

#Silence Hate

*Study on hate Speech Online in Belgium,
Czech Republic, Germany and Italy*

BRICKS *Building Respect on the Internet by Combating hate Speech*



Edited by Marcello Maneri

University of Milano Bicocca



www.brickspjroject.eu



Co-financed by the Fundamental
Rights and Citizenship Program
of the European Union.



This report is based on the four national reports carried out in the framework of the project “BRICKS – Building Respect on the Internet by Combating hate Speech” (JUST/2013/FRAC/AG/6090):

Andrea Svobodová, Hate speech online. Internet discussion in the Czech Republic, 2016.

Cécile Goffard, Anne-Claire Orban de Xivry (with contributions from Madeline Corazza and Romuald La Morté), Report on hate speech and participation in Belgium’s online French-language press, 2016.

COSPE, Hate is not an opinion. Study on hate speech, journalism and migration, 2016.

Grimme Institut Germany, Hate only? Reflections on German online media, 2015.

This publication has been produced with the financial support of the Fundamental Rights and Citizenship Program of the European Union. The contents of this publication are the sole responsibility of COSPE and its partners and cannot be considered to reflect an official opinion of the European Commission.



Grimme
Institut





Table of contents

Introduction	4
I. A definition of hate speech	5
II. The national contexts	6
1. Legal frameworks	7
2. Judicial practices	11
3. Prevention and monitoring	14
III. The case studies	17
1. The media selected for analysis	17
2. The topics and reporting that trigger hate speech	18
3. The rules of discussion. The amount and mode of moderation	19
4. Hate speech and moderation in different platforms	20
5. Discussion patterns	24
6. Rhetoric and functions of hate speech	25
7. The profile of hate speakers	27
IV. Editors, journalists and external players' opinions and strategies	29
1. Journalists' attitudes about online participation	29
2. The issues around freedom of speech	31
3. The different strategies of discussions' regulation	35
4. Innovative approaches to the regulation of internet discussions	38
V. Recommendations	41





Introduction

This report is the synthesis of four studies conducted in Belgium (the French-speaking part of the country), Czech Republic, Germany and Italy, and aims at comparing the various realities pinpointing differences and commonalities, with the aim of learning something from each national experience. In order to do this, the study will consider the national contexts and the case studies analysed by the national research teams. A special attention is given to the actions undertaken to encourage a positive and effective contrast to racist and xenophobic discourse on the web and to the opinions and suggestions given by social media managers, journalists, media editors, media activists and scholars.

Although the national research teams have used the same research design trying to answer to the same research questions, they did not deal at the same length with all the features under examination. Therefore, what will be done here is less a strict comparison and more the composition of a jigsaw puzzle, to which every country has given a more relevant contribution here and a less important one there.





I. A definition of hate speech

The definition of the concept of “hate speech” most often refers to the recommendations from the Committee of Ministers on hate speech included in the 1997 Council of Europe document. There, it is understood as: *“all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.”*¹ This definition does not cover all forms of hate speech: victims of hate speech may also be LGBT people, women, people with mental or physical disabilities, or any other stigmatized group. This definition also does not include hate speech directed at individuals, which is very common (such as hate speech directed at public figures).

In the four studies from which this report draws, hate speech has been defined according to the Council of Europe document, so hatred directed at LGBT people, women, people with mental or physical disabilities was not considered. Some cases of hate speech against individuals have been included.

1- Recommendation no.R (97) 20 of the Committee of Ministers to member states on “hate speech” [online]. Available at: http://www.coe.int/t/dghl/standardsetting/hrpolicy/other_committees/dh-lgbt_docs/CM_Rec%2897%2920_en.pdf





II. The national contexts

Hate speech has been an issue both at the political level and in public discussion for some time now, but it has taken on particular importance in 2015. In Italy in 2014, UNAR (*Ufficio Nazionale Antidiscriminazioni Razziali - National Office Against Racial Discrimination*) recorded 347 cases of racist expressions on social networks, 185 of which on Facebook and the remaining on Twitter and YouTube². According to that organization, the trend was increasing in 2015, in online media and social networks. Politicians and public figures pronounced discriminatory and xenophobic slogans and statements as well.

Before the summer of 2015, in Europe typical hate speech comments were predominantly directed at Muslims, for instance in the context of the Charlie Hebdo attacks, at LGBT people, at specific or unspecific (ethnic or religious) minorities. By early summer 2015, hate speech comments were increasingly directed at refugees, especially Muslim refugees, and at people acting in favour of them, like civic supporters, volunteers, politicians, activists. This development was clearly related to the European refugee crisis, which not only led to a surprising amount of civic support but regrettably also to an increase of hate speech in social media. Online media forums, comments on articles, and the Facebook pages of national and local newspapers are the virtual place in which hate speech that targets refugees and citizens of foreign origin has spread. Not only the quantity of comments that undeniably violated the boundaries of freedom of speech, but also the intensity of hate visible in these comments got to a point where politicians of almost all political factions demanded solutions and instruments that would help to minimise or control this form of destructive public discussion.

In Germany, Heiko Maas, the Minister of Justice, appealed to representatives of Facebook Europe (Dublin) and Germany to revise possible means to simplify procedures of deleting hate comments published on Facebook, since this platform had proven to be a hot spot for right-wing extremist and racist posts. The public prosecution department investigated several Facebook managers in Germany in order to find out whether the company could be held responsible for not dealing with hate speech in an appropriate manner. Facebook representatives announced that a revision of the rules and procedures would be agreed upon shortly (as of Dec. 2015). Additionally, a rising number of civic activists or groups as well as well-known public figures (actors, comedians, writers) raised their voices against hate speech. For the first time, individuals were investigated and / or prosecuted for publication of hate speech or lost their job.

2- P. Andrisani, Dal contagio 'virale' al web marketing dell'odio razzista, in: IDOS Dossier statistico immigrazione 2015.





1. Legal frameworks

The legal framework is very diverse in the four countries under study.

Germany has only a very general differentiation between acceptable and unacceptable expression of opinion, which makes it difficult to adequately define and come to terms with hate speech in a judicial sense. The constitution provides everyone with the right of “free expression of opinion via speech, scripture and picture”. The restrictions admitted are there to protect the youth as well as the personal honour of the individual in general. Therefore, hate speech can harm either someone’s honour or the public interest, but is not defined in any specific way.

An exception in German legislation is sedition [*Volksverhetzung*], which is an element of crime that is not “general” when punished because of a necessary national-socialistic motivation that leads to prosecution. Sedition however only covers a small part of the restriction of free expression of opinion.

In practical application, the important restrictions on free expression of speech are found in the right of personal honour / personal rights because opinions often violate those rights. Prosecutable violations include untruthful statements, libel and statements, which have lost all reference to the issue at hand and solely aim at attacking another person, but in the interest of freedom of speech prosecution always depends on the context of the matter, where each incident has to be examined carefully.

The **Czech** legal system, on the contrary, directly prosecutes hate speech, including the online form. The criminal offences share the common characteristic of defining hate speech as being aimed at a group of people who share an unchangeable characteristic; it must also be voiced publicly³. These offences include:

- defamation of nation, race, ethnic or other group of persons (§ 355 of the Criminal Code),
- incitement to hatred against a group of persons or restriction of their rights and freedoms (§ 356),
- establishing, supporting and promoting a movement aimed at suppressing human rights and freedoms (§ 403),
- expression of sympathy with these movements (§ 404),
- denial, questioning, approving and justifying of genocide (§ 405).

In the case of the offences (§ 355, § 356 and § 403 of the Criminal Code), starting from 2010 the new Criminal Code stipulates that instances of hate speech, committed by means of a publicly accessible computer network, fall under the category of strict liability in criminal law, thus allowing higher penalties⁴. Punishment ranges from 6 months to 3 years.

Outside the criminal code the Act no. 480/2004 Coll., “on certain Information Society Services”⁵ is particularly important. It governs the responsibilities, rights and obligations of persons who provide information society services and disseminate commercial messages.⁶

In Italy and in Belgium there is no specific legislation regarding hate speech, but there are a number of provisions related to incitement to racial hatred, to which other forms of discrimination have been added later.

3- Výborný, Š.(2011). Svoboda slova versus nenávistné projevy na internetu. Rigoróznípráce. Masaryk University, Brno.

4- Ibid.

5- Act no. 480/2004 Coll., on certain Information Society Services and on Amendments to some Acts.

6- § 1 Subject of the regulation, Act no. 480/2004 Coll., on certain Information Society Services and on Amendments to some Acts.



In **Italy**, at first the law⁷ ratified the International Convention on the Elimination of All Forms of Racial Discrimination, adopted by the General Assembly of the United Nations in 1969. In its original formulation, Article 3 of this law punished with imprisonment from one to four years “whoever spreads any form of ideas based on racial superiority or racial hatred”, and “who incites in any way to discrimination, or incites to commit or commits acts of violence or provocation to violence against some specific people only because they belong to a national, ethnic or racial group”.

Later, in 1993, the so-called *Mancino Law*⁸ extended the criminal repression to discrimination based on religion and distinguished the acts of “dissemination of ideas” and “incitement to discrimination”, for which a lighter penalty is provided, from those inciting to violence, or provocation of violence, that are more severely punished.

