.2 Community Standards of Facebook

https://www.facebook.com/communitystandards

A best practice guide of Facebook describing all content policies on the social media as well as to explain what kinds of things shouldn't be shared on Facebook. The Community Standards aim to find the right balance between giving people a place to express themselves and promoting a welcoming and safe environment for everyone.

Facebook removes hate speech, which includes content that directly attacks people based on their:

- race,
- ethnicity,
- national origin,
- religious affiliation,
- sexual orientation,
- sex, gender or gender identity, or
- serious disabilities or diseases.

Organisations and people dedicated to promoting hatred against these protected groups are not allowed a presence on Facebook. As with all of our standards, relying on the community to report this content to us.

People can use Facebook to challenge ideas, institutions and practices. Such discussion can promote debate and greater understanding. Sometimes people share content containing someone else's hate speech for the purpose of raising awareness or educating others about that hate speech. When this is the case, Facebook expect people to clearly indicate their purpose, which helps them better understand why they shared that content.

By allowing humour, satire or social commentary related to these topics, and Facebook believe that when people use their authentic identity, they are more responsible when they share this kind of commentary. For that reason, website administrators ask that Page owners associate their name and Facebook Profile with any content that is insensitive, even if that content does not violate their policies. As always, they urge people to be conscious of their audience when sharing this type of content.

While they work hard to remove hate speech, they also give tools to avoid distasteful or offensive content.

10.3.3 The Twitter Rules

https://support.twitter.com/articles/18311

A General Policies section of Twitter describing limitations on the type of content and behaviour that is allowed on the social media.

Any accounts and related accounts engaging in the activities specified below may be temporarily locked and/or subject to permanent suspension.

- **Violent threats (direct or indirect):**
  You may not make threats of violence or promote violence, including threatening or promoting terrorism.
- **Harassment:**
  You may not incite or engage in the targeted abuse or harassment of others. Some of the factors that we may consider when evaluating abusive behavior include:
  - if a primary purpose of the reported account is to harass or send abusive messages to others;
  - if the reported behaviour is one-sided or includes threats;
  - if the reported account is inciting others to harass another account; and
  - if the reported account is sending harassing messages to an account from multiple accounts.
- **Hateful conduct:**
  You may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.
- **Multiple account abuse:**
  Creating multiple accounts with overlapping uses or in order to evade the temporary or permanent suspension of a separate account is not allowed.
- **Private information:**
  You may not publish or post other people's private and confidential information, such as credit card numbers, street address, or Social Security/National Identity numbers, without their express authorization and permission. In addition, you may not post intimate photos or videos that were taken or distributed without the subject's consent. Read more about our private information policy here.
- **Impersonation:**
  You may not impersonate others through the Twitter service in a manner that is intended to or does mislead, confuse, or deceive others. Read more about our impersonation policy here.
- **Self-harm:**
  You may encounter someone considering suicide or self harm on Twitter. When we receive reports that a person is threatening suicide or self harm, we may take a number of steps to assist them, such as reaching out to that person expressing our

concern and the concern of other users on Twitter or providing resources such as contact information for our mental health partners.

10.3.4 Community Guidelines on YouTube

https://support.google.com/youtube/answer/2801939?hl=en

In relation to hate speech the section gives information about which videos are classified as promoting violence or hatred against individuals and how to report hateful content.

YouTube encourage free speech and try to defend your right to express unpopular points of view, but we don't permit hate speech.

Hate speech refers to content that promotes violence or hatred against individuals or groups based on certain attributes, such as:

- race or ethnic origin
- religion
- disability
- gender
- age
- veteran status
- sexual orientation/gender identity

There is a fine line between what is and what is not considered to be hate speech. For instance, it is generally okay to criticize a nation-state, but not okay to post malicious hateful comments about a group of people solely based on their ethnicity.

Reporting hateful content on YouTube

Keep in mind that not everything that's mean or insulting is hate speech. If you're upset by content that a specific person is posting, you may wish to consider blocking the user.

