

Rights, Equality and Citizenship (REC)
Programme of the European Commission
(2014-2020)



Monitoring and Detecting Online Hate Speech

Agreement Number: JUST/2014/RRAC/AG/HATE/6652

D1.5: Advisory Board Meeting 2 [†]

Abstract: This deliverable reports the proceedings of the 2nd MANDOLA Advisory Board.

Contractual Date of Delivery	30 September 2017
Actual Date of Delivery	18 December 2017
Deliverable Security Class	Public
Editor	Nikos Frydas
Contributors	Evangelos Markatos Meltini Christodoulaki Estelle De Marco
Quality Assurance	Vivi Fragopoulou

[†] This project is funded by the Rights, Equality and Citizenship (REC) Programme of the European Commission.

The *MANDOLA* consortium consists of:

FORTH	Coordinator	Greece
ACONITE	Principal Contractor	Ireland
ICITA	Principal Contractor	Bulgaria
INTHEMIS	Principal Contractor	France
UAM	Principal Contractor	Spain
UCY	Principal Contractor	Cyprus
UM1	Principal Contractor	France

Document Revisions & Quality Assurance

Internal Reviewers

1. Vivi Fragopoulou (FORTH)

Revisions

Version	Date	By	Overview
1.5.4	18/12/2017	Editor	Re-submitted to AB members for explicit permission to publish their details & contributions. Various amendments / corrections were received and incorporated.
1.5.3	18/11/2017	Editor	Submitted to AB members for comments, on 6/11/2017, but no further amendments / corrections were received to date.
1.5.2	6/11/2017	Editor	Amendments / corrections from Internal Reviewers
1.5.1	2/10/2017	Editor	First draft.

Table of Contents

Document Revisions & Quality Assurance	3
Table of Contents.....	4
1 Executive summary	6
2 Background to the MANDOLA project	7
2.1 MANDOLA objectives	7
2.2 MANDOLA activities	8
2.3 More MANDOLA material	8
2.3.1 Deliverables	8
2.3.2 Presentations	9
2.3.3 Publications in Journals & Conferences.....	9
2.3.4 Working Documents.....	9
3 Aims & Objectives of the Advisory Board (AB)	10
3.1 The Objectives of the MANDOLA AB	11
3.1.1 AB duties in general	11
3.1.2 AB duties in particular	11
3.2 AB Constraints.....	12
3.3 AB Membership	13
3.4 Methodology used to populate AB2	14
4 Proceedings of the AB2	15
4.1 Welcome/Introduction/Advisory Board	16
4.2 Short Review of the MANDOLA Results	17
4.3 Short Presentations by AB members	18
4.4 The MANDOLA Dashboard & Mobile Application	23
4.5 Privacy Impact Assessment (PIA) of the MANDOLA outcomes	24
4.6 A short review of the Landscape analysis and introduction to Mandola Stakeholder Survey	25
4.7 Brainstorming Panel.....	26
4.7.1 Question 1.....	26
4.7.2 Question 2.....	28
4.7.3 Question 3.....	30

4.7.4	Conclusions	32
5	Conclusions & Lessons Learned	34
6	Appendix A: Agenda of AB2 (Advisory Board Meeting 2)	35
7	Appendix B: AB2 presentation by Evangelos Markatos	38
8	Appendix C: The MANDOLA Dashboard & Mobile Application	43
9	Appendix D: Privacy Impact Assessment of the MANDOLA outcomes	63
10	Appendix E: A short review of the Landscape analysis and introduction to Mandola Stakeholder Survey.....	70
11	Appendix F: Brainstorming Panel / Question 1.....	84
12	Appendix G: Brainstorming Panel / Question 2	85
13	Appendix H: Brainstorming Panel / Question 3	86

1 Executive summary

The current document reports the proceedings of the 2nd MANDOLA Advisory Board which took place in Brussels, on 7 September 2017 (10-17:00), in the *Office of the Spanish National Research Council*. The Advisory Board comprised nine external and five internal members.

The aim of the Advisory Board was to discuss and offer feedback to selected areas of project deliverables, as well as to debate possible follow-up for MANDOLA.

This document, comprises the following chapters:

1. **Chapter 2** [*Background to the MANDOLA project* in p. 7]: This chapter offers the background to the MANDOLA project. It may be useful to readers unfamiliar with the project. More material about the project may be found at the project site (<http://mandola-project.eu/publications/>).
2. **Chapter 3** [*Aims & Objectives of the Advisory Board (AB)*, in p. 10]: This chapter describes the aims & objectives of the Advisory Board, as well as the practical constraints taken into consideration, when examining Advisory Board candidates.
3. **Chapter 4** [*Proceedings of the AB2*, in p. 15]: This chapter gives the proceedings of the Advisory Board Meeting 2 (**AB2**).
4. **Chapter 5** [*Conclusions & Lessons Learned*, in p. 34]: This chapter gives the conclusions and lessons learned from AB2. Important new issues that emerged from the discussion included the difficulty of defining and countering hate speech, as well as recent increased awareness about the subject.
5. The document includes the following **appendices**:
 - a. Appendix A: Agenda of AB2 (Advisory Board Meeting 2)
 - b. Appendix B: AB2 presentation by Evangelos Markatos
 - c. Appendix C: The MANDOLA Dashboard & Mobile Application
 - d. Appendix D: Privacy Impact Assessment of the MANDOLA outcomes
 - e. Appendix E: A short review of the Landscape analysis and introduction to Mandola Stakeholder Survey
 - f. Appendix F: Brainstorming Panel / Question 1
 - g. Appendix G: Brainstorming Panel / Question 2
 - h. Appendix H: Brainstorming Panel / Question 3

[All images used in this document have either been created by the Editor, or obtained through *creative commons*.]



2 Background to the MANDOLA project ¹

MANDOLA (Monitoring AND Detecting OnLine hAte speech) is a 24-months project cofounded by the Rights, Equality and Citizenship (REC) Programme of the European Commission, which aims at making a bold step towards improving the understanding of the prevalence and spread of online hate speech and towards empowering ordinary citizens to report hate speech.

2.1 MANDOLA objectives

The MANDOLA specific objectives are the following:

- To monitor the spread and penetration of online hate-related speech in the European Union (EU) and in the E.U. Member States using big-data approaches, while investigating the possibility to distinguish, among monitored contents, between potentially illegal hate-related speech and non-illegal hate-related speech;
- To provide policy makers with actionable information that can be used to promote policies for mitigating the spread of online hate speech;
- To provide ordinary citizens with useful tools that can help them deal with online hate speech irrespective of whether they are bystanders or victims;
- To transfer best practices among E.U. Member States;
- To set-up a reporting infrastructure that will enable the reporting of potentially illegal hate speech.

The MANDOLA project addresses the two major difficulties in dealing with online hate speech: the lack of reliable data and the poor awareness on how to deal with the issue. Indeed, it is difficult to find reliable data that can show detailed online hate speech trends (inter alia in terms of geolocation and in relation to the focus of hate speech). Moreover, available data generally do not distinguish between potentially illegal hate speech and not illegal hate speech. In addition, the different legal systems in various Member States make it difficult for ordinary people to perceive the boundaries between both these categories of content. In this context, citizens might have difficulties to know how to deal with potentially illegal hate speech and how to behave when facing harmful but not illegal hate content. The lack of reliable data also prevents to make reliable decisions and push policies to the appropriate level.

The two MANDOLA innovations are (1) the extensive use of IT and big data to study and report online hate, and (2) the research on the possibility to make a clear distinction between legal and potentially illegal content taking into account the variations between E.U. Member States legislations.

MANDOLA is serving: (1) policy makers - who will have up-to-date online hate speech-related information that can be used to create enlightened policy in the field; (2) ordinary citizens - who will have a better understanding of what online hate speech is and how it evolves, and who will be provided with information for recognising legal and potentially illegal online hate-speech and for acting in this regard; and (3) witnesses of online hate speech incidents - who will have the possibility to report hate speech anonymously.

¹ The content of this chapter has been taken from Deliverable D2.4b (http://mandola-project.eu/m/filer_public/d7/bd/d7bd3a35-f9b5-418e-af55-74539d17eddf/mandola_d24b4_20170930.pdf).

2.2 MANDOLA activities

In order to achieve its objectives, the project includes the following activities:

- An analysis of the legislation on illegal hate-speech at the European and international level and in ten E.U. Member States.
- An analysis of the applicable legal and ethical framework relating to the protection of privacy, personal data and other fundamental rights in order to implement adequate safeguards during research and in the products to be developed.
- The development of a monitoring dashboard, which aims to identify and visualise cases of online hate-related speech spread on social media (such as Twitter) and on the Web.
- The creation of a multi-lingual corpus of hate-related speech based on the collected data, to be used to define queries in order to identify Web pages that may contain hate-related speech and to filter the tweets during the pre-processing phase. The vocabulary is developed with the support of social scientists and enhanced by the Hatebase (<http://www.hatebase.org/>).
- The development of a reporting portal, in order to allow Internet users to report potentially illegal hate-related speech material they have noticed on the Internet.
- The development of a smart-phone application, in order to allow anonymous reporting of potentially hate-related speech materials noticed on the Web and in social media.
- The creation and dissemination of a Frequently Asked Questions document, to be disseminated via the project portal and the smart-phone app.
- The creation of a network of National Liaison Officers (NLOs) of the participating Member States. They are intended to act as contact persons for their country, to exchange best practices and information, and to support the project and its activities with legal and technical expertise when needed.
- The development of a landscape of current responses to hate speech across Europe and of a Best Practices Guide for responding to online hate speech for Internet industry in Europe.

2.3 More MANDOLA material

The project site (<http://mandola-project.eu/>) contains more information about the project, as well as all the publishable documents (<http://mandola-project.eu/publications/>):

2.3.1 Deliverables

1. D1.1: Dissemination Plan (3/2016)
2. D1.2: Midterm Dissemination Report (10/2016)
3. D1.3: Final Dissemination Report (9/2017)
4. D1.4: Advisory Board Meeting (10/2016)



5. D2.1: Intermediate Report - Definition of Illegal Hatred and Implications (7/2016)
6. D2.1b: Definition of illegal hatred and implications (final report) (9/2017)
7. D2.2: Identification and analysis of the legal and ethical framework (7/2017)
8. D2.3: Legal and ethical compliance of the MANDOLA research (9/2017)
9. D2.4a: Private Impact Assessment of the MANDOLA outcomes (7/2017)
10. D2.4b: Privacy Impact Assessment of the MANDOLA outcomes (final report) (9/2017)
11. D3.1: MANDOLA Monitoring Dashboard (9/2016)
12. D3.2: Reporting Portal (10/2016)
13. D3.3: Smartphone App (5/2017)
14. D4.1: FAQ on Responding to on-line hate speech (7/2016)
15. D4.1b: FAQ on Responding to on-line hate speech (9/2017)
16. D4.2: Best Practice Guide for Responding to Online Hate Speech for Internet Industry (3/2017)
17. D4.3: Mandola WS4 Workshop with Stakeholders (8/2017)
18. D4.4: Landscape and Gap Analysis (8/2017)
19. D4.5 Stakeholder Survey (9/2017)

2.3.2 Presentations

1. Evangelos Markatos. MANDOLA: Monitoring and Detecting on-line Hate Speech. MANDOLA Workshop, Montpellier France, February 2017.
2. Estelle De Marco. The criminalisation of Hate Speech: limits and comparative study of the laws from 10 European Union's member state. MANDOLA Workshop, Montpellier France, February 2017.
3. Demetris Paschalides. Technologies to detect, analyse and report online hate speech: the Mandola experience. MANDOLA Workshop, Montpellier France, February 2017.
4. Ioannis Inglezakis. The criminalisation of the criticism of religion. MANDOLA Workshop, Montpellier France, February 2017.
5. Ioannis Inglezakis. Hate and xenophobic speech on the Internet. In REDA 2015: Regulation and Enforcement in the Digital Era. Cyprus, November 2015.

2.3.3 Publications in Journals & Conferences

1. Marios Dikaiakos, George Pallis and Evangelos Markatos. Mandola: [Monitoring and Detecting Online Hate Speech](#). ERCIM News No. 107 (Special Theme: Machine Learning) p49. October 2016

2.3.4 Working Documents

1. D2.1: Intermediate Report - Definition of Illegal Hatred and Implications



3 Aims & Objectives of the Advisory Board ([AB](#))

This chapter describes the aims & objectives of the Advisory Board, as well as the practical constraints taken into consideration.

The *aim* of the task undertaken is to compose the *optimum AB*, under the *practical constraints* of the project.

The Chapter comprises the following sections:

1. The Objectives of the MANDOLA AB
2. AB Constraints
3. AB Membership



3.1 The Objectives of the MANDOLA AB

Setting up an Advisory Board “that will steer this project” is the goal of WS1.3. The delivery of the following outputs is part of the project’s contractual obligations:

1. D1.4 Advisory Board Meeting 1 Target group: ALL
2. D1.5 Advisory Board Meeting 2 Target group: ALL

The current document constitutes deliverable D1.5.

3.1.1 AB duties in general

In general, an Advisory Board provides non-binding strategic advice. Among the reasons for creating an AB are the following:

- Seek expertise outside MANDOLA.
- Complement existing strengths.
- Counsel on issues raised by MANDOLA.
- Become a resource for MANDOLA managers.
- Provide un-biased ideas.
- Monitor project performance.