However, in January 2006 the Parliament passed a new law that significantly weakened the punishments provided by the *Mancino law*, which are now alternative: imprisonment up to one year and six months, or a fine of up to 6.000 euro. Law 85/2006 further amended the previous law modifying the terms defining the crime: the punished crime is not “to spread in any form”, but “the propaganda of ideas based on superiority or racial or ethnic hatred”; no longer who “incites”, but who “instigates to commit or commits acts of discrimination based on racial, ethnic, national or religious grounds”.

Currently the Senate is discussing a bill (already approved by the Chamber of Deputies) that should add homophobia and transphobia among the already existing aggravating circumstances.

Since anything is specified in the law regarding the means used for the dissemination and promotion of racist ideas, this also applies to the Internet, but the problem of identifying the “ideas based on racial or ethnic grounds” spread on the web remains complex. Other instruments of the Italian legislation are the “insult” and “defamation” offences in the Criminal Code. Article 594 contains protection against the crime of “insult” and it states that who offends the honour or dignity of another person in his/her presence may be punished with imprisonment up to six months or condemned to a fine of up to 516 euro. The same punishment is applied to those who commit this kind of crime by telegraph or telephone, writings or drawings addressed directly to the victim. The punishment is increased if the offence is committed in the presence of many people. Article 595 of the Criminal Code focuses on “defamation” and provides for imprisonment up to one year or a fine of up to 1.032 euro for those who, talking to many people, offend the reputation of one or more persons. If the offence is made using the press or any other means of advertising, the penalty consists of imprisonment from six months to three years or a fine of at least 516 euro.

In the last years, there have been several efforts to regulate better hate speech on the web. The Senate approved in May 2015 the bill “Measures for protecting children for the prevention and fight against cyberbullying”⁹, which provides for the removal of offensive contents by the Internet Service Providers or the report to the Privacy Authority. Another bill (the so-called DDL Campana) provides for the imprisonment from six months to four years for anyone who “causes a persistent and serious state of anxiety or fear or generates founded fear for its own safety” through a series of acts known as “bullying or cyberbullying”¹⁰.

7- Law 13 October 1975, n. 654, <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:1975-10-13;654>.

8- Law 25 June 1993, n. 205, <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:1993;205>.

9- <https://www.senato.it/service/PDF/PDFServer/DF/302161.pdf>.

10- http://www.camera.it/_dati/leg17/lavori/stampati/pdf/17PDL0018680.pdf.





In July 2015 the Italian Charter of Rights on the Web (Carta Italiana dei Diritti di Internet) was born, with the goal to offer a high standard set of principles and values in the access to and the use of the Internet. In its Article 13, the Charter addresses the safety on the web that is properly seen both as protection of infrastructures and as defence of people. The article states that “no limitation of freedom of expression is allowed”, but “it must be guaranteed the protection of the people’s dignity from abuses related to behaviours such as incitement to hatred, discrimination and violence”.

In **Belgium**, four laws govern hate speech. Firstly, the Belgian Anti-racism Law passed in 1981 (also called the Moureaux Law), again based on the International Convention on the Elimination of All Forms of Racial Discrimination. This law defines “racial discrimination” as “any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life”. This law is chiefly aimed at exclusions based on race and therefore does not cover all the exclusions linked to religion, like Islamophobia. An EU Council Directive that includes discrimination based on religion or belief, disability, age or sexual orientation therefore supplemented it in November 2000. That Directive was transposed into Belgian national legislation by the Antidiscrimination Law of 10 May 2007, which extends those criteria to civil status, birth, wealth, religious or philosophical belief, political convictions, union convictions, language, current or future state of health, a physical or genetic characteristic or social origin¹¹.

Hate speech is also governed by the Men-women Anti-discrimination Law of 10 May 2007 and the Law against Revisionism of 23 March 1995. In the context of combating cyber hate, ‘incitement’ is a key concept when enforcing the above-mentioned four Belgian laws. “To be precise, the first two laws prohibit people from ‘inciting to’ discrimination, segregation, hatred or violence against individuals or groups on the basis of certain characteristics. ‘Inciting to’ is to be understood as any verbal or non-verbal communication that incites, stimulates, stirs up, encourages, accentuates, causes, impels or calls for others to have certain hate reactions. This is therefore more than a matter of mere ideas, information or criticism. However, it is not necessary for that incitement automatically to lead to a reaction¹².

11- Pax Christi (2015). Cadre de reference contre le racisme et les discriminations. Brussels.

12- Centre for Equal Opportunities (2009) October. Delete Cyberhate - Racisme et discrimination sur internet. <http://www.diversite.be/brochure-cyberhate>.





2. Judicial practices

One of the many practical problems in pursuing hate speech incidents is that the actual **authors can rarely be held accountable** since the identity of users on web pages, blogs and forums is hard to investigate and determine. In Italy, prosecutors receive every day complaints for defamation on the web, but in most cases they ask for the dismissal of the case on the basis of the so-called “objective desensitization” on the Internet: according to this reasoning, the choice to remain anonymous reduces the credibility of the message and severely limits its ability to damage reputation. Similarly, the context of the web is supposed to reduce the credibility of the defamatory statement as well as its damaging potential for the reputation of people, in this case sizing down the offensiveness.¹³ However, a recent verdict by the Supreme Court (Corte di Cassazione) about a post published on a Facebook wall considered “the potential and ability of the media used for the offence, involving and reaching a number of people (...) thereby causing a greater and more widespread damage to the victim”¹⁴ as an aggravating factor.

In general, a crucial issue is whether **who operates the web page can be held responsible to remove law-infringing content**. In Belgium, as François Deleu, a lawyer with the Cyberhate Unit of the Belgian Centre for Equal Opportunities explains, the rules of legal liability for content are different depending on whether the medium is regarded as host or publisher. The criterion is that if the moderator has a high-profile presence on forums, gets involved in conversations or pre-moderates, that is to say, checks the comments before they are put up, he will be considered to be the publisher of the posts and will therefore be held liable for any hate speech appearing on his site. If, on the other hand, the newspaper practises post-moderation it will be considered merely a host and will therefore be liable only for hate speech it is effectively aware of, that is to say, comments that have been reported to it. This rule applies throughout Europe, as it derives from e-commerce Directives establishing a specific system of liability for hosts. This partly explains why the majority of the media in Belgium prefer to post-moderate.

However, Germany seems to follow a different path. The Federal Constitution Court has decided that the forum operator is accountable only when they have neglected duties of necessary monitoring of content. If the operator is notified of law-infringing content, she/he is responsible for its removal. In this case the operator has to contact the author of the statement and ask for an adequate response. If the statement in question is not compliant with the German law, then the operator is required to delete it¹⁵.

13- ‘Simone Alliva, ‘Omofobia e razzismo: ecco perché minacce e insulti sul web restano impuniti’, <http://espresso.repubblica.it/attualita/2015/12/09/news/omofobia-e-razzismo-ecco-perche-minacce-e-insulti-sul-web-restano-impuniti-1.242753>

14- Supreme Court, Sentence no.24431, 12/02/2015,<http://www.penale.it/page.asp?IDPag=1198>.

15- <https://www.amadeu-antonio-stiftung.de/hatespeech/hate-speech-im-internet-eine-rechtliche-einordnung/>





It must be mentioned that a judgment delivered in 2015 by the European Court of Human Rights, known as the Delfi judgment, goes against this rule, which minimizes the host's liability. The Court upheld the liability of the operator of the large Estonian Internet portal Delfi for the offensive comments of its users. The European Court indeed appears to take the view that media hosting content must make every effort to avoid and prevent hate speech on their platforms. This position has the media assuming more responsibility than they do at present as hosts. The court decision might have an impact on the functioning of web portals in all member states of the Council of Europe.

This opinion goes against the rule that applies to the special case of hosts as regards liability (e-commerce Directives). There is therefore a difference between the Court of Justice of the European Union and the European Court of Human Rights. These two rules being contradictory, the law remains unclear in this respect. Judges can therefore choose the path they wish to take, namely that of the media's civil liability, recommended by the European Court of Human Rights, or that of legal liability as host, recommended by the Court of Justice of the European Union.¹⁶

If we look at the **actual cases brought to court**, it is clear that the legal system works better when (prominent) persons pronounce in public or post on their web page their explicit incitement to hatred or racist ideology. A groundbreaking court ruling regarding the promotion of extremism in the Czech Republic was issued in 2012, when five young men received a suspended sentence of three years for promoting Nazism on their Facebook profiles¹⁷. In September 2014, a former MP Otto Chaloupka was conditionally sentenced for insulting comments about Roma people posted on his Facebook profile¹⁸. Similarly, in Italy in 2009 the Supreme Court confirmed the conviction of the current mayor of Verona to two months of imprisonment for "racist propaganda"¹⁹. He had organized a petition to evict a Roma camp in Verona with a campaign called "Sign to send gypsies away from our city". For the first time an Italian Court considered racism not only as a biological-genetic theory but as one that also includes a cultural element. Among judgements specifically concerning hate speech cases on the web, it must be highlighted the sentence of the Court of Padua that in 2011 condemned a municipality councillor to the payment of a 4.000 euro fine and 6.000 euro for damages to the civil parties for carrying "propaganda of ideas based on racial hatred and incitement to commit racist acts". The councillor had written racist comments in his Facebook profile against Roma people. In 2013, the same Court sentenced to one year and half of imprisonment a Northern League councillor of a district of Padua: she committed the crime of incitement to racial violence against Cécile Kyenge, the former Minister for Integration, through a comment posted in her Facebook profile. Because of this episode, she was then expelled from the party²⁰.

16- Interview with as François Deleu, Centre for Equal Opportunities.