However, if you feel that content violates our hate speech policy, report it to YouTube for review in one of the following ways:

- Flag the video : You may report hateful content that you think may violate our community guidelines by flagging the video.
- File an abuse report : If you have found multiple videos, comments, or a user's entire account that you wish to report, please visit our reporting tool, where you will be able to submit a more detailed report.

10.3.5 Spain: Action Protocol for the security forces for hate crimes and behaviours breaching legal regulations on discrimination

http://fra.europa.eu/en/promising-practices/action-protocol-security-forces-hate-crimes-and-behaviours-breaching-legal

The protocol provides guidance for police officers on how to identify and treat hate crime incidents.

Spain's Comprehensive Strategy against Racism, Racial Discrimination, Xenophobia and Related Intolerance (2011) sets the objective of improving systems in Spain for collecting statistical information about racist incidents and xenophobia, racial discrimination and other forms of related intolerance.

The protocols contain:

- An index of behaviours on hate crime and breaches of the legal regulations governing discrimination;
- Indicators to identify hate incidents that must be included in police reports. This provides judges and prosecutors with the elements they need to describe the incident, bring charges and prosecute perpetrators. The indicators cover such elements as whether the victim belongs to a minority group, racist or xenophobic expressions or comments uttered by the aggressor, flags and signs exhibited by the aggressor, where and when the incident took place (did it coincide with an historical event), etc.
- Guidelines about police action on hate and discrimination incidents and the sequence of steps that need to be taken;
- Guidance on protecting victims and how to act;
- Managing guidelines on hate crime on the internet and in social networks;
- Guidance on taking action in case of violence in sports;
- Criteria on recording incidents according to the different grounds of discrimination. This is necessary for data collection and allows the Security Forces' Statistical Criminality System (SEC) to provide data classified by these different grounds;
- Guidance on managing relations with NGOs.

The overall objective of the practice is to improve hate crime recording in Spain; to improve police training so officers can identify and record hate crimes, as well as identify all elements that should be considered in police actions regarding these incidents.

10.4 Additional Websites

10.4.1 No Hate Speech Movement

http://www.nohatespeechmovement.org/

A youth campaign of the Council of Europe for human rights online, to reduce the levels of acceptance of hate speech and to develop online youth participation and citizenship, including in Internet governance processes.

The Campaign is part of the project Young People Combating Hate Speech Online running between 2012 and 2014. The project stands for equality, dignity, human rights and diversity. It is a project against hate speech, racism and discrimination in their online expression.

Objectives of the campaign:

- To raise awareness about hate speech online and its risks for democracy and for individual young people, and promoting media and Internet literacy;
- To support young people in standing up for human rights, online and offline;
- To reduce the levels of acceptance of online hate speech;
- To mobilise, train and network online youth activists for human rights;
- To map hate speech online and develop tools for constructive responses;
- To support and show solidarity to people and groups targeted by hate speech online;
- To advocate for the development and consensus on European policy instruments combating hate speech;
- To develop youth participation and citizenship online.

10.4.2 Information about Hate crime in the The European Union Agency for Fundamental Rights (FRA)

http://fra.europa.eu/en/theme/hate-crime

A specially dedicated section about Hate crime in the The European Union Agency for Fundamental Rights (FRA) website consisting:

- Giving victims a voice
- Publications
- Opinions
- Projects
- Surveys
- News
- Events
- Press Releases
- Photo Galleries
- Speeches
- Videos

- Compendium of Practices.

About the organization: The European Union Agency for Fundamental Rights (FRA) is one of the EU's decentralised agencies. These agencies are set up to provide expert advice to the institutions of the EU and the Member States on a range of issues. FRA helps to ensure that the fundamental rights of people living in the EU are protected.

Fundamental rights set out minimum standards to ensure that a person is treated with dignity. Whether this is the right to be free from discrimination on the basis of your age, disability or ethnic background, the right to the protection of your personal data, or the right to get access to justice, these rights should all be promoted and protected.