3.1.2 AB duties in particular

According to the MANDOLA project objectives, the Advisory Board should have the following characteristics:

- AB will **steer** the project.
- AB will help **spread** the project message well **beyond** participant Member States.
- AB will assist the **promotion** of the developed technologies and tools.
- AB will provide valuable **feedback & market guidelines** on progress & results.
- AB will further **enhance** impact & **dissemination** of MANDOLA’s ideas.
- AB will foster dialogue & **debate**.
- AB will serve as a source of **expertise**.



3.2 AB Constraints

Project constraints place an upper limit of **20** to the number of external AB members who reside outside Brussels. In addition, the **AB members must be EU residents**.

The meeting room made available has a capacity of 25. This implies that with a total of six internal AB members, the **external AB members should be restricted to 19**.

In addition, AB2 aims at discussing and offering feedback to selected areas of project deliverables, as well as debating possible follow-up for MANDOLA. Given that the members' participation was required on three distinct items of the Agenda [see *Appendix A: Agenda of AB2 (Advisory Board Meeting 2)*, p. 35], it was decided to restrict the number of AB2 members to 12-15.

MANDOLA project partners are grateful to the ***Office of the Spanish National Research Council***, Rue du Trône, 62, Brussels, who made their meeting room available, free of charge.



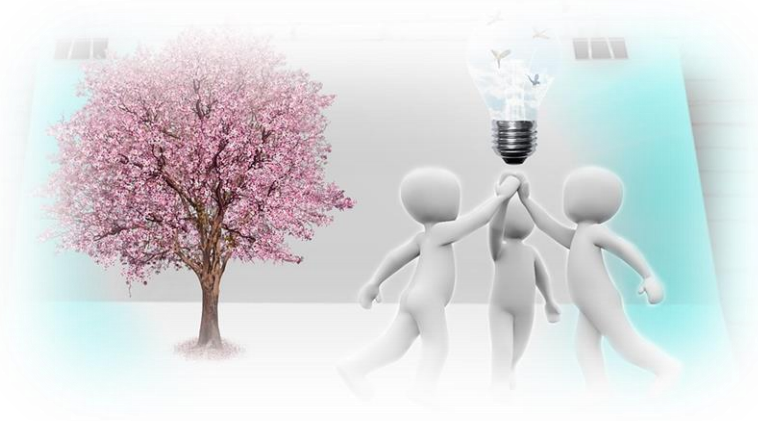
3.3 AB Membership

In general, AB members must be individuals

1. with personal qualities and
2. representing an *important* entity, where *important* is understood to mean *important for the project*, and
3. with knowledge of the issues the project deals with and
4. with good command of English and
5. with the ability to be present at the AB meetings in Brussels.

Given the above and the project objectives (see *The Objectives of the MANDOLA AB*, above), AB members shall then be drawn from:

- Academia
- NGOs
- LEA
- Internet Industry
- Government
- Other



3.4 Methodology used to populate AB2

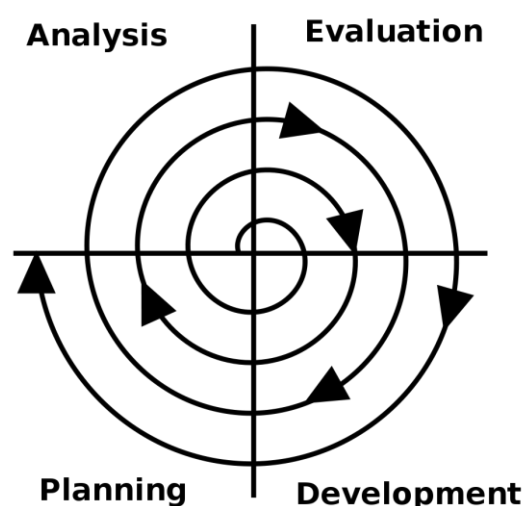
This section describes the methodology used to populate the Advisory Board. The material is taken from Chapter 3 of MANDOLA deliverable D1.4.

The process used to populate AB2 was simply to select from those individuals who were invited to AB1, the top 12-15. Chapter 3 of MANDOLA deliverable D1.4 discusses the AB candidates ranking methodology so that AB is balanced across four attributes: Personal expertise, type of members' organizations, nationality & gender. So, ranking DOES NOT reflect our opinion on candidates' competence. Only 12-15 were chosen so that there was time enough for all AB2 members to discuss, debate and contribute on three distinct items of the Agenda, as explained in Section 3.2 (see p. 12).

The methodology used to select AB candidates is to create a super-list through MANDOLA members' recommendations and Internet search and then narrow-down as following:

1. Create a super-list of 50-60 individuals, candidates for the AB.
2. Assess the suitability of each individual across a number of attributes.
3. Combine the marks/attribute into an overall score/individual.
4. Order the individuals according to their score.
5. Invite the top 16 individuals.
6. Once an individual accepts an invitation, the individual is moved to the top of the list.
7. Once an individual declines the invitation, the individual is moved to the bottom of the list.
8. Continue until you have 16 acceptances.

For more information, please see pp. 9-17 of Chapter 3 of MANDOLA deliverable D1.4.



4 Proceedings of the AB2

This chapter gives the proceedings of the Advisory Board Meeting 2 (**AB2**). The chapter will be partitioned into the AB2 Agenda items [see *Appendix A: Agenda of AB2 (Advisory Board Meeting 2)*, p. 35].



4.1 Welcome/Introduction/Advisory Board

Dr. Nikos Frydas from FORTH, a MANDOLA consortium partner, welcomed the AB members and went briefly through the Agenda items.

Following that, each AB member introduced him/her-self.

The Table below lists the AB2 members' surnames, names, organization, position and e-mail.

The field *Ext* indicates if the members is **Internal**, or **External**.

		
Second MANDOLA Advisory Board Meeting September 7, 2017 <i>Office of the Spanish National Research Council (Room 3), Rue du Trône, 62, Brussels</i> Meeting Agenda		
10:00-10:15	Welcome/Introductions/Advisory Board	Nikos Frydas
10:15-10:30	Short Review of MANDOLA results	Vangelis Markatos
10:30-11:15	Short Presentations by AB members	AB members
11:15-11:30	Coffee Break	

Surname	Name	Organization	Position	Ext	e-mail
Baider	Fabienne	University of Cyprus	Associate Professor	Ext	fabienne@ucy.ac.cy
Belavusau	Uladzislau	T.M.C. Asser Institute / University of Amsterdam	Senior Researcher in European Law	Ext	U.Belavusau@uva.nl
Callanan	Cormac	AIS, Ireand	CEO	Int	cc@aconite.com
Cummiskey	Siobhan	Facebook	Policy Manager, EMEA	Ext	scummiskey@fb.com
De Marco	Estelle	INTHEMIS, France	Director, Senior researcher	Int	estelle.de.marco@inthemis.fr
Dikaiakos	Marios	Univ. of Cyprus	Professor of Computer Science	Int	mdd@cs.ucy.ac.cy
Dzsinich	Gergely	CyCap	Partner	Ext	g@dzsinich.com
Frydas	Nikos	FORTH, Greece	External Consultant	Int	nfrydas@cantab.net
Inglezakis	Ioannis	Aristotelean University of Thessaloniki, Law School	Associate Professor	Ext	iingleza@law.auth.gr
Le Toquin	Jean-Christophe	CYAN, Cybersecurity and Cybercrime Advisers Network	President	Ext	jcletoquin@socogi.fr
Lemaire *	Sarah	www.ceji.org	Project assistant	Ext	sarah@ceji.org
Markatos	Evangelos	FORTH, Greece	Head, Distributed Computing Systems Laboratory	Int	markatos@ics.forth.gr
Mitrou	Lilian	Aegean University	Associate Professor	Ext	l.mitrou@aegean.gr
Van den Reeck	Mark	Hamogelo tou Paidiou	Head of International Coopera	Ext	marcvandenreeck@hamogelo.gr

- * **Ms. Sarah Lemaire** is not with www.ceji.org anymore. Any enquiries should be directed, instead, to Ms. Melissa Sonnino, *FacingFacts* Project Coordinator, at melissa.sonnino@ceji.org.

4.2 Short Review of the MANDOLA Results

Prof. Evangelos Markatos from FORTH, the project leader, made a short review of MANDOLA results. The main points of the presentation are:

1. What do we want to do in MANDOLA?
2. Why?
3. How is Hate speech measured?
4. Dashboard – Hatemap
5. Dashboard – Hotspot
6. FAQs
7. Reporting Portal
8. Legal issues

10:00-10:15	Welcome/Introductions/Advisory Board	Nikos Frydas
10:15-10:30	Short Review of MANDOLA results ←	Vangelis Markatos
10:30-11:15	Short Presentations by AB members	AB members
11:15-11:30	Coffee Break	

For the presentation see *Appendix B: AB2 presentation by Evangelos Markatos*, in p. 38.



4.3 Short Presentations by AB members

AB2 members were invited in advance to prepare a short (4-5 mins) presentation, or speech, on their work on hate speech.

Most of them kindly responded, and some sent their presentations in advance, even though they were not able to participate to AB2, due to last minute unforeseen complications.


10:00-10:15	Welcome/Introductions/Advisory Board	Nikos Frydas
10:15-10:30	Short Review of MANDOLA results	Vangelis Markatos
10:30-11:15	Short Presentations by AB members ←	AB members
11:15-11:30	Coffee Break	

Most members gave written permission to make available their presentation. This material has been uploaded in the MANDOLA common space in the cloud.

The following presentations / speeches were made:

Baider	Fabienne	University of Cyprus	Associate Professor, Coordinator of the
--------	----------	----------------------	--

**Expressing Hate:
Some Cross-European
Perspectives on Hate speech**

Co-funded by the Rights, Equality and
Citizenship (REC) Programme
of the European Union

Identity and Goals

1. Prof. **Baider** introduced and discussed C.O.N.T.A.C.T., Creating an Online Network, monitoring Team and phone App to Counter hate crime Tactics. This is a “DG Justice Action project (Oct. 2015-Sept. 2017) aiming to address hate crime, with particular focus on hate speech”. It is Coordinated by the University of Cyprus.

The consortium comprises five Universities and seven NGO partners from Cyprus, Malta, Greece, Spain, Italy, UK, Romania, Poland, Lithuania & Denmark.

C.O.N.T.A.C.T.,’s goals were to develop reporting mechanism, to conduct training and to undertake research.

The presentation is available to AB

members, here:

<https://www.dropbox.com/sh/b8uye9l84xh6lwy/AABEu325cQs5R7Rm425GUFcPa?dl=0>

Belavusau	Uladzislau	T.M.C. Asser Institute / University of Amsterdam	Senior Researcher in European Law
-----------	------------	--	-----------------------------------




ASSER INSTITUTE   **UNIVERSITY OF AMSTERDAM**

MELA
Memory Laws in European and Comparative Perspective

Hate Speech and Genocide Denialism in European and Comparative Perspectives

Presentation during MANDOLA Expert Meeting
Dr. Uladzislau Belavusau
7 September 2017

Today 

- Introduction
- Major research themes
- Hate speech in my research
- Hate speech and phenomenon of historical revisionism
- Major publications

2. Dr. **Belavusau's** presentation was on *Hate Speech and Genocide Denialism in European and Comparative Perspectives*.

MELA (**ME**emory **LA**ws in European and Comparative Perspective), a consortium of four organizations was briefly introduced.

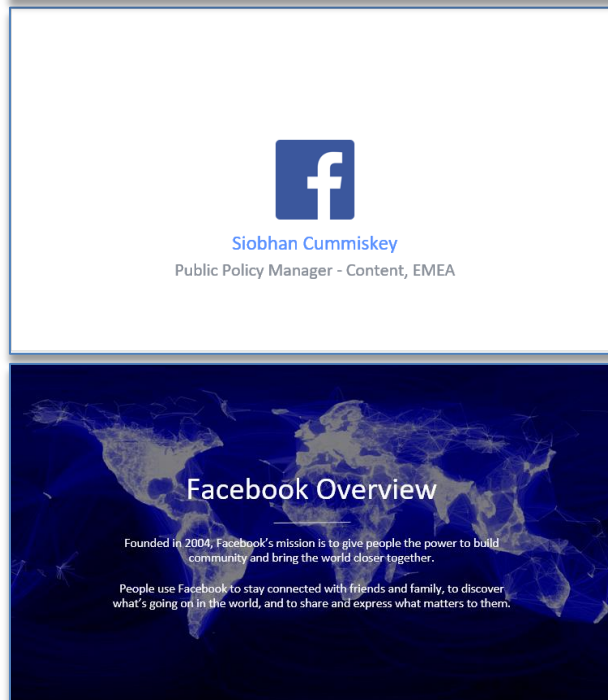
A distinction was made between the American (protecting the freedom of expression) and the European (protecting non-discrimination) model.

Historical Revisionism was also discussed, as well as its relation to Hate Speech.