17- <http://tn.nova.cz/clanek/pozor-co-pisete-za-komentare-na-facebooku-muzete-jit-do-vezeni.html>

18- <http://zpravy.aktualne.cz/domaci/soud-potrestal-chaloupku-podminkou-za-vyroky-o-romech/r-43a23506336611e49b2b002590604f2e/>

19- Supreme Court of Cassation, fourth criminal section, Sentence of July 10, 2009, no. 41819.

20- 'Tribunale di Padova: l'invito a commettere violenza sessuale nei confronti della Ministra Kyenge espresso attraverso i social network costituisce un'istigazione pubblica alla violenza razziale', in: <http://www.cirdi.org/?p=4788>.





An important decision of the Supreme Court²¹ in Italy extended the application of Article 416 of the Criminal Code on “criminal conspiracy” to hate speech perpetrated within virtual communities, blogs, chats and social networks. The Supreme Court stated that social networks and the internet are apt to spread messages aimed at influencing ideas and behaviours of the public opinion and, therefore, this kind of crime (which provides imprisonment from 1 to 5 years) applies to propaganda of ideas based on racial hatred and racial discrimination through these means. The first sentence that recognized the existence and the prosecution of a criminal organization constituted through the web was pronounced by the Court of Rome against the managers of Stormfront, a neo-Nazi website. This sentence, which established an important precedent in the struggle to cyber-crime, convicted to three years in prison the managers of the forum who incited to commit violence based on racial, ethnic and religious prejudice, and exalted the superiority of the white race in the Italian section of the website²².

Hate speech still represents a **controversial issue in the field of freedom of expression**. The definition of hate speech and its legal limits is intrinsically linked to freedom of speech, in that it restricts it. As emphasized by Bart Cammaerts, a Doctor of Social Sciences at VUB (Brussels), “Inevitably the hate-speech debate also brings into question ideological differences in relation to how absolute the freedom of speech is, the nature of democracy and how or to what extent a balance needs to be struck between different rights, including respect for and recognition of difference and the right not to be discriminated against.”²³

While the concept of ‘incitement’ can be relatively well explained, for example in Belgian laws, it appears that, in fact, the distinction between incitement to hatred and simply expressing an opinion is not always so clear, as several social media managers told researchers in Belgium in their interviews. It thus regularly happens that if there is any doubt about this distinction, the moderators err on the side of caution, preferring to remove or hide posts.

As François Deleu, of the Belgian Centre for Equal Opportunities, explained, there are several levels of hate speech – the ultimate one being that prohibited by law – including stigmatizing remarks that accentuate the tendency to lump things together and are hurtful to a section of the population. In the lawyer’s view, the way to combat such stigmatizing remarks is not through legal channels but rather by education. Also in Italy, most lawyers are contrary to face the problem of hate speech by using the criminal justice system, although there is a favourable trend of the Courts to incriminate hate speech cases.

In Germany, the legal framework implies that - apart from sedition - important restrictions on free expression of speech are found only in the right of personal honour and personal rights. Prosecutable violations include untruthful statements, libel and statements that have lost all reference to the issue at hand and solely aim at attacking another person, but in the interest of freedom of speech prosecution always depends on the context of the matter, where each incident has to be examined carefully.

21- Supreme Court, Sentence no. 33179/2013, available at: http://www.asgi.it/wp-content/uploads/public/cassaz_33179_2013.pdf.

22- Court of Justice of Rome, Sentence no. 884 del 8/04/13, available at: http://osservatorioantisemi-c02.kxcdn.com/wp-content/uploads/2013/11/Sentenza-884-13-dell8.04.2013-1_stormfront.pdf.

23- Cammaerts, B. (2009). “Radical pluralism and free speech in online public spaces: the cases of North Belgian extreme right discourses”. *International Journal of Cultural Studies* (12), pp. 555–575.





3. Prevention and monitoring

In the **Czech Republic**, the Ministry of the Interior and the Police are engaged with the issue of hate speech spread over the Internet, especially in the context of extremism and the manifestation of extremist violence. The Interior Ministry issues an annual *Report on extremism in the territory of the Czech Republic* and a policy paper called *the Conception in combating extremism* and its evaluation²⁴. Both documents also deal with hate speech on the Internet, although only marginally. In the policy paper, *Conception in combating extremism*, point 1.2 (Internet without hate propaganda) hate speech is mentioned in connection with one of the first five pillars of combating extremism (Using communication against demagogu). A concrete step in achieving this goal is the operation of an Internet Hotline²⁵. It receives reports from users regarding illegal and inappropriate content on the Internet. The Hotline evaluates the content of the reports and passes information to such institutions as the police, mobile operators, Internet service providers or the international organization INHOPE.

The Information Crime Department, of the Police Headquarters, and the Department for Combating Organised Crime are responsible for the prosecution and monitoring of hate speech on the Internet. However, there is no special team: the activity falls within the job description of all investigators in the department. In addition to the investigation of reports submitted through a special "Hotline", detectives also have to search actively the problematic posts. Their activities do not confine themselves to print and Internet resources, but are dedicated to all open sources dealing with extremism, including television broadcasts. The police is also trying to establish direct cooperation with the operators of social networking websites to come to agreements determining who would be responsible for the removal of hateful messages, etc. According to one of the police officers, hate speech is especially used as a resource for monitoring extremist groups. Collected hate speech manifestations then serve as evidence for later prosecution of those groups or individuals.

A good number of non-governmental organizations is concerned with hate speech²⁶. They publish annual reports, provide free counselling and legal assistance to victims of hate crimes, promote educational projects and campaigns against hate speech, provide online media and social network monitoring.

In Italy there are two main public bodies dealing with discrimination complaints: OSCAD (*Osservatorio per la Sicurezza Contro gli Atti Discriminatori* - Observatory for Security Against Discriminating Acts) within the framework of the Department of Public Security and UNAR (*Ufficio Nazionale Antidiscriminazioni Razziali*- National Office Against Racial Discrimination), created in 2004. UNAR is part of the Department for Equal Opportunities of the Presidency of the Council of Ministers.

24- <http://www.mvcr.cz/clanek/extremismus-vyrocn-zpravy-o-extremismu-a-strategie-boje-proti-extremismu.aspx>

25- <http://www.horkalinkaczi.cz/>.

26- An overview of organizations dealing with the issue of hate speech in society, including relevant projects, is included in Appendix No. 1 of the Czech Study.





OSCAD receives reports and complaints by telephone or by e-mail from institutions, associations and private citizens; it also monitors the discrimination phenomena on grounds of "race" or ethnicity, nationality, religious belief, gender, age, language, physical or mental disability, sexual orientation, gender identity. Among its tasks, OSCAD also activates, according to reports received, targeted interventions on the ground through the Police Forces and follows the evolution of discriminatory acts reported, besides developing measures of prevention and contrast of hate crimes²⁷.

UNAR makes available other reporting tools, collecting reports, complaints and witnesses through a dedicated Contact Centre, accessible through a toll-free number and a website²⁸. It also provides assistance to victims of discrimination in proceedings conducted both at administrative and legal level. A Memorandum of Understanding between the Ministry of the Interior and the Department of Equal Opportunities establishes that the Police must report to the UNAR Contact Centre discrimination cases that fall under the civil protection and do not constitute crimes, while UNAR is obliged to report to the Police all hate crimes and racist acts of criminal relevance²⁹.

The Postal Police is the section of the State Police that specifically investigates on crimes committed on the web and consequently on hate crimes, even if there is not a specific section dedicated to hate speech. Reports to the Postal Police can be made in person, sent by e-mail or through a special online form. UNAR and OSCAD refer to the Postal Police all reports concerning discrimination and racist acts on the web³⁰.

It is worth noticing that there are also several NGOs and associations, which play an important role in the prevention and fighting against discrimination. They are all enrolled in the Register of associations operating in the field of fight against discrimination and to the list of bodies entitled to legally act for protecting victims of racial discrimination.

27- <http://www.poliziadistato.it/articolo/22017/>.

28- http://www.unar.it/unar/portal/?page_id=201.

29- <http://www.poliziadistato.it/articolo/view/22039/>.

30- <https://www.denunceviaweb.poliziadistato.it/wfintro.aspx>.





III. The case studies

This section draws on the more than twenty case studies analysed in the different countries in order to find some common features and patterns and to try to sketch a few generalisations. Some insights come from interviews conducted with web editors as well. What do the several examples of hate speech teach about its content, function and rhetoric, its dynamics, the forces that heat it or cool it down, the different roles of the actors involved, the relationship between news content and comments? Of course, the actual analyses presented in the national reports are far more informative and detailed than what can be presented here.

1. The media selected for analysis

A three-month monitoring of online news websites preceded the selection of articles and discussions. The websites were selected trying to pick:

- An online version of a national newspaper
- An online news outlet at the national level, with no paper version
- An online newspaper that has experimented with new ways of interacting with its readers
- A popular online newspaper, with lightweight news
- A local online newspaper, with a big readership in a particular area

Five articles were then selected for analysis. The objective of the case studies was to analyse significant cases of racist hate speech or of incitement to hate developed among the community of readers. Articles deemed “significant” were those that well illustrated some forms of “hate speech”, trying to highlight different dimensions of the issue. In addition, a criterion of “popularity” was adopted, measuring the number of people involved in the discussion and number of shares and likes. For the sake of comparison, a discussion about the Charlie Hebdo case has been included in all countries.

In Germany the articles were also chosen in accordance to the impact they had in German media, to the presence of celebrities, or if the topics themselves regularly generate a debate in Germany.