Through the collection and analysis of data in the EU, the FRA assists EU institutions and EU Member States in understanding and tackling challenges to safeguard the fundamental rights of everyone in the EU. Working in partnership with the EU institutions, its Member States and other organisations at the international, European and national levels, the FRA plays an important role in helping to make fundamental rights a reality for everyone living in the EU.

10.4.3 True Vision - the online reporting site run by the police in England, Wales and Northern Ireland.

http://www.report-it.org.uk/what_is_hate_crime

True Vision is the website which gives information about hate crime or incidents and how they can be reported. In addition True Vision is available as an app – police hate crime app. The app such as the website gives information how to report hate crime incidents. Both the app and the website platform are UK's Police projects.

On this website, you can:

- find out what hate crimes or hate incidents are.
- find out about the ways you can report them.
- report using the online form.
- find information about people that can help and support you if you have been a victim.

10.4.4 Factsheet on Hate speech

http://www.echr.coe.int/Documents/FS_Hate_speech_ENG.pdf

A Factsheet on Hate speech by European Court of Human Rights. In the document there is a list of different examples of hate speech such as:

- Ethnic hate
- Negationism and revisionism
- Racial hate
- Religious hate

- Threat to the democratic order
- Circulating homophobic leaflets and many others.

10.4.5 Council of Europe's website and European Commission against Racism and Intolerance (ECRI)

http://www.coe.int/t/dghl/monitoring/ecri/legal_research/national_legal_measures/

Portal section at Council of Europe's website and European Commission against Racism and Intolerance (ECRI) indicating legal measures to combat racism and intolerance in the member States of the Council of Europe.

10.4.6 American Bar Association

http://www.americanbar.org/groups/public_education/initiatives_awards/students_in_action/debate_hate.html

Section in American Bar Association regarding hate speech and different actions. It is published in the Students in Action and Student Central directories.

About the ABA: The American Bar Association is one of the world's largest voluntary professional organizations, with nearly 400,000 members and more than 3,500 entities. It is committed to doing what only a national association of attorneys can do: serving our members, improving the legal profession, eliminating bias and enhancing diversity, and advancing the rule of law throughout the United States and around the world.

10.4.7 Take Back The Tech!

https://www.takebackthetech.net/about

Take Back The Tech! was initiated in 2006 by the Association for Progressive Communications (APC) Women's Rights Programme and has grown into a diverse movement of individuals, organisations, collectives and communities. It is the result of research papers published in 2005 that looked at the connection between ICT and VAW, an issue that received little attention or discussion at that time. After sharing the findings with women's rights and communication rights advocates in different spaces, APC found this to be a critical issue that compelled further attention and deeper engagement.

Take Back the Tech! sets out to:

- Create safe digital spaces that protect everyone's right to participate freely, without harassment or threat to safety.
- Realise women's rights to shape, define, participate, use and share knowledge, information and ICT.
- Address the intersection between women's human rights and the internet, especially VAW (Violence Against Women)
- Recognise women's historical and critical participation and contribution to the development of ICT.

The campaign has been taken up, adapted and owned by individuals, groups, networks and organisations in places such as Bangladesh, Bosnia and Herzegovina, Brazil, Cambodia, Canada, Democratic Republic of the Congo, Germany, India, Kenya, Macedonia, Mexico, Malaysia, Pakistan, Philippines, Rwanda, South Africa, Uganda, UK, Uruguay, USA.

In the website of the project people could also learn more about:

- Blackmail
- Cyberstalking
- Hate speech
- How to help
- How to be safe online

10.4.8 OSCE Office for Democratic Institutions and Human Rights (ODIHR) Tolerance and Non-Discrimination Department

http://hatecrime.osce.org/italy

OSCE Office for Democratic Institutions and Human Rights (ODIHR)

Tolerance and Non-Discrimination Department website is giving information about hate crimes and recording of hate crimes in many participating countries. In the link above a statistic could be found regarding recorded hate speech crimes in Italy.