The presentation is available to AB members, here:

<https://www.dropbox.com/sh/b8uye9l84xh6lww/AABEu325cQs5R7Rm425GUFcPa?dl=0>

Cummiskey	Siobhan	Facebook	Policy Manager, EMEA
-----------	---------	----------	----------------------

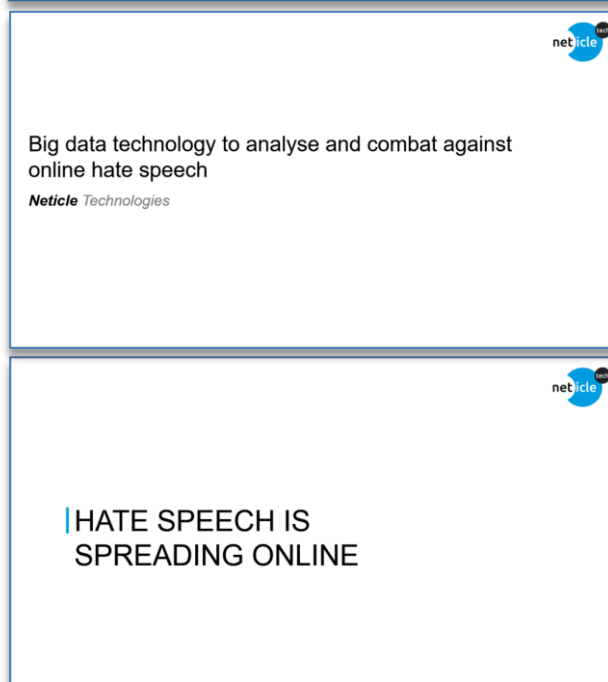


3. Ms. **Cumminskey's** presentation touched

upon the following issues:

- Facebook's community standards,
- Stakeholders Facebook cooperates with.
- How Facebook removes hate speech.
- EU Code of Conduct on Hate Speech
- Reporting posts
- How to improve?

Dzsinich	Gergely	CyCap	Partner
----------	---------	-------	---------



4. Dr. **Dzsinich** spoke on behalf of Neticle

Technologies about *Big data technology to analyse and combat against online hate speech*.

Issues discussed include:

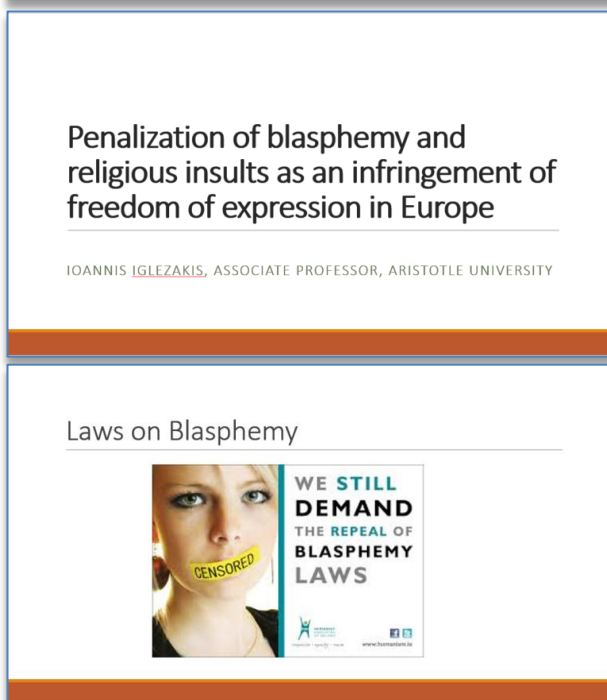
- Hate speech spread online:
Measurements of anti-jewish, anti-roma and anti-immigrant content on the Hungarian web, versus time.
- Measure and take actions based on media data.
- Neticle – A potential solution to measure hate speech online.
- Neticle media Intelligence – How it works.

The presentation is available to AB

members, here:

<https://www.dropbox.com/sh/b8uye9l84xh6lww/AABEu325cQs5R7Rm425GUFcPa?dl=0>

Inglezakis	Ioannis	Aristotelean University of Thessaloniki, Law School	Associate Professor
------------	---------	---	---------------------



5. Prof. **Inglezakis** discussed the topic *Penalization of blasphemy and religious insults as an infringement of freedom of expression in Europe*.

The topics of his presentation were:

- a. Reactions against Blasphemy
- b. The purpose of blasphemy laws
- c. Limitations of the right to freedom of expression
- d. European Court of Human Rights
- e. Council of Europe - Report of the Committee on Culture, Science and Education of 8 June 2007
- f. Venice Commission

The presentation is available to AB members, here:

<https://www.dropbox.com/sh/b8uye9l84xh6lww/AABEu325cQs5R7Rm425GUFcPa?dl=0>

Lemaire	Sarah	www.ceji.org	Project assistant
---------	-------	--------------	-------------------



6. Ms. **Lemaire**, presented *CEJI – a Jewish Contribution to an Inclusive Europe*.

Topics covered include:

- a. Diversity education
- b. Engaging Jewish communities
- c. Intercultural dialogue
- d. Advocacy
 - i. The eMORE project
 - ii. Facing facts – “online courses on monitoring hate speech and hate crime - www.facingfacts.eu”

The presentation is available to AB

members, here:

http://prezi.com/smvzbaki1fup/?utm_campaign=share&utm_medium=copy&rc=ex0share

Van den Reek	Mark	Hamogelo tou Paidiou	Head of International Cooperations
--------------	------	----------------------	------------------------------------

7. Mr. **Van den Reek** spoke about *The Smile of*

the Child ('Smile') and its direct interest on hate speech, as the national operator for Greece of child assistance and emergency lines:

“*Smile* is initiator and currently holding secretariat and presidency of EAN (European Anti-bullying Network, set up in 2014), where a debate is coming up as to the question whether or not there is/ought to be a tendency towards osmosis between cyberbullying and hate speech”.

“The development of bullying is remarkably on the rise as well because of the cyber phenomenon. It has significantly lowered the threshold for perpetrators and has somehow brought both phenomena of bullying and hate speech closer to one another”.

“The question lies hence in the dilemma whether hate speech could or should be incorporated in antibullying programs as 'new forms of bullying'. Some believe it should, others are very reticent to negative. As said, the debate will shortly be on the agenda of EAN as well. *Smile*', for one, believes that the answer tends to be negative. Handling of individual bullying is an issue on its own and should not be blurred nor further complicated by broader and complex issues of hate speech”.



presentations

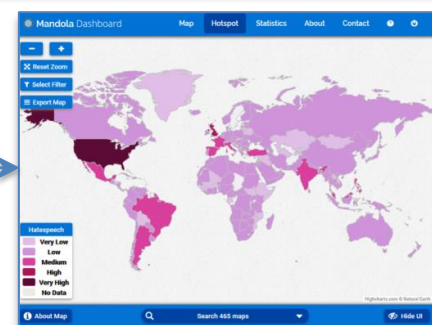
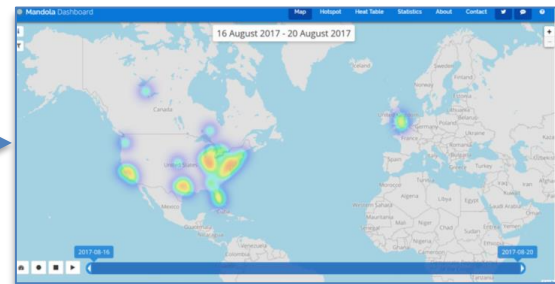
4.4 The MANDOLA Dashboard & Mobile Application

Prof. Marios Dikaiakos from UCY, a MANDOLA consortium partner, presented the MANDOLA Dashboard and the MANDOLA Mobile applications.

The presentation, available to MANDOLA members, comprised the following topics:

11:15-11:30	Coffee Break	
11:30-11:55	Short Presentations by AB members (continued)	AB members
11:55-12:10	Short Presentations of the MANDOLA Dashboard & Mobile applications	Marios Dikaiakos & George Pallis
12:10-13:10	Privacy Impact Assessment (PIA) of the MANDOLA outcomes (short presentation & discussion) ¹	Estelle De Marco

1. Hate speech definition
2. Major Web platforms on Hate-speech
3. Monitoring Dashboard
4. Data Stream Collections
5. Twitter Data Stream
6. Web Data Stream
7. Hate-speech Data Analysis
8. Hate-speech Data Flow
9. Preprocessing
10. Hate-speech Classifier
11. Classification Algorithm
12. Hate-speech Data Storage
13. Multi-lingual Corpus
14. Social Scientists
15. Monitoring Dashboard Web application
16. Dashboard Hatemap
17. Dashboard Hotspot
18. Dashboard Heat-table
19. Dashboard Statistics
20. Dashboard Administrator Panel
21. Mobile Application
22. The Mandola “Bubble”
23. Reporting while browsing YouTube Mobile Application



For the presentation, see *Appendix C: The MANDOLA Dashboard & Mobile Application*, in p. 43.

4.5 Privacy Impact Assessment (PIA) of the MANDOLA outcomes

Dr. Estelle De Marco from INTHEMIS, a MANDOLA consortium partner, made a short presentation on the *Privacy Impact Assessment of the MANDOLA outcomes*.

During the 1st phase of her presentation, Dr. De Marco discussed the **Privacy Impact Assessment (PIA)**. In a broad sense, PIA is understood to mean:

Assessment of risks posed by a project, to the right to private life and to personal data protection, and more widely to the other rights and freedoms either exercised by individuals in their respective personal spheres, or restricted by extension because of a privacy limitation or a personal data processing.

The method used on PIA was based on other methods, work and recommendations, like methods designed in several projects (ePOOLICE, PIAF, VIRTUOSO), Guidelines on risk management (ENISA, EBIOS), The Article 29 Data Protection Working Party Guidelines on DPIA, the Article 35 of the GDPR / 26 of the Directive 2016, etc.

MANDOLA outcomes, subject to the PIA include

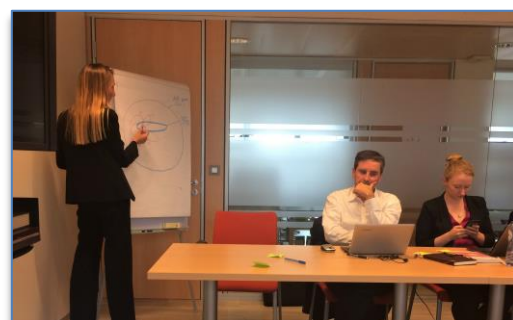
- the monitoring dashboard
- the smartphone app
- the reporting portal
- information dedicated to policy makers and the Internet Industry and
- information dedicated to Internet users

The discussion aimed at obtaining members' feedback on:

1. Section 4 Recommendations,
2. The elements of the content of the PIA, such as the identification of risks &
3. The methodology

For the presentation, see *Appendix D: Privacy Impact Assessment of the MANDOLA outcomes*, in p. 63.

11:15-11:30	Coffee Break	
11:30-11:55	Short Presentations by AB members (continued)	AB members
11:55-12:10	Short Presentations of the MANDOLA Dashboard & Mobile applications	Marios Dikaiakos & George Pallis
12:10-12:30	Privacy Impact Assessment (PIA) of the MANDOLA outcomes (short presentation & discussion) ¹ ←	Estelle De Marco



4.6 A short review of the Landscape analysis and introduction to Mandela Stakeholder Survey

Mr. Cormac Callanan from ACONITE, a MANDOLA consortium partner, made a short presentation on the Landscape analysis and also introduced the Mandela Stakeholder Survey.

13:10-14:10	Lunch Break	
14:10-14:25	A short review of the Landscape analysis and introduction to Mandela Stakeholder Survey (a short presentation) ³	Cormac Callanan
14:25-15:15	Brainstorming Panel on above topic	Cormac Callanan & Nikos Frivas

The *Landscape Document* focuses on the ongoing initiatives and on the current activities in Europe. It also includes a brief Gap Analysis.

It examines the following five countries:

- Bulgaria
- France
- Greece
- Ireland
- Spain

For each of the above countries, best practice in this field were highlighted, areas which need focus were determined and differences between EU member states were identified of different punishment for similar behaviour.



Finally, the Stakeholder Survey was introduced and explained. It comprises 29 questions and is also available in Spanish.

For the presentation, see *Appendix E: A short review of the Landscape analysis and introduction to Mandela Stakeholder Survey*, in p. 70.



4.7 Brainstorming Panel

In this session, four questions were given to the AB. For each question, the members wrote their answers on sticky notes, which were then collected, read, displayed and recorded for processing.