The articles were published on various news sites and on several social media sites such as Facebook, Twitter, and Instagram. This ensured a large amount of comments for analysis.

2. The topics and reporting that trigger hate speech

The topics of articles that cause negative and hateful reactions are **dependent on current political events in society** and are quite variable. In the Czech Republic for example, in 2014 the posters were mostly targeting the *Roma and the Russian-Ukrainian* conflict in their hateful comments. On the contrary, in January 2015, after the attacks on the Parisian magazine Charlie Hebdo, the most discussed topic with the most frequent hate speech was *Muslims and Islamic terrorism*. A large share of the commenting appeared as part of the movement *Islám v České republice nechceme* (We do not want Islam in the Czech Republic). Since the outbreak of the so-called “refugee crisis” (in May 2015) one of the most represented topics, and the one who attracted the most frequent aggressive comments, has been *refugees and migrants*.





As the analysis of one of the Italian case studies has shown, when news deal with a **violent event** the emotional involvement is stronger and therefore even the reaction of the community presents more dramatic and drastic tones.

However, although the news stories can intensify and boost aggressive comments (the migration issue being the “hottest” according to Italian editors), **no specific topic is free from such risk**, so that several cases of hate speech in Italy were found in a wide array of issues: sports, social policies, demographic and cultural aspects. Racist hatred can arise in response to a statement pronounced by a public body, a political or institutional actor and so within a particular framework of the issue given in the article, but also as an instinctive expression of a reader who expresses a feeling of hatred regardless the way in which the topic was expressed.

In Belgium, however, some topics are considered more sensitive than others are: the Israeli-Palestinian conflict, topics relating to migrants and refugees, Islam and the Arab world, Jews and, finally, LGBT issues. Many hate comments can be also found on articles about sports and in particular football, where clashes occurred between clubs.

The case studies analysed in Belgium do not seem to show any direct relationships between the way an article is written and the hate comments it attracts. It may be noted nonetheless that the **words used** in the article’s headline to describe the people involved are often **attention-grabbers** or even sure fire-triggers for comments (Molenbeek, Muslim, Islam, veiled women). These words lead to racist reactions because they have to do with religion (Muslim or Jewish) or they refer to districts with a high population density of foreign / North African origin.

Similarly, in Germany, matters related to keywords such as Holocaust, 3rd Reich, Nazi Germany, connoted with remembrance, guilt, identity, responsibility are extremely controversial, leading to high emotional and often very aggressive comments.

3. The rules of discussion. The amount and mode of moderation

The emergence of the earliest forms of participation in the online press brought with it the question of moderation, a question that was generally answered in one of two ways. Firstly, by **pre-moderation**, which is similar to readers’ letters: each comment is read and approved before being published. This form of moderation curbs communication between participants in that their posts are not published instantly, but it does have the benefit of greatly cutting down abuse. It is evident that pre-moderation is becoming increasingly rare today, as it requires a lot of human resources, is far from the trend to instantaneous communication and also because the use of Facebook as a tool for reader participation is becoming increasingly common, either on the medium’s Facebook page or on its website. The editors who use the Facebook platform cannot pre-moderate.

The second option is **post-moderation**, where moderators can remind users of the netiquette or to comment only on the issue at hand. They can also insert themselves into the debate, cancel individual posts or the whole discussion, or even ban individual posters. This was the most common form of moderation in the Belgian case studies and was applied to various degrees by different news outlets also in the Czech Republic and Germany, but only seldom in Italy. As a German case study shows, new approaches by newspaper’s editorial staff, e.g. the usage of irony and sarcasm against trolls and haters, could yield to a new culture of commenting, although this requires a more in-depth analysis that could confirm it as an effective and sustainable method. A common problem with post-moderation, as will be explained below, is that while it is generally performed on the news website comments, it is far rarer in newspapers Facebook pages or Facebook plugins.





Some online newspapers used **'reactive moderation'**, which consists of intervening only when abuse is reported by a user. This can be done either on the website or via the Facebook page (see below).

Another common option was **not to open discussions** on highly controversial issues, something that has been frequent in the Czech Republic after the refugee crises, with the related considerable increase of hateful comments, and two court decisions that upheld the liability of the editors for the offensive comments of their users.

A common finding of all the studies is that when moderation is **absent** or too loose, either because the newspaper is understaffed or by editorial choice, or also due to the technical limitations of Facebook, hate speech spreads without obstacles. In Italy, moderation was infrequent even when the language was becoming heavily offensive and so it happened in some case studies in the Czech Republic. Other readers' moderation has proven generally ineffective and the simple link to Terms of Use is not sufficient.

Some newspapers have managed to ensure that posters behave in a relatively decent manner building a **community of readers** via a loyalty program, or simply asking commenters to register.

4. Hate speech and moderation in different platforms

For all the editors interviewed in Belgium, not being on Facebook is suicide for the online press today. Readers' practices have altered with time and nowadays many web users get their news via that platform. The growing tendency of the news media to use Facebook raises some questions, notably in the area of moderation, where the moderators are very dependent on the tools Facebook sees fit to put in place to support them in their work.

Some media continue using their own system of moderation on their website while others have opted to install the Facebook plugin and to use only that system for reader interaction from now on. Lastly, a third point of moderation occurs on the media's own Facebook pages. These choices have a big influence on the types and levels of moderation.

Posting articles on its **Facebook page** forces the newspaper to submit to the American social network's technical features, one of these being that pre-moderation is not possible. Admittedly, that promotes direct interactions between users and thus increases the number of reactions. The Facebook pages of the media interviewed in Belgium receive between 1,000 and 4,000 comments per day on average. At most newspaper offices there is only one person in charge of moderation, or else journalists take turns at checking comments. Therefore, it is impossible to read all the comments in terms of human resources. This is the reason why hate speech appears in Belgium more likely on Facebook than on press websites.

Being unable to keep an eye on all the comments, many media therefore rely on the readers to bring abusive posts to their attention. However, if a user decides to report abusive language, she or he can only report an entire article, and not one particular comment below an article. To do so, she/he has to check the option 'I don't like this post' next to the article and state 'what is the problem with this post'. Facebook then explains the available options to the user: she/he can either submit the article to Facebook for examination, block the newspaper, hide all for that newspaper, or send the newspaper a message.

If she/he wants to avoid the 'lengthy' process of reporting to Facebook, the user's other option is to send a message directly to the newspaper responsible for the page. The moderator will then be alerted that problem content has been reported.





These various ways of reporting abusive language are not as simple as the 'Inappropriate' buttons on the media's own websites, or the 'Undesirable reaction' buttons that alert the moderator directly without having to go through a series of questions or send a message.

Adopting Facebook's system of moderation also means accepting that editors cannot take down comments; all they can do is make them invisible to other users. The person whose remarks have been hidden therefore does not realise this; there is no way for them to know they have overstepped the limits. All they will realise, at most, is that their comments are less 'liked' or that they have stopped getting reactions. This feature makes life easier for moderators, in that they no longer need to give reasons for taking down comments, but on the other hand, the people concerned are not given a chance to reassess their views. If someone posts problem comments too often, the media page can also opt to ban them. Anyone who is banned cannot then comment, or like, or do anything else on the page. The person will not be notified, but they will not take long to realise what has happened if they visit the page frequently.

Apart from reporting abuse, the only other moderation tool Facebook offers moderators is the use of filters. The way filters work is that when certain words are used, such as abuse for example, the comments are automatically made invisible.

Facebook has "Community Standards" on its page, where, among other things, it describes the rules regulating hate speech. The Standards establishes that the following will be deleted "verbal expressions, which include content that directly attacks people based on their: race, ethnicity, national origin, religious affiliation, sexual orientation, gender, gender identity or serious physical disability or disease." Further, it declares that "organizations and people dedicated to promoting hatred against these protected groups are not allowed a presence on Facebook".

Many, however, believe that the social networking site Facebook is not fighting hate speech vigorously enough. Recently, as noted in the Czech report, the Federal Republic of Germany, especially the Minister of Justice Heiko Maas, has been involved in this issue the most. In August of this year he sent a letter to Facebook's management in which he requested a reconsideration of the policy on racist and xenophobic posts directed at refugees, and suggested an approach similar to contributions depicting nudity. Repeated arson attacks on refugee hostels in Germany, which had been preceded by a seditious campaign on Facebook, led the minister to take this step³¹. The result of the meeting that took place at the company's European headquarters was the establishment of a "Task Force" based at the German branch of Facebook and financially supported by the Ministry. The group members' job description includes the evaluation of hate speech on the Internet and, mainly, monitoring whether or not these statements conflict with German law. Facebook also took action and organized a meeting with representatives of local communities, with the aim of finding a solution to this "complex" issue. At the same time Facebook introduced a policy that favours the promotion of "counter-statements" (counter speech) rather than the removal of problematic comments.

31- Kabátová, Š.(2015). (online) Nenávist vůči uprchlíkům na síti zaměstnává policii. Denněpřijdeudání. The article was published on 9.9.2015 in Lidovénoviny. Available at: http://www.lidovky.cz/nenavist-vuci-uprchlikum-na-siti-zamestnava-policii-denne-prijde-udani-13h-/zpravy-domov.aspx?c=A150908_115325_In_domov_sk





In essence, as the Czech report highlights, this means that Facebook's intention is to support users in commenting even more. The founders of Facebook also believe that "people can use Facebook to challenge ideas, institutions, and practices. Such discussion can promote debate and greater understanding." The question is whether this proclamation is an attempt to strengthen independent political debates on controversial issues, or a deliberate business strategy.