Effectively tackling hate crime requires a comprehensive effort targeting various levels. Governments need to integrate strategies to combat hate crime into education, law-enforcement and social policies, as well as to collect and publish relevant data. The gravity of these crimes needs to be reflected in law. Police, judges and prosecutors must be able to identify hate crimes and deal with them accordingly. Civil society play an important role in monitoring and reporting incidents, supporting victims, fostering good inter-community relations and raising awareness in society.Intergovernmental organizations can help set standards, promote best practices internationally and assist states in meeting these goals.

ODIHR has been tasked to assist OSCE participating States in their efforts to counter hate crime. In line with this mandate, ODIHR has partnered with participating States, civil society and international organizations to produce the following programmes:

- **Hate crime recording**
  ODIHR supports government officials in designing and developing monitoring mechanisms and data collection on hate crime.
- **Police training**
  Training against Hate Crimes for Law Enforcement (TAHCLE) is a programme designed to improve police skills in recognizing, understanding and investigating hate crimes,

interacting effectively with victim communities, and building public confidence and co-operation with other law-enforcement agencies.

- **Supporting Law Makers**
  ODIHR helps participating States design and draft legislation that effectively addresses hate crimes. To that end, ODIHR has developed a practical guide assisting law makers in fulfilment of this role. On the request of the participating States, ODIHR also reviews and comments on draft versions of hate crime legislation.
- **Training prosecutors**
  ODIHR provides training that builds the capacity of participating States' criminal justice systems. PAHCT is designed to improve the skills of prosecutors in understanding, investigating and prosecuting hate crimes. In doing so, it helps prevent hate crime and build constructive ties with marginalized groups.

  The programmed is tailored to the needs and experiences of each country in which it is used. PAHCT is short, compact and flexible. It is designed to be integrated into existing training efforts.
- **Working with civil society**
  Civil society plays a crucial role in monitoring and reporting hate crimes. Data provided by NGOs form an important part of ODIHR's hate crime data collection and offer indispensable context to participating States' reporting on hate crimes.

  ODIHR helps raise awareness of hate crimes among civil society and international organizations. It provides information about the characteristics of hate crimes and their impact on the stability and security of the community. ODIHR also supports civil society efforts to monitor and report hate crimes, NGOs outreach efforts in their communities and foster relationships between community groups and law enforcement so that victims feel confident to report crimes. ODIHR also encourages civil society advocacy for better hate crime laws.
- **Working with educators**
  Educators play a fundamental role in countering intolerance and discrimination. ODIHR works to support participating States that have committed themselves to promoting educational programmes that counter intolerance and promote mutual respect and understanding.

  ODIHR, together with the Council of Europe and UNESCO, has developed guidelines for educators to counter intolerance and discrimination against Muslims. As well, in co-operation with national experts, ODIHR and the Anne Frank House in Amsterdam have developed teaching materials to combat anti-Semitism. ODIHR continues to develop educational tools and strategies to counter the biases that can lead to hate crime.

10.4.9 The Legal Project

http://www.legal-project.org/issues/european-hate-speech-laws

Section in the Legal Project's website concerning European hate speech laws, reports, further reading and useful links.

About the Legal Project: protects researchers and analysts who work on the topics of terrorism, terrorist funding, and radical Islam from lawsuits designed to silence their exercise of free speech.

10.5 Research

This section contains additional research reviewed during the preparation of this report.