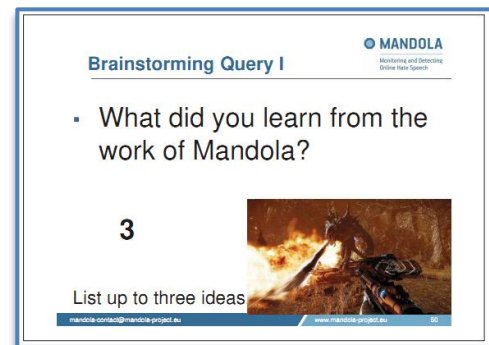
13:10-14:10	Lunch Break	
14:10-14:25	A short review of the Landscape analysis and introduction to Mandola Stakeholder Survey (a short presentation) ³	Cormac Callanan
14:25-15:15	Brainstorming Panel on above topic ←	Cormac Callanan & Nikos Evridas

4.7.1 Question 1

Question: *What did you learn from the work of Mandola? List up to three ideas.*

Answers:²

1. **Q1-1:**
 - a. **Cooperation** between various disciplines – aspects private and academic sector is extremely useful. **C**
 - b. Difficult to strike a **balance** between detection of hate speech and freedom of speech. **B**
2. **Q1-2:**
 - a. It is difficult to **measure** hate speech. **A**
 - b. It is difficult to **define** hate speech. **A**
 - c. It is difficult to **counter** hate speech. **B**
3. **Q1-3:**
 - a. Definition will also remain **difficult** after MANDOLA. **A**
 - b. MANDOLA offers a **platform** to act in practice to combat. **D**
 - c. Facts collection need to be translated into an evolutionary picture. MANDOLA contributes greatly. **D**
4. **Q1-4:**
 - a. There is **no easy** method to identify hate speech. **A**
5. **Q1-5:**
 - a. **Complexity** of legal difficulties. **A**
 - b. Possibility to develop innovative **apps**. **D**
 - c. Importance of review by people – automation is difficult. **F**
6. **Q1-6:**
 - a. **Complexity**. **A**
 - b. Variety of **stakeholders**. **C**
 - c. **More** work to do. **F**
7. **Q1-7:**
 - a. Is law the answer? **F**
 - b. **Interdisciplinary** needed. **C**
 - c. But each one selects their own field (only?). **E**



² Every participant has one 'vote'. Hence, if a participant gives n answers (n=1,2,...) to a question, then each of the member's answers carries a weight of 1/n. **Emphasis** is placed by the author and indicates the perceived keyword(s). Letters, e.g. **A**, etc., are added to indicate categorization of the participant's response.

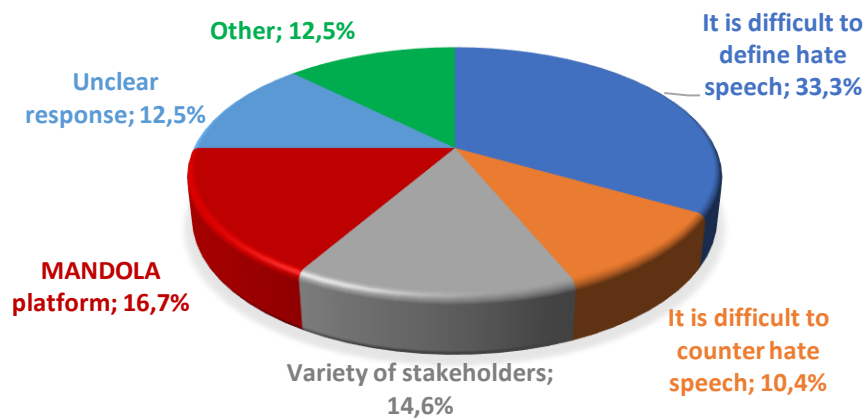
8. Q1-8:

- Issue of policy confidentiality. E
- Quick screenshot **button** to report hate speech for citizens. D
- Mixing type of hatred issue (ethnicity – nationality – sexual – gender). E

The above findings may be grouped as following [*What did you learn from the work of Mandola?*]:

A.	<i>It is difficult to define hate speech:</i>	$\frac{1}{3} + \frac{1}{3} + 1 + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 2\frac{2}{3}$	---	33.3%
B.	<i>It is difficult to counter hate speech:</i>	$\frac{1}{2} + \frac{1}{3} = \frac{5}{6}$	-----	10.4%
C.	<i>Variety of stakeholders:</i>	$\frac{1}{3} + \frac{1}{3} + \frac{1}{2} = 1\frac{1}{6}$	-----	14.6%
D.	<i>MANDOLA platform:</i>	$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1\frac{1}{3}$	-----	16.7%
E.	<i>Unclear response:</i>	$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$	-----	12.5%
F.	<i>Other:</i>	$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$	-----	12.5%

***What did you learn from the work of
Mandola?***



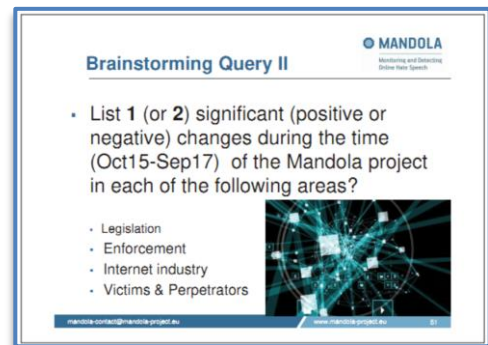
4.7.2 Question 2

Question: “List 1 (or 2) significant (positive or negative) changes during the time (Oct15-Sep17) of the Mandela project in each of the following areas:

- Legislation
- Enforcement
- Internet industry
- Victims & Perpetrators”

Answers: ³

- Q2-1:**
 - Legislation Net3 DG (negative). **E**
 - Internet industry **Code of Conduct** 2nd monitoring period results (positive). **A**
- Q2-2:**
 - There is much **more awareness** about hate speech. **C**
- Q2-3:**
 - Much stronger **public concern on illegal** content and need to discuss **proactive** measures. **C**
- Q2-4:**
 - Risk of **anti-migrant** hate speech in Europe. **D**
 - Tendency to **extend** grounds of **hate speech**, e.g. protection against homophobic speech in a number of European countries. **C**
- Q2-5:**
 - Lots of **projects** went on to make users report. **C**
 - Code of conduct** for Internet industry. **A**
 - Enforcements** proven not to be done. **B**
- Q2-6:**
 - Trump **F**
- Q2-7:**
 - Negative: Victims & perpetrators. **D**
 - Positive: Enforcement (restricted). **B**

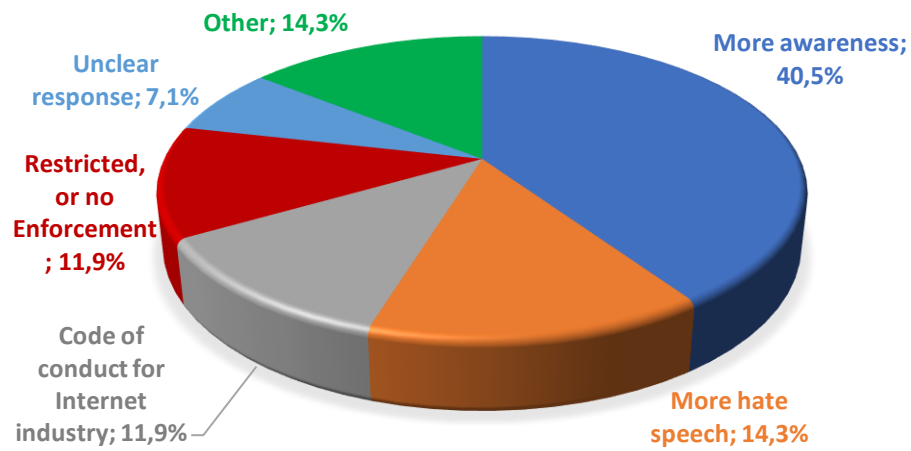


The above findings may be grouped as following [List 1 (or 2) significant (positive or negative) changes during the time (Oct15-Sep17) of the Mandela project]:

- | | | | | |
|----|---|--|-------|-------|
| A. | <i>Code of conduct for Internet industry:</i> | $\frac{1}{3} + \frac{1}{2} = \frac{5}{6}$ | --- | 11.9% |
| B. | <i>Restricted, or no Enforcement:</i> | $\frac{1}{2} + \frac{1}{3} = \frac{5}{6}$ | ----- | 11.9% |
| C. | <i>More awareness:</i> | $1 + 1 + \frac{1}{2} + \frac{1}{3} = 2\frac{5}{6}$ | ----- | 40,5% |
| D. | <i>More hate speech:</i> | $\frac{1}{2} + \frac{1}{2} = 1$ | ----- | 14.3% |
| E. | <i>Unclear response:</i> | $\frac{1}{2} = \frac{1}{2}$ | ----- | 07.1% |
| F. | <i>Other:</i> | $1 + = 1$ | ----- | 14.3% |

³ Every participant has one ‘vote’. Hence, if a participant gives n answers (n=1,2,...) to a question, then each of the member’s answers carries a weight of 1/n. **Emphasis** is placed by the author and indicates the perceived keyword(s). Letters, e.g. **A**, etc., are added to indicate categorization of the participant’s response.

List 1-2 significant changes during the time of the Mandola project

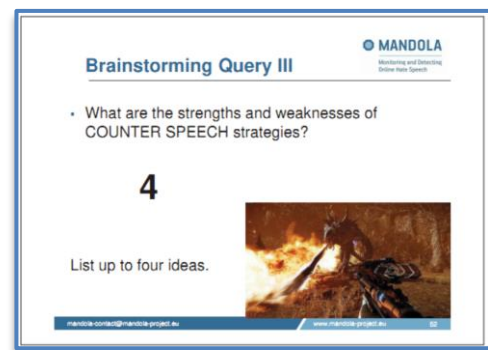


4.7.3 Question 3

Question: “What are the strengths and weaknesses of COUNTER SPEECH strategies? List up to four ideas”.

Answers: ⁴

1. **Q3-1:**
 - a. **Weaknesses:**
 - i. **Not enough** of it. E
 - ii. Hard to **measure** success. D
 - b. **Strengths:**
 - i. Potential to **change** hearts & mind. A
 - ii. More **effective** than delete. B
2. **Q3-2:**
 - a. **Weaknesses:**
 - i. It is defence, **not effective**. E
 - b. **Strengths:**
 - i. Response of a **community**. A
3. **Q3-3:**
 - a. **Weaknesses:**
 - i. Hard to know what to counter exactly. D
 - b. **Strengths:**
 - i. Way to make active and **responsible citizens**. A
 - ii. To not let hate speech not respond. C
 - iii. Remove hate speech won't make people think differently. B
4. **Q3-4:**
 - a. (Weaknesses): -
 - b. **(Strengths):**
 - i. Laugh 😊 (counteract hate speech with humour and statistics). B
5. **Q3-5:**
 - a. **Weaknesses:**
 - i. May be **difficult to mobilize** in countries where problems / awareness / edu is low. D
 - b. **Strengths:**
 - i. Probably the **best** & most effective **way** to combatting hate speech. B
6. **Q3-6:**
 - a. **Weaknesses:**
 - i. May lead to **confrontation**, flame wars. Not easy to implement. D
 - b. **Strengths:**
 - i. Seems to be **working** more than other approaches. B
7. **Q3-7:**
 - a. **Weaknesses:**
 - i. **Difficult** to implement. D



⁴ Every participant has one ‘vote’. Hence, if a participant gives n answers (n=1,2,...) to a question, then each of the member’s answers carries a weight of 1/n. **Emphasis** is placed by the author and indicates the perceived keyword(s). Letters, e.g. A, etc., are added to indicate categorization of the participant’s response.

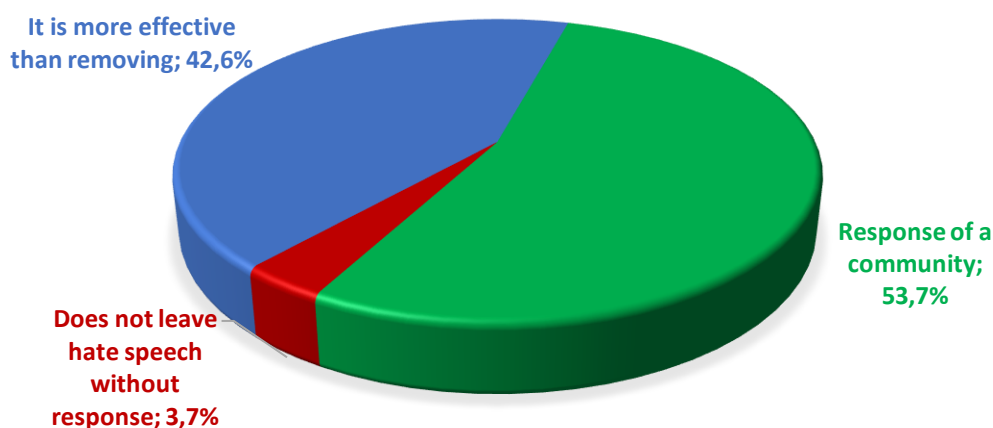
- ii. It has to come from people you trust. G
- b. **Strengths:**
 - i. Crowdsourcing approaches. A
- 8. **Q3-8:**
 - a. **(Weaknesses):**
 - i. **Difficult** to use the proper language / argumentation. D
 - ii. **Uncertain** if it reaches the right public/audience. E
 - iii. **Legitimizing** hate speech. F
 - b. (Strengths): -
- 9. **Q3-9:**
 - a. **(Weaknesses):**
 - i. No legal enforcement in severe cases. F
 - b. (Strengths): -
- 10. **Q3-10:**
 - a. **(Weaknesses):**
 - i. People can use hate speech words as **metaphor** without hate specific intent, without paying attention to it. D
 - ii. Current initiatives are **not going far enough**. G
 - iii. Governments and media share a huge part of **liability** in spreading. G
 - b. **(Strengths):**
 - i. **Education** to respect of others & others' right is fundamental in a multicultural society. A
- 11. **Q3-11:**
 - a. (Weaknesses): -
 - b. **(Strengths):**
 - i. Short (to be read), Concise and To the point (to have **impact**). B

The above findings may be grouped as following [*What are the strengths and weaknesses of COUNTER SPEECH strategies?*]:

- A. *Response of a community*: $1 + \frac{1}{3} + \frac{1}{2} + 1 + 1 = 3\frac{1}{6}$ ----- 42.6%
- B. *It is more effective than removing*: $\frac{1}{3} + \frac{1}{2} + 1 + 1 + 1 + 1 = 4\frac{1}{6}$ ----- 53.7%
- C. *Does not leave hate speech without response*: $\frac{1}{3} =$ ----- 03,7%



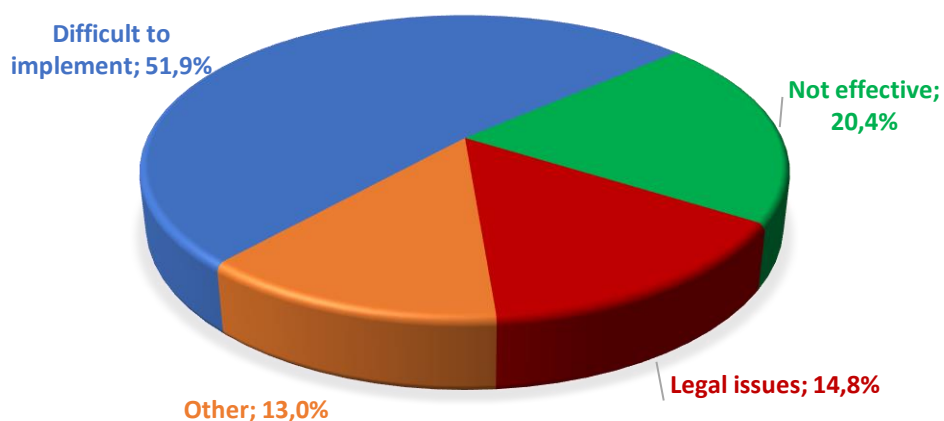
What are the strengths of COUNTER SPEECH strategies?