Nowadays, the influence of Facebook is enormous and it will increase in the future. In May 2015, nine media companies (The New York Times, BBC News, Bild, National Geographic, BuzzFeed, NBC News, The Atlantic, The Guardian and Der Spiegel) concluded an agreement with Facebook, in which they committed that they would publish some reports through Facebook exclusively. The new service will be called Instant Articles. Content publishers will be able to either insert adverts into articles themselves, sell it and keep all the revenue, or allow Facebook to sell advertising and keep 30 percent. Facebook will also allow content publishers to collect data about people who read the articles.

A number of editors have opted to integrate the **Facebook plugin** into their websites, thus choosing not to use their own system and thereby relinquishing control over moderation options. For some of them, going over to the Facebook system meant to eliminate gradually hate commenters from their page and to select carefully the articles where the comments option is permitted. In their opinion, it would be more complicated to create a number of false Facebook accounts than to create false accounts on a press website: so people who have been banned from the page come back less often. Besides, false accounts would be easy to spot on Facebook: those are the ones with no profile picture, few friends, crazy names, etc.

Facebook also offers the possibility of promoting comments on the media website; however, it would appear that this option is used to move less interesting comments lower down the list, rather than to give prominence to the best ones (for a few interesting exceptions see further on).

Those editors who have chosen to use Facebook on their own websites explain that it has its advantages for the user. Indeed, as the majority of readers have a Facebook account, when they read an article they do not need to go through the registration process, and that makes posting comments easier and quicker. It stimulates interaction even more, in that it is an appeal for personal accounts and interaction: the users' profile pictures appearing below the article invite them to comment on the article via their Facebook account.

Other editors, have chosen not to use the Facebook plugin on their site because the social network does not provide the same control over comments as they currently have with their **own Content Manager System**. In some editors' opinion, there are too many uncertainties surrounding the moderation options provided by Facebook, which are deemed far more cumbersome in terms of procedures. Using Facebook for the simplicity of instantaneous communication and the viral effect of the conversation forces the media to lose some control over the management of the moderation system and the content. The primary users' charter on the Facebook pages is that of the social networking corporation, not that of the media. Often comments posted on the news outlet Facebook page have a lower quality than those posted on the website itself.





5. Discussion patterns

As the “SılaŞahin” case study has shown in Germany, both the events that trigger “shitstorms” (German expression for online backlash)³² and “shitstorms” themselves can move from one platform to the other, in particular when traditional and established news media report on an event or a “shitstorm” producing even more attention and comments in the social media channels.

As far as the dynamic of the single discussion is concerned, the majority of users in Belgium post one to two comments, confining themselves to stating their opinions or expressing their emotions without getting into a genuine discussion with the other surfers. Nonetheless, in the Italian case studies this form of “communication to” coexists with another form of “communication with”: some posters address their message to a specific member of the community. It may be noted however that in all the case studies conducted in Belgium, three or four commenters monopolise the conversation, accounting for nearly one-third of the total comments. These lead commenters either hold a discussion among themselves (a sort of debate *in camera* with little interference from the other commenters), or else they make a point of systematically challenging the other users’ comments, either taking a xenophobic stance or, conversely, acting as moderator. The lead commenters generally defend very black-and-white positions and are backed up by other web surfers who rally to one or other of the two dominant opinions. By no means all the exchanges are confined to trading insults, as some of the discussions are conducted in a friendly tone and even result in some people modifying their points of view. However, such instances of friendly exchanges are a minority compared with the comments that are abusive or not open to dialogue. As the Belgian and the Italian studies have shown, if conversations take place, they happen when some users counter discriminating remarks. Otherwise, it’s more a case of a continuous thread (a safety valve) of ‘borderline’ remarks, to borrow the term used by a community moderator, which are difficult to home in on and position within the framework of the laws.

It sometimes happens that a user attempts to introduce some shades of grey or to calm down the debate by calling for less abuse and verbal attacks. This pattern of two opposing groups and one (or more) person(s) trying to calm things down can be found in nearly all the case studies.

According to the Italian study - where positive comments have been very frequent and often written in opposition to aggressive comments from which people want to dissociate themselves - those comments aiming at achieving a position of mediation between two extreme points of view, as well as more objective and aware interventions, can shift the tone of the conversation and weaken the intensity of statements. However, such behaviours, in order to be effective, must be supported by specific “structural” moderation instruments, activated by editorial staffs and not entirely left to the good will of a single reader. It also happens frequently that one reader will appeal to the newsgroup moderator when the remarks go too far.

Many of the insults target the surfers’ assumed nationality or religion (according to their profile name or the content of their posts). In the Czech Republic, also attacks on the editorial team or specific members of the editorial staff were frequent. Many commenters use data provided by the other users’ profiles (photographs, age, previous posts, etc.) to strip what the person says of all credibility or to ridicule them. The commenters do not appear to be loyal to one publication exclusively, as the same people turned up in several articles in different newspapers in Belgium, and users sometimes refer to the quality of rival publications’ newsgroups as a way of criticizing the standard of the posts.

32- “Shitstorm” was first recorded in German usage in 2010 and it refers to widespread and vociferous outrage expressed on the internet, especially on social media platforms. <https://www.theguardian.com/books/booksblog/2013/jul/04/shitstorm-german-dictionary-duden-shitschturm>.



6. Rhetoric and functions of hate speech

In the Italian case studies, researchers found some frequent rhetorical moves, used to legitimize harsh statements:

- **"I am not racist, but ..."** which, as many experts have pointed out, only confirms the label from which the authors of the comments want to escape. "I'm sorry, I do not feel racist but as a mom I don't dare to think what that poor woman is feeling now...let's get them come back home..."

- "I talk like this **because I know them**" is used to validate what follows. Some readers preface it by saying that the point of view expressed is the result of a personal experience and therefore it is more aware. The author of the comment presents himself/herself as "expert on the issue" and on this condition, he/she bases the credibility of his/her point of view and tries to gain trust and confidence within the community of readers. However, this supposed "knowledge" of immigrants is a way to draw distinctions between "Us" and "Them", with a superficial description of the "Otherness" described. In the Belgian case studies, a similar reference to one's own authority was found: "I know because I've experienced it", "I am Sicilian and so...", "I'm speaking from experience...", "Some of my friends are Muslims and so..."

- **I say what other people think.** "If other people do not express racist ideas, it is because they are hypocrites. I am not. So I do". Allegations like this are addressed towards other citizens, who hide behind 'too indulgent' a thinking, or towards the politicians who do not want to see the obvious and who, due to opportunism, do not say what they "really" think. Those who express their contempt and dissent, even if in a "hot-blooded" way, are more sincere and honest and so more legitimated to speak.

- **They force me to be racist.** "I would not be racist, but other people force me to be like that ..." The author of the comment feels the urge to express his/her own hostility, in an offensive way, by forces external to his/her will and thereby legitimizes an aggressive and violent behaviour. The author needs to react to the silence or to the immobility of the institutions, considering migrants as a "punishment" that Italian people deserve for their weak and passive country. In this rhetoric the first target is the migrant who managed to seize a chance within the 'flaws' of the country and the second is the "system" - political, institutional, judicial, economic, the Government, specific politicians - to whom the blame is put and responsibilities are attributed.

In the Belgian case studies, researchers found a typical *topos* of hate speech, one that is very frequent in racist discourse in general. Posters tend to see a **victimisation of their own community** (religion, sexual orientation, ethnicity, etc.) and thus aim at defending it, notably by accusing exogenous communities.

In the majority of the xenophobic posts, they also detected a use of **inaccurate generalisation**, where things that belong to different domains are lumped together.





In the Italian case studies the authors, reworking Malinowski's classification, found three communicative functions that seem to be involved in hate speech:

- **Pragmatic function.** Communication may intend to produce or coordinate actions, to act in a concerted way with other people or it may be itself an action that will produce real effects. In the reported comments, all the exhortations to "do" something towards (against) immigrants are gathered into this kind of function; they are imperative statements like "we rise up!", "let's kick out them", "send them away" that can have even stronger and violent tones.

- **Identifying function.** Communication may be a way to perform identity: we show others who we are, presenting an image of ourselves. We do it intentionally when we use terms and concepts that reveal the way we would like others to see us or talk about our status, our social roles; but it also occurs in an indirect and unintentional way whenever we show constitutive aspects of our identity. Most of the collected comments tend to perform this identity function showing a varied array of elements: the bravery of saying and the clarity of mind – "they should all be exterminated", "death penalty is necessary!"; "I would beat up all of them" –; the knowledge of the person who supports his/her own ideas with data and historical circumstances; sarcasm aimed at inspiring a feeling of sympathy in the community – "Foreigners go?!? Noooo, now that we could fire them as we want!"; wisdom – "we are not surprised if racism is rampant among Italians: foreigners have never brought anything good".

- **Social function.** Interpersonal communication can serve to establish relationships; it can be used to create, maintain or change social relations and the mutual recognition among people. Similarly, in online chats, the author of a comment may need to express his/her own opinion because he/she is trying to share some ideas creating a sort of community that uses the Other as a mirror. This function seems to be well expressed in these comments aiming at "opening the eyes" to other readers: "but don't you realize that they are invading us?!", "don't you notice that we cannot even walk in peace on the street?".