- "Hate-Speech Protocol to Cybercrime Convention", in The American Journal of International Law, Vol. 96, No. 4 (Oct., 2002), pp. 973-975
- Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, Council of Europe, Strasbourg, 28 November 2003;
- Akdeniz, Y. (1998). Who watches the watchmen? Internet content rating systems and privatised censorship. The Australian Library Journal, 47(1), 28-42.
- Akdeniz, Y. (2001). Controlling Illegal and Harmful Content on the Internet. In D. (. Wall, Crime and the Internet (pp. 113-140). London: Routledge.
- Anne Weber, Manual on hate speech, council of Europe Publishing, September 2009, https://www.coe.int/t/dghl/standardsetting/hrpolicy/Publications/Hate_Speech_EN.pdf (on 09/26/2016);
- Anti-Defamation League, "Cyberhate response, Best practices for responding to cyberhate", available at http://www.adl.org/combating-hate/cyber-safety/best-practices/#.V-UvzDWEq5k
- Awan, I., & Blakemore, B. (2012). Policing cyber hate, cyber threats and cyber terrorism.
- Bailey, J. (2006). Strategic Alliances: The inter-related roles of citizens, industry and government in combating Internet hate. Canadian Issues, 56-59.
- Banks, J. (2010). Regulating hate speech online. International Review of Law, Computers and Technology, 24(3), 233-239.
- Banks, J. (2011). European Regulation of Cross-Border Hate Speech in Cyberspace: The Limits of Legislation. European Journal of Crime, Criminal Law and Criminal Justice, 1-13.
- Bartlett, J., & Krasodomski-Jones, A. (2016). Counter-speech on Facebook UK and France. DEMOS.
- Bartlett, J., Reffin, J., Rumball, N., & Williamson, S. (2014). Anti-social Media. DEMOS.
- Berger, J. (2016). Nazis Vs. ISIS on Twitter: A Comparative Study of White Nationalists and ISIS Online Social Media Networks. George Washington University, Program on Extremism, Washington D.C.
- Brennan, F. (2009). www.hatecrime.co Laws Uncritical Engagement.
- CEJI under the Facing Facts project, November 2012, "Guidelines for Monitoring of Hate Crime and Hate Motivated Incidents".
- Christopher D. Van Blarcum, Internet Hate Speech: The European Framework and the Emerging American Haven, 2005.

- Citron, D., & Norton, H. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. Boston University Law Review, 91, 1435.
- Citron, D., & Norton, H. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. Boston University Law Review, 91, 1435.
- Cohen-Almagor, R. (2011). Fighting Hate and Bigotry on the Internet. Policy & Internet, 3(3), 1-26.
- Cormarc Callanan, "Best Practices for Internet Hotlines" in Hate Speech on the Internet, http://www.osce.org/fom/13846?download=true (on 09/26/2016);
- Council Framework decision 2008/913/JHA of 28 November 2008, on combating certain forms and expressions of racism and xenophobia by means of criminal law.
- European Commission, June 2016, "Code of Conduct on countering illegal hate speech online.
- Guichard, A. (2009). Hate crime in cyberspace: the challenges of substantive criminal law. Information & Communicatinos Technology Law, 201-234.
- Harris, C., Rowbotham, J., & Stevenson, K. (2009). Truth, law and hate in the virtual marketplace of ideas: perspectives on the regulation of Internet content. Information & Communication Technology Law, 18, 155-184.
- Iginio Gagliardone et al., Countering Online Hate Speech, UNESCO Series on Internet Freedom, UNESCO Publishing, 2015;
- James B. Jacobs and Kimberly Potter, Hate Crimes: Criminal Law and Identity Politics (New York, Oxford University Press, 1998);
- Josang, A., Ismail, R., & Boyd, C. A. (2007). A survey of trust and reputation systems for online service provision. Decision Support Systems, 43(2), pp. 618-644.
- McGonagle, T. (2013). The Council of Europe against online hate speech: Conundrums and challenges.
- No hate speech movement, 2014, "Starting points for Combating hate speech online"
- No hate speech movement, 2016 " A manual for combating hate speech online through human rights education"
- Perry, B., & Olsson, P. (2009). Cyberhate: The globalization of hate. Information & Communications Technology Law, 18(2), 185-199.
- Recommendation No. R (97) 20 of the Committee of Ministers to Member States on "Hate Speech", 30 October 1997;
- Richards, R. D., & Calvert, C. (2000). Counterspeech 2000: A New Look at the Old Remedy for "Bad" Speech. Brigham Young University Law Review, 553.
- Sandy Starr, "Understanding Hate speech" in Hate Speech on the Internet, http://www.osce.org/fom/13846?download=true (on 09/26/2016);
- UNESCO (United Nations, Educational, Scientific and Cultural Organization), 2015 "Countering Online Hate Speech"