The above findings may be grouped as following [What are the *strengths and weaknesses* of COUNTER SPEECH strategies?]:

- D. *Difficult to implement*: $\frac{1}{2} + \frac{1}{3} + 1 + 1 + \frac{1}{2} + 1 + \frac{1}{3} = 4\frac{2}{3}$ ----- 51.9%
- E. *Not effective*: $1 + \frac{1}{2} + \frac{1}{3} = 1\frac{5}{6}$ ----- 20.4%
- F. *Legal issues*: $\frac{1}{3} + 1 = 1\frac{1}{3}$ ----- 14,8%
- G. *Other*: $\frac{1}{3} + \frac{1}{3} + \frac{1}{2} = 1\frac{1}{3}$ ----- 13,0%

What are the weaknesses of COUNTER SPEECH strategies?



4.7.4 Conclusions

One in three AB2 members, when asked *what did they learn from the work of Mandela*, answered that *it is difficult to define hate speech*. A further 17% referred to the platform developed by MANDOLA.

When asked about the *significant changes that occurred during the course of the MANDOLA project*, 40% of the AB2 members thought that there is now *more awareness* about hate speech, while a further 14% indicated that there is *more hate speech*.

Regarding the **strengths of the COUNTER SPEECH strategy**, AB2 members were almost split between “it is more effective than removing” (43%) and “it is good because it appears as the response of the community” (54%).

Finally, regarding the **weaknesses of the COUNTER SPEECH strategy**, half the members (52%) thought that it is *difficult to implement*, while a further 20% thought that it is *not effective*.



5 Conclusions & Lessons Learned

This chapter gives the conclusions and lessons learned from AB2.


The size and composition of an Advisory Board is very crucial for meetings its aims and objectives. During the course of the MANDOLA project, two methods were employed with success:

In **AB1**, participation was wide. As a result, and given the number of agenda items, AB1 members had a relatively short period of time to debate issues extensively. The ‘sticky-notes’ brainstorming sessions were very productive, though, and allowed for the collection of the sectoral experience of each member in a short period of time. In AB1 it was possible to conduct eight such sessions.

In **AB2**, participation was somehow restricted, in order to give the opportunity to the meeting to debate selected issues of interest to the project. The debate was very successful, and members took home a valuable feedback which is recorded in the final version of selected project deliverables. In addition, there were three sticky notes sessions, which focused on the project as a whole. For details see Chapter 4, in p. 15-34.



6 Appendix A: Agenda of AB2 (Advisory Board Meeting 2)

 Monitoring and Detecting Online Hate Speech		
Second MANDOLA Advisory Board Meeting		
September 7, 2017		
<i>Office of the Spanish National Research Council (Room 3), Rue du Trône, 62, Brussels</i>		
Meeting Agenda		
10:00-10:15	Welcome/Introductions/Advisory Board	<i>Nikos Frydas</i>
10:15-10:30	Short Review of MANDOLA results	<i>Vangelis Markatos</i>
10:30-11:15	Short Presentations by AB members	<i>AB members</i>
11:15-11:30	Coffee Break	
11:30-11:55	Short Presentations by AB members (continued)	<i>AB members</i>
11:55-12:10	Short Presentations of the MANDOLA Dashboard & Mobile applications	<i>Marios Dikaiakos & George Pallis</i>
12:10-13:10	Privacy Impact Assessment (PIA) of the MANDOLA outcomes (short presentation & discussion) ¹ The Discussion aims at obtaining members' feedback on: ² <ol style="list-style-type: none"> 1. Section 4 Recommendations, 2. The elements of the content of the PIA, such as the identification of risks & 3. The methodology 	<i>Estelle De Marco</i>
13:10-14:10	Lunch Break	
14:10-14:25	A short review of the Landscape analysis and introduction to Mandola Stakeholder Survey (a short presentation) ³	<i>Cormac Callanan</i>
14:25-15:15	Brainstorming Panel on above topic Questions to be discussed: <ol style="list-style-type: none"> 1. What did you learn from the work of Mandola? List up to three items. 2. List one (or two) significant (positive or negative) changes during the time (Oct15-Sep17) of the Mandola project in each of the following areas: <ol style="list-style-type: none"> a. Legislation b. Enforcement c. Internet Industry d. Victims and Perpetrators 3. What are the strengths and weaknesses of counter speech strategies? List up to four items. 	<i>Cormac Callanan & Nikos Frydas</i>
¹ The deliverable to be discussed (D2.4b) will be forwarded separately. ² See also attached APPENDIX. ³ The deliverable to be discussed will be forwarded separately.		



Monitoring and Detecting
Online Hate Speech

15:15-15:30 Coffee Break

15:30-15:45 Future Activities – Brief Introduction

Vangelis Markatos

15:45-16:50 Brainstorming Panel – Future Activities

*Vangelis Markatos
& Nikos Frydas*

Questions that may be discussed:

1. What shall we **do** with the developed MANDOLA monitoring technology? *Make it available for research, monitoring, or combating online hate speech, or evolve it to include more capabilities, etc.* Name up to three ideas.
2. What other IT tools may be useful in combating hate speech online more effectively? Name up to three ideas.
3. What other approach against hate speech online would be useful, apart from IT monitoring tools? *For example, a universal definition of hate speech, combating the origins of hate speech, focusing on the most dangerous categories of hate speech, etc.* Name up to three ideas.
4. How will hate speech evolve? Name up to three ideas.
5. What is the role of IT in combating future hate speech? Name up to three ideas.

16:50-17:00 Closing session

Vangelis Markatos



Monitoring and Detecting
Online Hate Speech

APPENDIX: Questions on the Privacy Impact Assessment

1. Please let us have your opinion in relation to the **Privacy Impact Assessment (PIA)** that has been performed on the MANDOLA outcomes (this is presented in Section 3 of the Deliverable D2.4b, forwarded separately).

You may express a general opinion (e.g. on its adequacy, structure, completeness, etc.) and/or focus on issues that are more important to you (such as the identification of risks), according to your experience, area of expertise and interest.

In case you would need further explanations on the PIA method that has been used, please refer to the MANDOLA [Deliverable D2.4a \(intermediate\) - Privacy Impact Assessment of the MANDOLA outcomes](#).

In case you would need further explanations on the notion and protection of fundamental rights considered within the framework of this PIA, please refer to the MANDOLA [Deliverable D2.2 - Identification and analysis of the legal and ethical framework](#).

2. Please let us have your opinion on the recommendations that conclude this PIA and which are summarised in Section 4 of Deliverable D2.4b.

You may express a general opinion and/or focus on issues that are more important to you, according to your experience, area of expertise and interest. In particular, we are interested in any positive or negative comments relating to one or more of the safeguards proposed in these recommendations (appropriateness, adequacy, lack of safeguard in relation to a particular issue...).

3. You are encouraged to share with us any other comment you would like to make in relation to Deliverables D2.4b, and / or Deliverables D2.4a and D2.2.

4. Please, in relation to the deliverable D2.4b, let us know:

- (1) If you would like to be named in relation with your comments, or prefer that these comments remain anonymously aggregated with the comments of the other experts, and
- (2) If you would like to be named at the end of the report in the list of the Advisory Board members who contributed to the PIA.
- (3) Accordingly, let us have the exact way you would like to be referred (ex. Dr.X, researcher, laboratory, company or University, country).

7 Appendix B: AB2 presentation by Evangelos Markatos

MANDOLA

**MANDOLA:
Monitoring and Detecting on-line
Hate Speech**

**Evangelos Markatos
FORTH and U of Crete**

Our Background

MANDOLA
Monitoring and Detecting
Online Hate Speech

- Partners of the project are active in
 - Cyber security
 - Cybercrime research and education
 - Illegal Internet Content Reporting



mandola-contact@mandola-project.eu www.mandola-project.eu 2



Our Background – Cyber security

MANDOLA
Monitoring and Detecting
Online Hate Speech

- SysSec: European Network of Excellence in Cybersecurity
 - Founding coordinator
 - Editor of Red Book in Cybersecurity
- NIS (Network and Information Security) Platform
 - Strategic Research and Innovation Agenda
- ECSO: European Cybersecurity Organization
 - Founding members (Spain and Greece)
 - (alt.) member of the Partnership Board







mandola-contact@mandola-project.eu www.mandola-project.eu 3

3

Our Background – Cybercrime Research and Education

MANDOLA
Monitoring and Detecting
Online Hate Speech

- National Centers of Excellence in cybercrime
 - Founding coordinators in
 - Spain, Greece, and Bulgaria
- Safeline: Greek Hotline to report illegal Internet content
 - Founder and first coordinator
- SENTER: European Network of National Centers of Excellence in cybercrime
 - Founding member and WP leader



mandola-contact@mandola-project.eu www.mandola-project.eu 4

4

What do we want to do in MANDOLA?

MANDOLA
Monitoring and Detecting
Online Hate Speech

- Monitor
 - the spread and penetration of
 - on-line hate speech in EU
- Call for proposals (from 2014) reads:
 - “Monitoring and reporting on hate crime and on-line hate speech (HATE)”


Mock up data

mandola-contact@mandola-project.eu www.mandola-project.eu 5


Why?

MANDOLA
Monitoring and Detecting
Online Hate Speech

- *If you can not measure it you can not improve it*

Lord Kelvin

- Measurements are at the core
 - of solid conclusions
 - and sound actions



mandola-contact@mandola-project.eu www.mandola-project.eu 6

How to measure it?

Twitter-based approach

1. Collect tweets
2. Apply a hate-detection filter
3. Apply sentiment analysis
 - To find hate speech (<http://www.nltk.org/>)



Mock up data

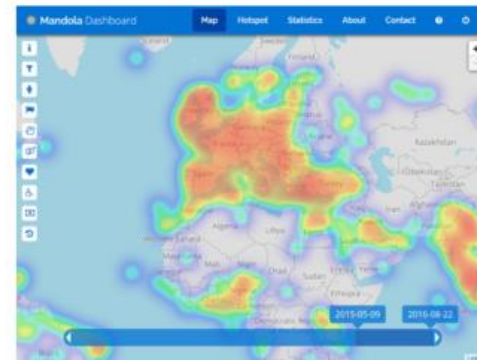
mandola-contact@mandola-project.eu

www.mandola-project.eu

7

7

Dashboard - Hatemap



These Visualizations are based on mock data and have nothing to do with real hate-speech analysis.

mandola-contact@mandola-project.eu

www.mandola-project.eu

8

8

Dashboard – Hotspot



These Visualizations are based on mock data and have nothing to do with real hate-speech analysis.

mandola-contact@mandola-project.eu

www.mandola-project.eu

9

9

Other activities? Frequently Asked Questions

- FAQ Book
- Landscape Analysis
 - Of current and future activities
- More at the talk at 14:10



mandola-contact@mandola-project.eu

www.mandola-project.eu

10

10

Reporting Portal

MANDOLA
Monitoring and Detecting
Online Hate Speech

- Where can you report Hate Speech?

European organizations where you can report hate speech incidents.



mandola-contact@mandola-project.eu www.mandola-project.eu 11

11

Other activities? Legal

MANDOLA
Monitoring and Detecting
Online Hate Speech

- Definition
 - What is hate speech
- Legal Framework
 - In Member States
- Policy-Impact Assessment
 - More at the talk at 12:10




mandola-contact@mandola-project.eu www.mandola-project.eu 12

12

Who?