7. The profile of hate speakers

The authors of aggressive comments do not seem to have particular characteristics: they are men and women, with different political ideas, resident in different regions, of different ages.

In the Italian study, researcher identified different “profiles” of hate speakers:

- *Resigned*: their expressions reveal a criticism to the country, a disappointment towards politicians, who should have done something for them and did not improve their quality of life. The resigned are those who, in this specific context and mood of opinion, pick on immigrants, but in the future, they could express hatred against any other “scapegoat”. Some readers, for example, attribute to foreign people some generic “sins” such as: stealing jobs to Italians, littering the city, exploiting the country, etc.
- *Angry*: behind some comments, it is possible to read a polemical and rancorous approach towards a lack of management of migration flows. The angry do not insult the migrant in him/herself, but what he/she represents. They express an angry point of view and vent their feelings, emotions, and the discomfort of a forced cohabitation that, above all, institutions do not manage. The angry tend to highlight the “more favourable” treatment that local and national administrators concede to immigrants in comparison with natives: if they commit crimes, nobody punishes them adequately, if they are involved in tax evasion, they are not persecuted, etc.
- *Aggressive*: this attitude does not stand on actual contents. Migrants are violently attacked - without making a specific argument - with a derogatory and offensive “verbal” behaviour and negative labels. The aggressive person rails against the foreigner, with a particularly abusive terminology, criticizing: health conditions, morality, civic behaviour, somatic and physical traits, (mainly economic) opportunism, cultural marginality and low intelligence, occupation and invasion of a country which does not belong to them and much more. The racist hate speech of these readers is more difficult to manage, because the aggressive does not see in the migrant a person to deal with, but only an obstacle to be removed.





IV. Editors, journalists and external players' opinions and strategies

The following section is based on more than 60 qualitative, semi-structured interviews, conducted in 2015 with editors/editors in chief, journalists/bloggers, social media experts or managers, researchers, journalists' associations, community managers, representatives of public bodies (or the Ministry of the Interior in the Czech Republic), and anti-racism and anti-discrimination associations and NGOs.

Interviewees were asked to describe and discuss their attitudes about online participation, its impact and its control, the limits to freedom of speech, their strategies when it comes to regulation of discussions and what they deem to be the most innovative and promising approaches.

1. Journalists' attitudes about online participation

Editors tend to have one of two attitudes to sensitive topics: either they think people should be able to express their views on any topics, even the most sensitive ones, or they consider that as some topics invariably trigger hate reactions, it is better not to enable people to pour out their hatred. Among experts and representatives of NGOs and public bodies instead, there is a far wider consensus about the need of a careful management of public discussions.

For newspapers, the "traffic" generated by negative comments can be **economically convenient**. As the editor of a minor news website in the Czech Republic put it: *"I feel that for large sites, discussion is just a tool to increase web traffic [...] Some people do not even read articles and go straight to the discussions. It is enough for those sites because the user is on their website. He/she sees the ads, can click on them, and read them. I guess we do not all have the same goals. Of course, traffic matters if that is a site that makes money. The question is how far one should go."* None of the respondents representing online newspapers with a large number of readers would admit it, although all of them presumed that the increase in readership is one of the reasons why readers are allowed to discuss on the websites. According to most respondents, in the case of smaller newspapers this argument is not very relevant.

The argument often put forward by editors who are in favour of readers participation regards the wish to provide a space for readers to be able to **actively participate** and express their own opinions, so enriching the contribution of the news website. One of the respondents said he was confident about the purpose of the discussions, adding: *"The discussions often bring added value to the article, both on the informative side (which does not happen often) and by refining the point of view and opinions on the issue."* A well-represented opinion in the Czech Republic was that discussions help people to get out of their "bubbles" and get to know opinions other than those they are used to hearing due to communicating with like-minded people. In addition, the benefit of certain comments was noted. Comments may provide interesting information, refer to new facts or notify the author of the article/editorial team of a factual mistake. On the other hand, most respondents also agreed that valuable contributions that further develop the discussion are not very frequent. The editor in Chief of *Lesoir.be* (Belgium) is quoted saying that out of one hundred comments, only ten relate to the article and only one is interesting. So-called trolling, when people try to sow discord among posters, also does not contribute to the meaningfulness of the discussions. Therefore, some respondents add that discussions can only be meaningful if they are regulated.

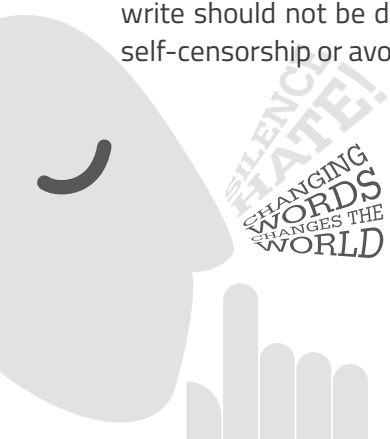


Apart from the issue of meaningfulness, we can ask whether discussions can **influence the overall atmosphere in society**. According to some respondents in the Czech Republic, discussions receive too much importance, e.g. the statement “gypsies to gas chambers” somewhere in the discussion below an article does not mean that someone would go out and kill Roma people. Nevertheless, according to most respondents, Facebook and some popular blogs in particular have great importance in this respect. *“Facebook has big influence, such statements receive wide publicity there,” said one respondent, adding: “With the advent of social networks racism did not increase, but it has been given far greater room to manoeuvre. These statements clearly create an atmosphere in which some idiot could snap and go and murder a Roma; the impression is created that all this is something normal and acceptable.”* One respondent, who is an officer of the Police of the Czech Republic and works in a team that is in charge of combating racist and extremist manifestations, said that anti-Roma demonstrations that took place during 2013 reached such a scale thanks to the Internet. The Internet became a means of connection and mobilization for people, as well as a means to assert one’s opinions.

Another question with regard to the impact of discussions relates to the extent to which they can retroactively affect readers’ perception of the article. Some of the respondents believe that this happens quite often. Unregulated discussions full of racist and xenophobic comments could also cause distaste and distrust among some readers towards certain online newspapers. Therefore, they would stop reading the articles appearing on them out of principle.

Impact on the reputation of the new outlet, instigation to discrimination and racism, meaningless discussions, off topic comments, all these common characteristics of online discussions bring some editors to consider online participation as a **disturbing element**, a “noise” to be avoided. Several topics in the Czech Republic are closed to comments in some news websites. In Belgium, all the media agree that the quality of comments is currently very mediocre and they either are resigned to that or are still trying to improve it. Interesting comments in fact are not very common, in part because of the comment format – which prevents a well-constructed argument from developing due to the limited number of characters (on some press websites) –and because everything happens instantly (on Facebook), which does not promote deep thinking.

In Belgium journalists explained that they do not generally adapt the content of their articles according to the hate comments they might attract, but they do **handle certain topics** with caution. Debates take place at *l’Avenir* offices about the use of the words ‘migrants’ or ‘refugees’ or the photographs used to illustrate their articles. For instance, they take care not to select a picture of a burka to accompany an article about wearing the veil. At *RTLinfo* for instance, they are careful not to mention the nationality of people involved in crime reports unless necessary. As many pieces published on Facebook receive comments without the users even reading the article, the headline and the splash headline introducing the news are very important. A tricky balance therefore has to be struck between attracting readers’ attention while avoiding descending into sensationalism, especially as articles are often published to a tight timescale. The editor interviewed at RTBF therefore has no hesitation in amending some splash headlines after they appear on Facebook, if some users react and she realises that her splash headline is too provocative or not sufficiently neutral. However, the way journalists write should not be dictated by possible hate reactions from their readers, nor should they have to practise self-censorship or avoid mentioning over-sensitive topics for fear of over-the-top outbursts, as *Le Vif* stressed.





2. The issues around freedom of speech

As we saw, the different stances about online discussion involve an assessment of the nature and the limits of freedom of speech. As Benoit Frydman wrote in a paper on freedom of speech published by the Human Rights League (Belgium): "The question of the status of speech inciting to hatred in democratic societies is a classic one which is both difficult and dangerous (...) because it requires us to probe deep into our system and our democratic passions. Traditionally classed under the philosophical question of whether freedoms should be granted to the enemies of freedom, the dilemma of the status of hate speech comes up in very concrete terms in the realms of politics and law"³³.

In this article, Frydman explains the different concepts **Europe** and the **United States of America** have with regard to freedom of speech. He uses the terms 'slippery slope' and 'fatal slope' to describe that difference. In Europe, our laws are based on the premise of the 'fatal slope' which "maintains that hate speech often, even inevitably, leads to acts of violence and sometimes to mass murder and it must therefore be punished to prevent or limit such violence"³⁴. The 'slippery slope' argument, on the other hand, which underlies the American position, "maintains that, in the absence of a precise criterion for determining what comes into the category of hate speech, [prohibiting hate speech] is likely to cover an infinitely extendable field, thus threatening public debate"³⁵. As they state on their embassy website, "the United States consider the most effective weapon against hate speech is not eliminating it, but having counter-speech which is tolerant, sincere and intelligent"³⁶.

Both of these concepts were discernible in most of interviews and in the journalists' opinions and questions about freedom of speech in Belgium. Similarly, in Italy, **open positions** toward readers' comments, in the name of freedom of expression, are **as equally expressed as restrictive attitudes** towards readers' participation.