MANDOLA
Monitoring and Detecting
Online Hate Speech

- FORTH (coordinator), GR
- U of Cyprus, CY
- AIS, IE
- ICITA, BG
 - Bulgarian Center of Excellence in cybercrime
- UAM, SP
 - Spanish Centers of Excellence in cybercrime
- UMO, FR
 - Member of the French CoE in cybercrime
- INTHEMIS, FR



mandola-contact@mandola-project.eu www.mandola-project.eu 13

Why are we here today?

MANDOLA
Monitoring and Detecting
Online Hate Speech

- We want your **advice**
 - Advisory Board
- You know a lot about this area
 - Can you share some of your **knowledge**?
 - Can you share some of your **experience**?
 - Some of your **wisdom**?



mandola-contact@mandola-project.eu www.mandola-project.eu 14

How are we going to get this advice?

MANDOLA
Monitoring and Detecting
Online Hate Speech

- Interactive Brain storming sessions
- Post it notes
- Share your advice
 - Even half-baked ideas
 - All ideas!
 - In research all ideas are welcome



mandola-contact@mandola-project.eu www.mandola-project.eu 15

15

Brain storming sessions

MANDOLA
Monitoring and Detecting
Online Hate Speech

- We are going to ask you questions:
 - e.g. "If you could pass one law in hate speech, what would it be?"
 - Think "out of the box"
 - Share your knowledge
 - Share your ideas



mandola-contact@mandola-project.eu www.mandola-project.eu 16

16

A final note

MANDOLA
Monitoring and Detecting
Online Hate Speech

- If you want to make a short presentation
 - Of your activities
 - Let me know
 - We have some slots
 - < 5 minutes long...



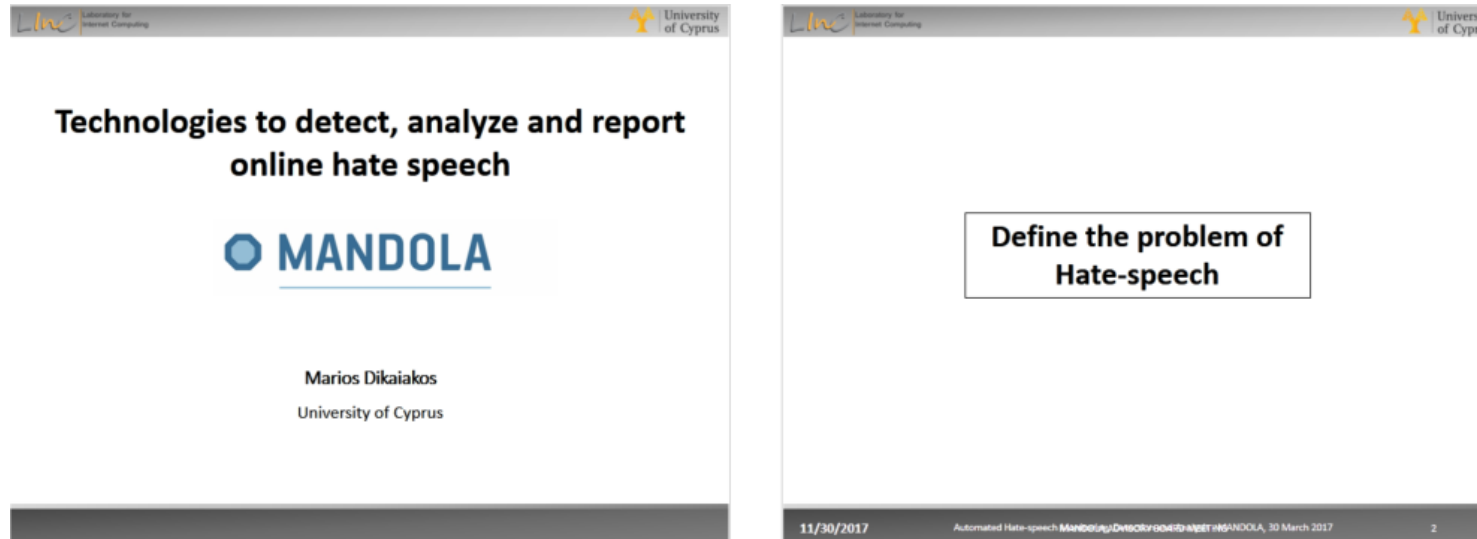
mandola-contact@mandola-project.eu www.mandola-project.eu 17

MANDOLA

Thank you

Evangelos Markatos
FORTH and U of Crete


8 Appendix C: The MANDOLA Dashboard & Mobile Application



University of Cyprus

Challenge

- Online **hate speech** represents a major challenge for Online Social Networking services and Web sites.



11/30/2017 MANDOLA ADVISORY BOARD MEETING 3

3

University of Cyprus

Barack Obama WIRED Magazine Editorial (11/2016)

"Now is the Greatest Time to Be Alive"



*That's how we will overcome the challenges we face: by unleashing the power of all of us... **Not just making our social networks more fun for sharing memes but also harnessing their power to counter terrorist ideologies and online hate speech.***

[Barack Obama: "Now is the greatest time to be alive" Wired, November issue, 2016](#)

11/30/2017 MANDOLA ADVISORY BOARD MEETING 4

4

University of Cyprus

What is Hate-speech?

- In the Oxford dictionary hate is defined as an emotion of "intense dislike", for someone or something, of an "aversion to" something.
- This feeling of "hatred or intolerance" can be verbalized in a speech (becoming this way a "hate speech") and used to express "hatred or intolerance of other social groups, especially on the basis of race or sexuality".
- As there cannot be a universal definition of Hate Speech, some web pages and social platforms have **adopted their own definitions of hate speech.**

11/30/2017 MANDOLA ADVISORY BOARD MEETING 5

University of Cyprus



"@USERNAME absolute bastard!"

No hate speech



"These faggot ass niggas..."

Hate speech

11/30/2017 MANDOLA ADVISORY BOARD MEETING 6

Major Web platforms on Hate-speech

- **Facebook** defines the term "hate speech" as "direct and serious attacks on any protected category of people based on their race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability or disease".
- **Twitter** does not provide its own definition, but simply forbids to "publish or post direct, specific threats of violence against others."
- **YouTube** website clearly says it does not permit hate speech, which it defines as "speech which attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status and sexual orientation/gender identity."
- **Google** makes a special mention on hate speech in its User Content and Conduct Policy: "Do not distribute content that promotes hatred or violence towards groups of people based on their race or ethnic origin, religion, disability, gender, age, veteran status, or sexual orientation/gender identity."

11/30/2017 MANDOLA ADVISORY BOARD MEETING 7

7

★

Our Objectives

- **Monitor** the spread of online hate-related speech.
- **Analyse** its content and the **categories** to which it might belong (*Ethnicity, Nationality, Religion, Gender, Sexual, Class, Disability*).
- **Store** and **visualize** actionable information for **policy makers**, to promote policies against online hate speech, and **citizens**, to raise their awareness.
- Do that **without** holding any user's sensitive data, by processing data on the fly, following a procedure approved by the Cyprus Data Protection Commissioner.

Article 7(1)(2) of the Personal Data Protection, N. 138(I)/2001

11/30/2017 MANDOLA ADVISORY BOARD MEETING 8

8



MANDOLA Tools

- A bundle of tools to tackle online hate-speech and meet our objectives.

Monitoring Dashboard

Mobile Application

11/30/2017 MANDOLA ADVISORY BOARD MEETING 9

Monitoring Dashboard

- **Consists of the following modules:**
 - Data Streams Collection (Twitter and Web pages)
 - Hate-speech Data Analysis (Multi-lingual Classifier)
 - Data Storage (Hate-speech statistical database)
 - MANDOLA Application Interface – API
 - Monitoring Dashboard Web Application


```

graph LR
    A[Data Streams Collection] --> B[Hate-speech Data Analysis]
    B --> C[Data Storage]
    C --> D[API]
    D --> E[Monitoring Dashboard]
  
```

11/30/2017 MANDOLA ADVISORY BOARD MEETING 10

Data Stream Collections

- Consist of two sub-modules :
 - Twitter data stream** is collected via Twitter Stream API using a framework to efficiently retrieve large collection of tweets per day (developed by University of Cyprus).
 - Web data stream** is collected via meta-search engine that crawls possible hate-related web pages (developed by University Autonomous of Madrid, Spain).
- Each collected stream is fed to Apache Kafka, a queuing system for supporting streaming data processing.

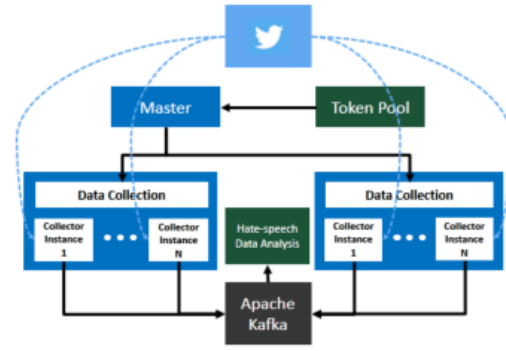


"Distributed Large-Scale Data Collection in Online Social Networks" H. Efstathiades et al. IEEE CIC 2016

11/30/2017 MANDOLA ADVISORY BOARD MEETING 11

11

Twitter Data Stream

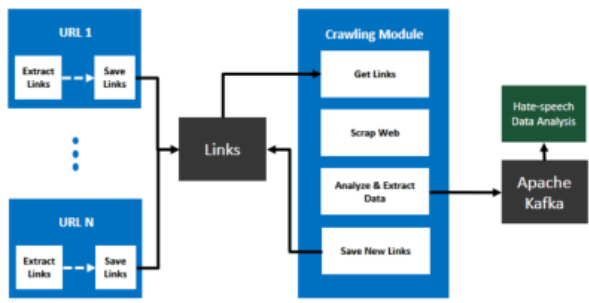


"Distributed Large-Scale Data Collection in Online Social Networks" H. Efstathiades et al. IEEE CIC 2016
"Online Social Network Evolution: Revisiting the Twitter Graph" H. Efstathiades et al. Best student paper award, IEEE Big Data 2016

11/30/2017 MANDOLA ADVISORY BOARD MEETING 12

12

Web Data Stream

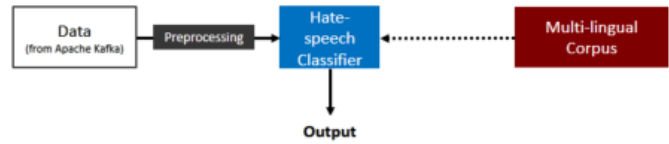


Prof. Alvaro Ortigosa
Marcos Hernando

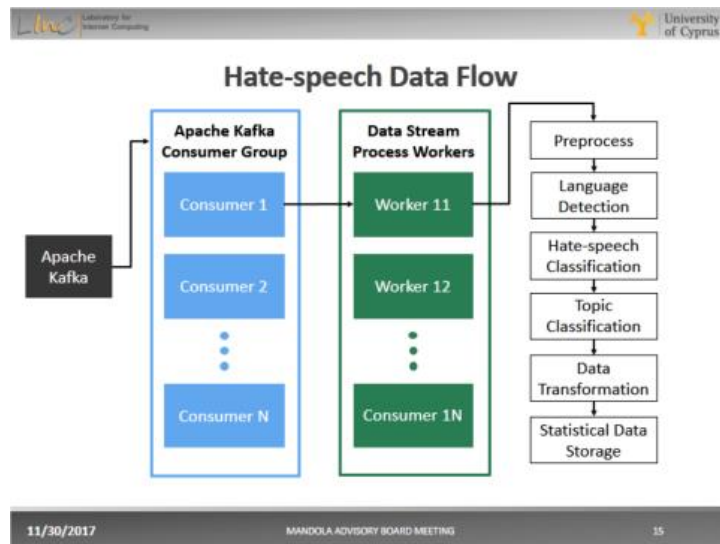
11/30/2017 MANDOLA ADVISORY BOARD MEETING 13

Hate-speech Data Analysis

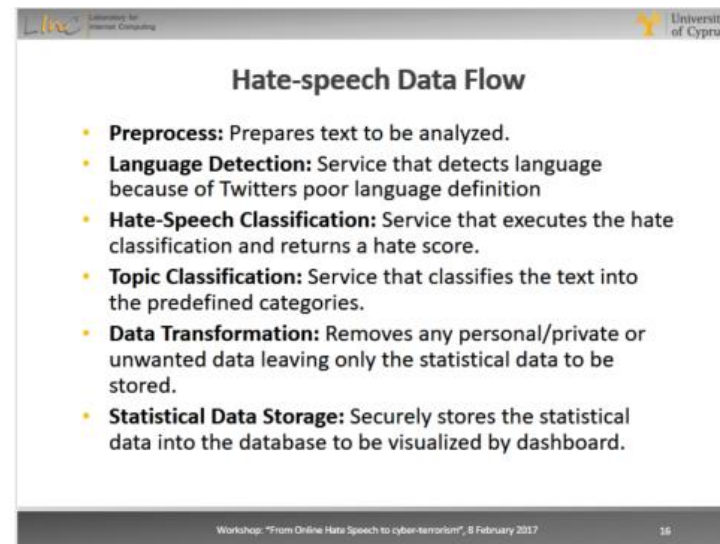
- Input data comes from Apache Kafka are preprocessed to enter the Classifier.
- Multi-lingual Hate-speech Corpus, continuously enriched, supports and retrain the Classifier.
- The Classifier receives the input, and outputs the Hate-speech analysis to be stored.



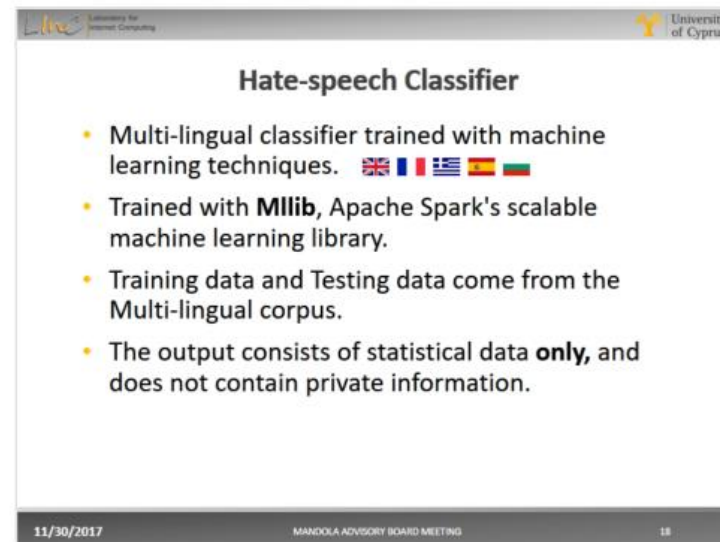
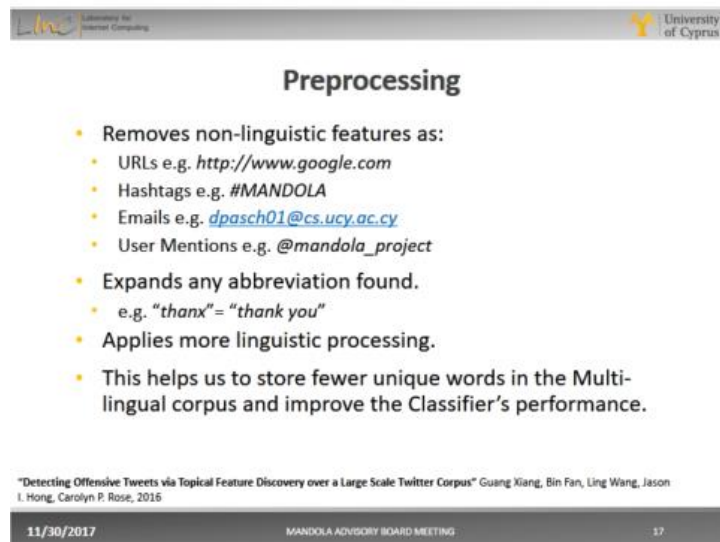
11/30/2017 MANDOLA ADVISORY BOARD MEETING 14



15



16



Classification Algorithm

- **Class:** **Hate** or **No Hate**
- **Target Labels:** 7 categories of hate speech (ethnicity, nationality, religion, gender, sexual, disability, class)
- **Model**
 - In multiclass-multilabel classification, the goal is to assign one or more labels to each instance in an instance space
 - Each label associates an instance with one of 2 possible classes (hate / no hate)
 - **Stochastic Gradient Descent (SGD)** Classifier supports multi-class classification by combining multiple binary classifiers in a “one versus all” (OVA) scheme.

Offer Deket, Ohad Shamir (Microsoft Research): Multiclass-Multilabel Classification with More Classes than Examples. AISTATS 2010
Bottou L. (Facebook AI Research) Large-Scale Machine Learning with Stochastic Gradient Descent. Proceedings of COMPSTAT'2010.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 19

19

Hate-speech Data Storage

- In the hate-speech database we store only the statistical output of the hate-speech analysis module, without any vulnerable information.
- This is in line with the article 7(1)(2) of the Personal Data Protection (Protection of the Human) Laws of 2001 to 2012 in Cyprus (N. 138(I)/2001 as it was modified with N. 37(I)/2003 and N105(I)/2012).

Property	Description
hatescore	A number between 0 and 1, representing the hate strength of the tweet or Google page content, with 1 being the most hateful and 0 being the least.
posted_on	The date in UTC format of when the specific data was posted or last updated.
language	The language in which the hate speech is written.
country	The country from which was posted.
city	The city from which was posted.
tags	An array of the hate categories that have been detected in the input. The probable categories are ethnicity, nationality, gender, religion, sexuality, class, disability and history.
geohash	The encoded location of the data that is used in the grouping for the map visualization as well as for the protection of the user personal information.

Workshop: "From Online Hate Speech to cyber-terrorism", 8 February 2017 20

20

Why Stochastic Gradient Descent (SGD)?

- SGD has been successfully applied to large-scale and sparse machine learning problems often encountered in **text classification** and **natural language processing**
- Given that the data is sparse, the classifiers in this module easily scale to problems with more than 10^5 training examples and more than 10^5 features
- The advantages of Stochastic Gradient Descent are **efficiency** and **ease of implementation**.
- **Evaluation Methodology:** conduct 10-fold cross-validation on the dataset.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 21

11/30/2017

MANDOLA ADVISORY BOARD MEETING

21

Multi-lingual Corpus 1/2

- Comprises **multi-lingual sets** of hate-related speech in the different hate categories.
- It is produced from the **Filtering Mechanism** and by having **Social Scientists** manually annotate text.
- The system will continuously enrich this corpus so to be up to date.

```

graph LR
    SD[Sample Datasets] --> HF[Hate Filtering Mechanism]
    HF --> SS[Social Scientists]
    SS --> MC[Multi-lingual Corpus]
    MC --> ES[Enrichment from System]
    ES --> HF
  
```

11/30/2017 MANDOLA ADVISORY BOARD MEETING 22

11/30/2017

MANDOLA ADVISORY BOARD MEETING

22

Multi-lingual Corpus 2/2

- **Subjectivity Classifier:** Hate-speech tends to be opinionated, thus subjective sentences are more likely to contain hate-speech than objective.
- **Hate-base Classifier:** Sentiment lexicon for hate specific terms which often express hate against the referent.
- **Polarity Classifier:** Hate-speech tends to have negative meaning, thus sentences with negative polarity are more likely to contain hate-speech.

"Hatebase," Mobocracy, Sentinel Project for Genocide Prevention, 2008
"A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," F. Å. Nielsen 2011.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 23

23

Social Scientists

- Social Scientists annotate if text excerpts presented through a web-based **MANDOLA annotation system** contain hate-speech and the category to which they belong.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 24

24

Social Scientists

- **Social Scientists:**
 - MANDOLA appointed a total of 8 social scientists, 2 for Spanish, 2 for Greek, 2 for Bulgarian, 1 for French and 1 for English.
 - Each one annotated text both in his/her native language and in English.
 - **Annotated as Hate-speech:** 3,212
 - **Annotated as not hate-speech:** 20,203

Language	Content	Hate-speech [Agreed On]	Not Hate-speech [Agreed On]
English	26422	63	23836
Greek	4533	165	3646
Bulgarian	4791	12	1231
French	911	0	0
Spanish	12267	901	10165

11/30/2017 MANDOLA ADVISORY BOARD MEETING 25

11/30/2017

MANDOLA ADVISORY BOARD MEETING

25

Social Scientists Results

- **Problem:**
 - The annotated content does not contain sufficient amount of hate-speech in order to successfully train a model for each language.
- **Solution for English:**
 - Several publicly available annotated datasets from other published works used to expand hate-speech corpus.

No Hate-speech	Hate-speech	Offensive	Racist	Sexist
22747	3783	24006	1957	3217

"Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter," Waseem, Zeerak & Hovy, Dirk, 2016
"Automated Hate Speech Detection and the Problem of Offensive Language", T. Davidson, D. Worrmsley, M. Macy, I. Weber 2017.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 26

11/30/2017

MANDOLA ADVISORY BOARD MEETING

26

Social Scientists Topic Labeling

- Manual Analysis on Topic Labeling
- Many miss-classifications between *Ethnicity – Nationality*, *Sexual – Gender*.
- Very few classifications on *Class* and *Disability* categories.
- *Other* category is really noisy and cannot be inferred.
- *Personal* category is also really noise and cannot be inferred

11/30/2017 MANDOLA ADVISORY BOARD MEETING 27

27

Solutions


- Remove *Class*, *Disability*, *Other* and *Personal* categories.
- Combine *Ethnicity* and *Nationality* into one category *Ethnicity-Nationality*.
- Combine *Sexual* and *Gender* into *Sexual-Gender*.
- Data will be classified in *Other* category, if none of the rest categories exceed their respective threshold t .

11/30/2017 MANDOLA ADVISORY BOARD MEETING 28

28

Topic Modelling

- Topic Modeling is a set of algorithmic techniques that aim to discover and annotate large archives of documents with thematic information.
- LDA – Latent Dirichlet Allocation is the most popular statistical topic modeling algorithm.
- Train using MALLET - Java-based package for statistical natural language processing.

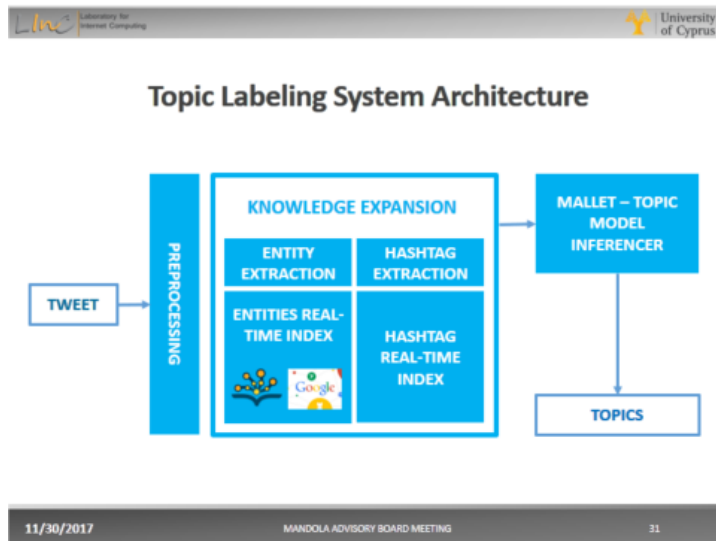


11/30/2017 MANDOLA ADVISORY BOARD MEETING 29

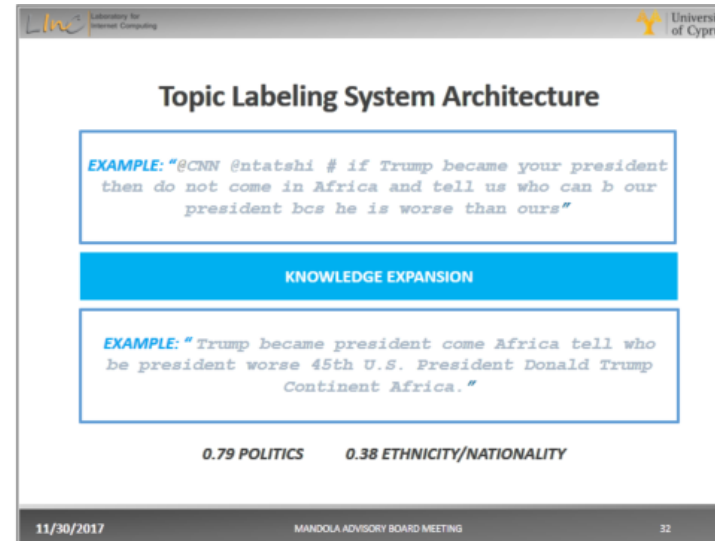
Word Cleaning Algorithm

1. Remove shortened URLs in tweets.
2. Suppress three or more repeated letters into one e.g. *hellooooo* to *hello*.
3. Replace slang words and phrases.
4. Apply Entity Extraction and perform Normalization e.g. *trump* to *Donald Trump*, *45th President of the United States*.
5. Expand any #HASHTAG.
6. Remove @USERNAME.
7. Remove stop-words.
8. Remove any non-letter character other than '-' and "'".