In the opinion of the General Secretary of the Council for Ethical Journalism in Belgium, some of the accusations of censorship made against journalism are based on a **misconception of the nature of journalism** and the limits to freedom. Journalists are bound to comply with rules that are justified but which restrict their freedom of speech and this is something the public is often unaware of. The General Secretary explained that in fact, journalists have not the liberty to say whatever they wish, quite simply because journalism is not a matter of saying whatever one wants. In his view, journalism is not to be equated with exercising one's freedom of speech. He believes a journalist has a specific social function as an intermediary between whatever happens in society and the public, who have the right to know about things of relevance to society.

Journalists' job is not merely to relay everything that happens in society: they must assess, select, decide what is of general interest and check the details, because a lot of the information going around is false.

33- Frydman, B. (2013). "Pente glissante ou pente fatale". La Chronique de la Ligue des droits de l'Homme asbl, 155, pp. 9-10.

34- Ibid.

35- Ibid.

36- <http://iipdigital.usembassy.gov/st/french/pamphlet/2013/04/20130424146340.html#axzz3qwZhNBOA>.





Therefore, putting a halt to such false information is not an infringement of freedom of speech, but a socially useful thing to do. Among the limits imposed on journalists, there is the journalistic function, i.e. sorting, checking, putting things into perspective, etc.; the limits originating in the law, such as not identifying minors involved in court cases and not spreading hate speech; and lastly, the limits deriving from the code of ethics, such as sticking as closely as possible to the truth (not spreading rumours or unchecked information) and respecting other people's privacy.

A slightly different position comes from another Belgian expert, the Administrator of the Human Rights League, who stresses that as long as we remain within the sphere of debating ideas, there should be no limits to freedom of speech. It is only **when a text incites people to act**, advocates discriminatory attitudes **or spreads hate** that we leave the sphere of debating ideas and enter a realm that has to do with disturbing the peace and therefore justifies the intervention of the legislator and, then, the judge.

Remarks not amounting to 'incitement' may perhaps seem absurd, outrageous or unacceptable, but we must not expect the judge to take the place of the historian, the politician or the polemicist: remarks that are unacceptable or deliberately distorted must be condemned and combated with means from the sphere of debating ideas. He points out that it is extremely perilous to legislate on the limits to freedom of speech, as the substance is thus likely to be lost. It is important to bear in mind that freedom of speech is, in a way, freedom to harm: the freedom to shock, scandalise, appal others – otherwise, it is meaningless. This special feature of freedom of speech, which unlike other freedoms does not stop where that of others begins, explains why it is one of the freedoms challenged most often.

On certain conditions, therefore (existence of a law, measures that are necessary in a democratic society and measures aimed at upholding basic rights), freedom of speech can be restricted. However, there is at present in Belgium no law or legal text a judge can take as a basis in order to prohibit the publication or broadcasting of texts, sound or pictures *before the event*. This is the meaning of the judgment delivered by the European Court of Human Rights on 29 March 2011 in the case of *RTBF versus the Belgian State*. Prohibiting publication before the event therefore violates the Belgian constitution (articles 19 and 25) and the European Convention on Human Rights (Article 10 (1)). However, penalties can be imposed after the event, notably in cases of incitement to hatred, discrimination, segregation or violence on the ground of protected criteria and revisionism.

While the limits to freedom of speech appear straightforward, in practice it is more complicated than it appears to differentiate between an opinion and incitement to violence. In addition, the editors in chief and moderators interviewed in Belgium all agreed that moderating involves a large degree of subjectivity, even though the law is quite clear as regards incitement to hatred.

That **subjectivity** has its origin, notably, in the fact that some editors decide to remove comments that contain stereotypes, incorrect information or abuse or which are simply poorly expressed. At RTBF for example, racist comments like "Send them all back!" will be hidden, but comments that express an opinion, like "Personally I think they should all go back home" will be left. On the other hand, people inciting hatred with comments such as "Let them all drown!" will be banned. At *Le Vif*, someone noted that quite often, certain remarks do not necessarily fall foul of the law, but they will be removed all the same because the journalist doing the moderating considers them too shocking or too hurtful. If a comment is taken down, the reason is that it is offensive or that it does not contribute anything to the debate. Therefore, the decision will depend on the journalist who is moderating at the time.





The editors at the Belgian daily *L'Avenir* have decided not to tolerate comments that spread mistaken ideas (for example, the misconception that refugees do not immigrate into Muslim countries, but only into Europe), so their red line is quite a bit higher than the legal limit. According to the editors, that decision has created a **problem of representativeness** in terms of the range of comments. For instance, taking the refugee issue as an example, those who argue against the refugees generally comment in large numbers, but they either leave out some of the facts, they use abuse and stereotypes, or they are aggressive. As a result, their comments are regularly removed, whereas those who argue in favour of the refugees generally post comments that are worded in a calm and measured way and well argued, and so they will not be removed. This ultimately creates a problem in terms of how opinions are represented in the final depiction of the comments, as only the best-constructed messages will be left up, whereas there may have been ten times as many with a different opinion that could not be left up because of problems with the way they were expressed. Moderating in this way often causes some conscience searching, as it creates an imbalance in the perception reflected by the readership.

Let us see now what editors and journalists actually do or think it should be done.

3. The different strategies of discussions' regulation

Almost all the interviewees stand by the opinion that all online article comment sections must be moderated, with a very few exceptions in Italy and the Czech Republic. The tools of regulation mentioned by the interviewees are the following:

- **Public Rules or Policy.** To begin with, probably the most common practice is to establish a set of provisions – a “netiquette” or a “public policy” – with criteria that regulate comments and modalities of interaction between the editorial staff and the readers. The establishment of explicit rules allows replying to users according to shared principles, avoiding arbitrariness, and most if not all of the media outlets studied use it.

- **Pre-moderation.** A few newspapers decided to analyse all comments before their publication. This is the case of the weekly *Respekt* in the Czech Republic and of *Il Fatto Quotidiano* in Italy: starting from spring 2014, all readers' comments must be first approved. Pre-moderation is facilitated by a platform / alert that is based on a customised algorithm and a users' categorization, corresponding to different levels of control.

- **Active moderation.** It is very rare to see moderators really interacting with readers on matters of substance, i.e. the content of the article. At most, readers are thanked for reporting an error or, on some press websites, the moderator steps in to calm things down if the remarks start to go off track, but overall the interaction remains at a very superficial level. However, even this minimal form of proactive moderation shows that there is someone behind the web page and people cannot just say whatever they want. If readers have the perception that a form of control by the newspaper exists, they are encouraged to express themselves in a more correct and respectful way, at least according to the experience of some social media managers. Interventions must be prompt, so that overheated comments do not reinforce each other in an escalation of verbal violence.





■ **Removal of comments.** In most news websites in Italy, this practice is usually limited only to extreme cases, because of the implications in terms of freedom of expression, while in the other countries it seems to be more common. In Belgium, and probably in Germany, for most editors managing reader participation means precisely removing comments that do not comply with the law or the newsgroup rules. On the opposite side, in Italy and the Czech Republic, some editors have chosen not to remove posts thinking that cancelling comments is too severe a limit to freedom of speech.

■ **Forbidden words.** Most of the newspapers whose journalists have been interviewed in Italy and in the case studies in the Czech Republic use this family of computer management systems. They allow detecting hate speech and replying automatically or blocking an account, using customized and constantly updated keywords, this way simplifying the work of the staff in charge of the community management.

■ **Time limited discussions.** Contributions to discussions are permitted in some websites only for a certain period (e.g. 24-48 hours).

■ **Users' 'classification'.** Many newspapers support the decision to make registration to the news site mandatory before being able to comment. Some social media strategists feel that knowing the community of users is an important step, in order to become aware of their own public via a database of the "hard core" of commenters being therefore able to prevent difficult situations.

■ **Rewards and loyalty programs.** High-quality texts can be facilitated by using loyalty programs, in which users have a certain number of points that can be increased or removed if they violate the rules of discussion. Rewarding well-constructed comments is a way of setting an example for other users to follow and encouraging the more hesitant users to give their views. There are various ways of 'rewarding' or giving prominence to comments, such as having a vote for the 'comment of the week' or taking a screenshot of the most interesting comments and posting it on the social media. This kind of approach has been attempted only seldom.

■ **Not opening or closing a discussion.** Recently in the Czech Republic, editors have decided in several cases not to open a discussion on 'hot' topics or, in Belgium, not to public an article on a sensible topic on the Facebook page, where moderation is more difficult. Curiously, La Stampa in Italy followed an opposite strategy, closing comments on the website moving all moderation efforts on Facebook. The motivation behind this decision is that a newspaper is legally responsible for the contents of its website. According to one of the experts interviewed, this is a resigned choice, since it means that the newspaper is not able to handle interactions with readers on its website. However, other newspapers adopted this strategy in other countries, in some cases changing their minds.

37- <http://www.wan-ifra.org/reports/2013/10/04/online-comment-moderation-emerging-best-practices>





All these strategies have in common a “No” approach, in that they limit their scope to the prevention, the dissuasion or the punishment of undesirable behaviour. In Belgium, a different, proactive approach has been experimented. Why not to sort all the comments and publish only the most interesting ones, the ones that truly fuel the debate? This idea is in line with the old-style readers’ letters in the printed newspapers, and it provides as many questions as it does answers. Firstly, because pre-moderation takes much time and thus human resources; secondly, because it raises all the questions about freedom of speech: is not there the risk that only those opinions that concur with the newspaper’s views will receive space?

The RTL info website enables readers to display, within the thread, the ‘RTL info selection’, which is chosen by the editorial team because the comments contribute to the debate. This feature of the site shows a more positive view of reader participation, even if it turns out that the site uses it seldom and that most of the time it promotes no comments. People at La Libre say they also use regularly this feature of promoting comments when they think they add something to the topic under discussion. As they point out, moving certain comments higher up is above all a way of moving other, more disreputable, ones to a position further down the thread.

Appealing for readers’ reactions and giving them prominence would be a way of raising the standard of quality and cutting down on hate speech. However, this solution takes too much time and effort to be pursued on a daily basis and it is not applicable to every topic (it has to be a major social issue that many people have something to say about).

Is it the journalist’s role to chair debates and online communities? Can journalists be administrators and community moderators, or they should give the job to other professionals? The various online press editors have different solutions. Some newspapers prefer an in-house management, others entrust the management to an external team or a specialized company. Some use professionals who have a specific expertise on the web, for example for the management of their social networks. Others prefer to have journalists who write the articles, or journalists who are members of the editorial staff, in order to have the same “tone of voice” and to involve them directly in launching news, also taking advantage of their ability to reply to readers. In websites where thousands of posts appear daily the management of such a large number of comments cannot be handled by people who would do this as a secondary activity in addition to editorial work. In the case of smaller sites, reporters are generally mainly responsible for regulation of discussions.

An original approach, one based on a mixed group of journalists and web professionals, was adopted in Italy by Fan Page, which created a “social media staff” composed of 10 people with a strong experience in both journalism and social networks who regularly exchange and collaborate with the editorial staff. The social media staff identifies the trending topics and the editorial staff stresses the strategic contents.





4. Innovative approaches to the regulation of internet discussions

It can happen that moderators do not remove comments that leave out some of the facts or that contain incorrect information because other users have corrected or challenged the information. They consider this form of social control important as it results in a better balance within the comments. Administrators and community moderators often report this method.

More in general, some journalists think that **readers can also play a part** in cutting down the amount of hate speech online. This is the case at *La Libre*, in Belgium, where the thinking is that web users should step up to their responsibility as citizens by refusing to accept hate speech and reporting abuse to moderators. In their view, readers are the initial filter against hate speech, as they are the ones reading the articles and the comments. Using that 'social control' factor, in other words, relying on interactions between readers to moderate, self-check, denounce racist remarks and report abuse, needs to be put into perspective though because, as indicated above, it is more complicated for users to report problem comments on Facebook than it is on the press websites that have their own systems. The risk here is that if it is not easy for the user to access a channel for reporting problematic content, this removes even more responsibility from the media, which can always claim they were unaware that they were hosting illegal content. It is thus clear that it is necessary to reach a balance, as both the media and the readers have a share of responsibility with regard to hate speech appearing online.

Nevertheless, there are other ways of giving readers' contributions prominence apart from press forums or Facebook posts. The aim of some Belgian websites has been **diversifying the forms of participation**. One example is the Opinions page at *La Libre* or *Le Soir's* Carte Blanche page, which allow readers to develop their views in a more structured format than a comment, as their text is published on the website like an article. The *RTL* website has an "Alert us" page, where it invites readers to send in their first-hand accounts of events that concern them personally, and these sometimes form the basis of further stories. It would indeed appear to be the case that allowing readers to share their experiences on specific topics encourages them to express themselves in a more constructive way.

According to many respondents, the purpose of the discussions is not only getting feedback, but also **building a community**. One of the most important tools against hate speech is a good relationship with one's online community, as a healthy and critical community is capable of reacting to hateful comments on their own, without a constant interference of a mediating party. A community that shows solidarity and speaks up loudly against hateful comments, while staying factual and objective, is a powerful tool.





So how can we support serious posters and build a community when, in the opinion of most respondents, a large amount of people commenting on their sites do not really care about the real debate? How to attract the “decent and serious” debaters who would have something to say on the topic? According to several Czech respondents, it is important that the sites give up on a marketing strategy when trying to get as many readers as quickly as possible. Instead, they should focus on qualitative changes such as building a community of loyal and decent readers. This can only be achieved with very active discussion moderation. The moment the community establishes itself and develops some standards, it will all work through self-regulation. Among the monitored news websites, *iDnes.cz* is doing relatively well in this respect. As one of the survey respondents said, ultimately, this strategy increases readership, because a serious discussion stimulates the motivation to read the article and find out more information on the matter.

According to some respondents, interference with anonymity and introduction of complicated registration processes are also mistakes. Complicated registration and being incapable of posting anonymously not only deter haters and trolls, but also serious debaters who could sway the discussion in another direction.

In Italy, some experts and community managers pointed out another reason why newspaper should take care of the community to select the type of users they want: “a large number of readers who post many negative comments is not useful, not even in terms of marketing. The reputation is fundamental, because only faithful readers buy the newspapers and make subscriptions, not those who write comments on the web page to denigrate articles”. Nonetheless, this statement does not seem to put into the picture online advertising, the most important source of revenues of news websites and one that probably does not discriminate between faithful/decent readers and occasional/aggressive ones.

For some journalists, deleting comments should be used as a last resort, as a hate comment can be a good opportunity to “educate”. According to a German media expert, people are very “brave” behind a computer screen but when their comments are made public, when they are laughed at and commented on by the community, then they usually stop to post hate speech in the future or change platforms. The representatives of a consumers’ advisory centre and of an initiative for civil society and democratic culture in Germany said that users who stand up to haters and react by using factual information should be rewarded by the media organisations, by liking the comments and highlighting them, so that the favoured tone on the platform is clear to any new user and commentator. The initiative for civil society has created several brochures, manuals and many more publications as educational tools that users can use to react to hate speech. They put a big focus on creating pamphlets that have an all-embracing view on topics such as racism, homophobia, immigration and more. This gives users factual arguments when they want to react to hate comments.





V. Recommendations

In the framework of this project, a two-days meeting gathering European social media managers was held in Brussels on May 11 and 12, 2015. This meeting aimed at exchanging experiences and acquiring new knowledge about the management of discriminatory and racist posts and comments. The concrete output of this exchange consists in guidelines with recommendations for community and social media managers.

Guidelines for community and social media managers

- 1. Define your policy and make it known:** it is fundamental to establish the rules of exchange in the community. These rules should be simple and clear. When abusive comments are made, do not hesitate to remind the contributor of the rules. While moderating, taking decisions based on publicly known rules enhances transparency.
- 2. Engage and value your community:** being a journalist nowadays is as much about community building as it is about news reporting. If community members play an active and positive role, then they will be more likely to protect it against hate speech, by reporting offensive comments and feeding it with interesting and relevant contents.
- 3. Establish and maintain good relations with your community:** develop a good relationship with members of your community, for example, by welcoming newcomers. It is a good idea to promote positive and respectful relations from the beginning.
- 4. Be part of the discussion:** monitoring and fostering comments is a good strategy to guide the debate and positively influence users. Asking a question can also guide users in their online participation.
- 5. Showcase good practices by users:** rewarding good comments is very important: it sets an example for other users and encourages shy users to express their views. There are several ways of rewarding comments, for example, by establishing an election of the "comment of the week", or taking a screenshot of the best comments and posting them on your social media pages.

38 - The following Social Media Managers and experts took part in the European meeting: Simona Berankova (Czech Republic - aktualne.cz), Nadia Ferrigo (Italy - La Stampa), Martin Hoffmann (Germany - welt.de), Michal Ischia (Czech Republic - Respekt), Vittorio Pasteris (Italy - Expert/blogger), Marcus Pelz (Germany - Via Verkehrsgesellschaft), Anthony Planus (Belgium - Le Vif), David Schmidt (Germany - Zeit Online), Damien Vanachter (Belgium - Labdavanac), Valentina Vellucci (Italy - La Stampa), Michal Zlatkovsky (Czech Republic - Respekt). The meeting was facilitated by Média Animation and four experts in online hate speech (Reeta Pöyhtäri - University of Tampere, Finland) and online audience engagement (David Domingo, Sylvain Malcorps, Laura Ahva - University of Brussels, Belgium).

A stylized illustration of a hand with a pointing finger. The hand is grey and has a simple, curved line for a smile. The text 'CHANGING WORDS CHANGES THE WORLD' is written in a bold, sans-serif font, with 'CHANGING' and 'WORLD' in a larger size. Below it, the words 'SILENCE IS HATE!' are written in a smaller, bold, sans-serif font. The background is white with a light blue gradient.

CHANGING
WORDS
CHANGES THE
WORLD

SILENCE
IS
HATE!



6. Share positive experiences with your newsroom: share with your colleagues in the newsroom particularly interesting or sensitive exchanges and questions that arise in the community, in order to increasingly involve them in discussions on good moderating practices and their benefits.

7. Use irony and be creative: using irony and self-irony in moderating exchanges and responding to abusive comments and hate speech can be an effective way of defusing tensions and keeping discussions under control. Be creative in the management of the community and find original ways of reminding participants of the rules of exchange.

8. Know where you are: be aware of the differences between commercial social networking sites and media organizations' social networking sites. Develop moderating practices and exchange / comments standards for each type.

9. Differentiate public and private conversations: in order to maintain the public comments less tense in tone and contents, consider the possibility of isolating hate speech through private contacts with those directly involved. You can, for example, suggest further exchanges via private e-mail or telephone.

10. Take your decisions: do not be afraid to take all necessary steps to ensure that the rules of the community are respected including, for example, banning users who repeatedly post racist and violent comments.



www.bricksproject.eu



Co-financed by the Fundamental Rights and Citizenship Program of the European Union.