11/30/2017 MANDOLA ADVISORY BOARD MEETING 30



31



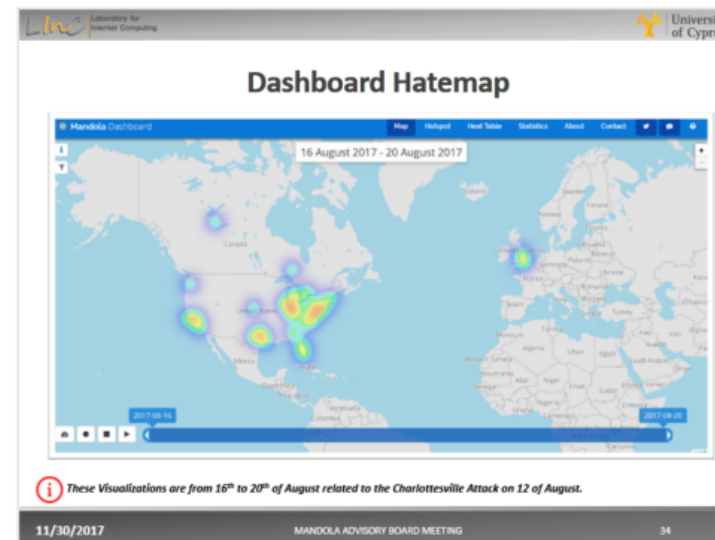
32

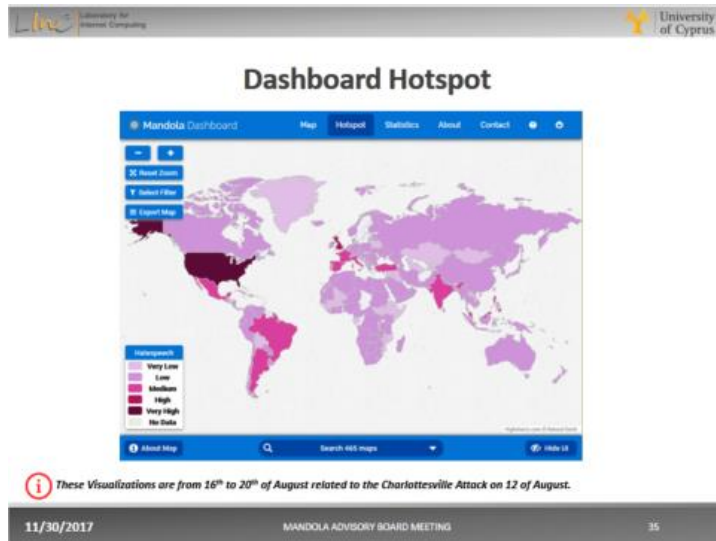
Monitoring Dashboard Web application

- Built with **HTML5**, **CSS3** and **JavaScript**.
- Compatible with any web browser.
- Responsive (**Twitter's Bootstrap**) and compatible with any mobile device.
- Can be found at: <https://goo.gl/CnXttH>
- Password: m@nd0la_2016

HTML JS CSS B

11/30/2017 MANDOLA ADVISORY BOARD MEETING 33

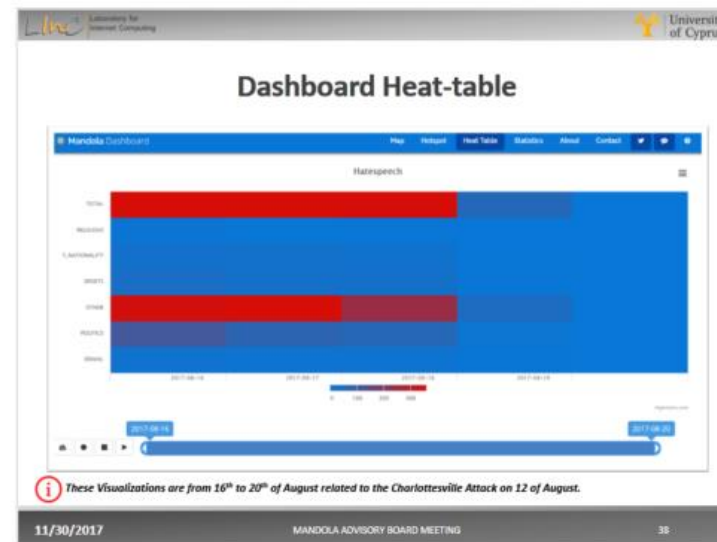


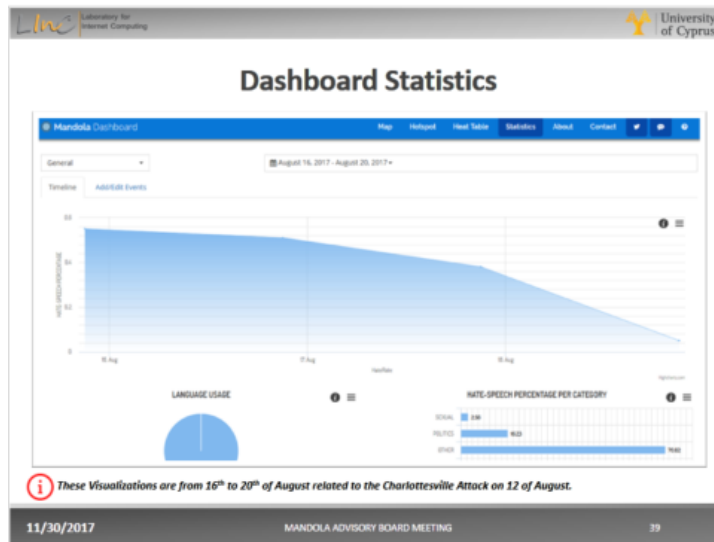


35

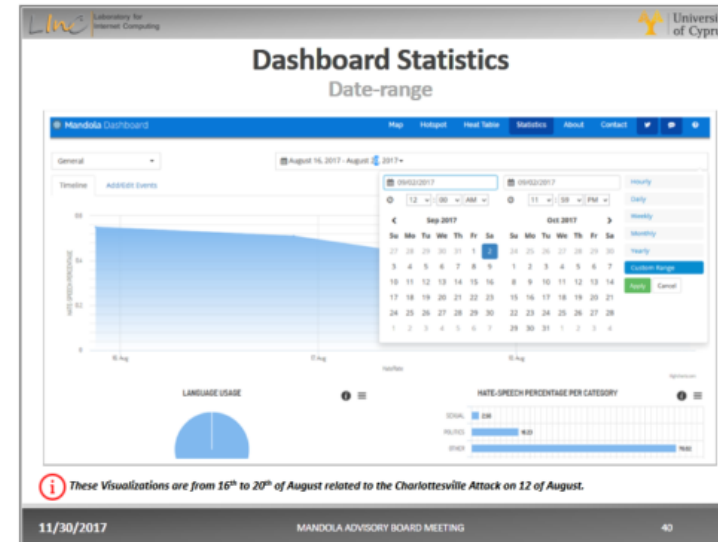


36

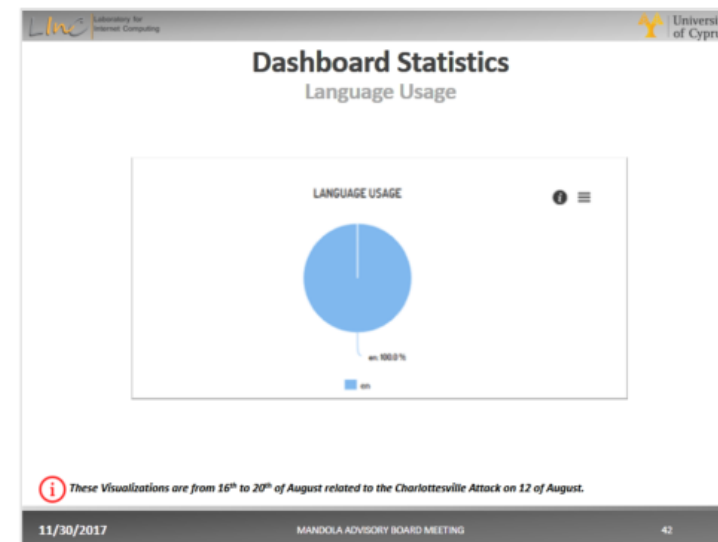
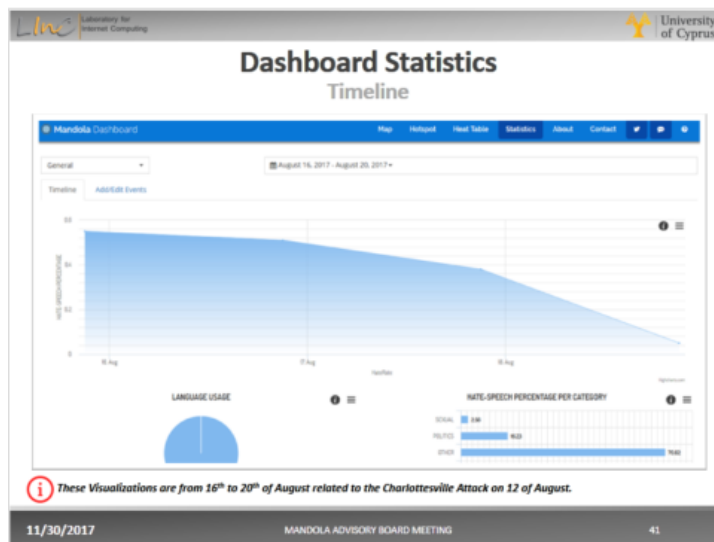


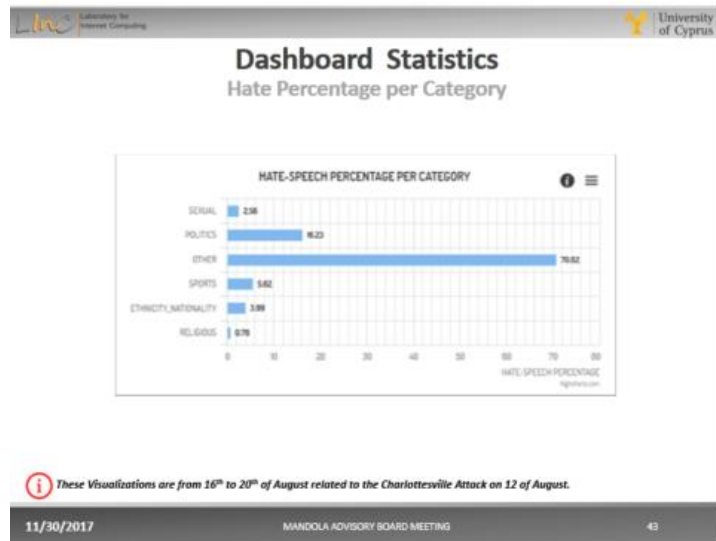


39

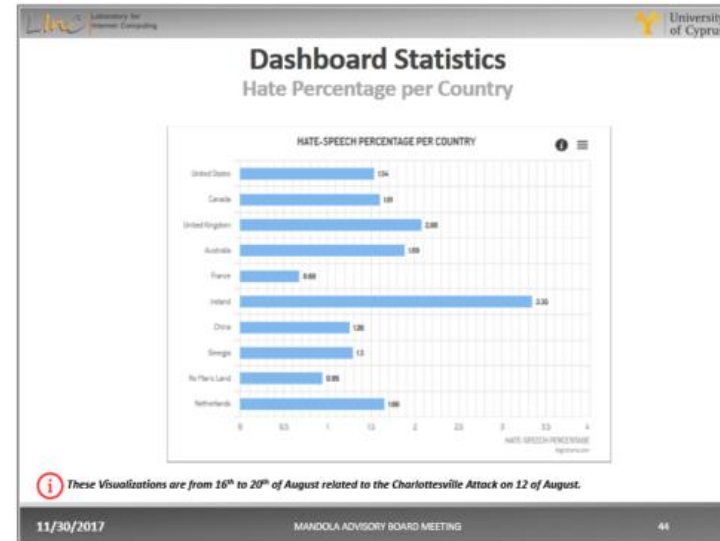


40

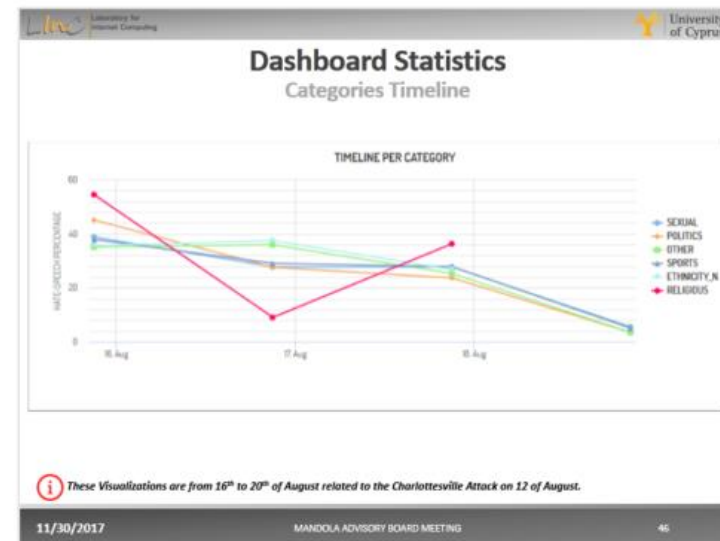
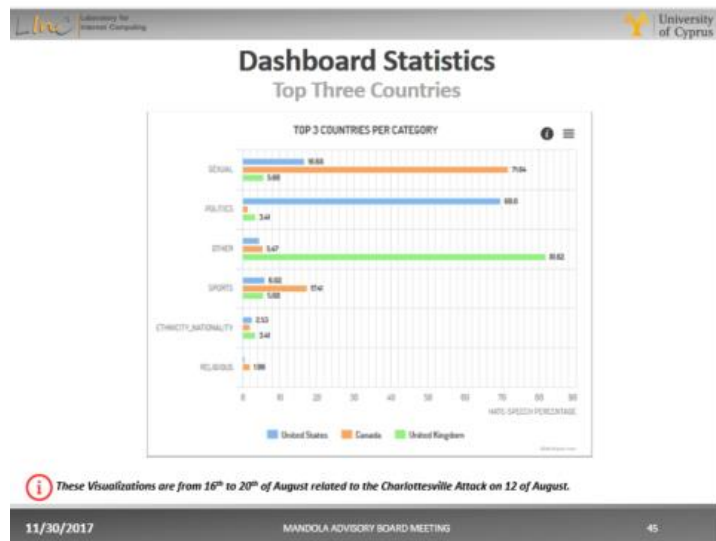


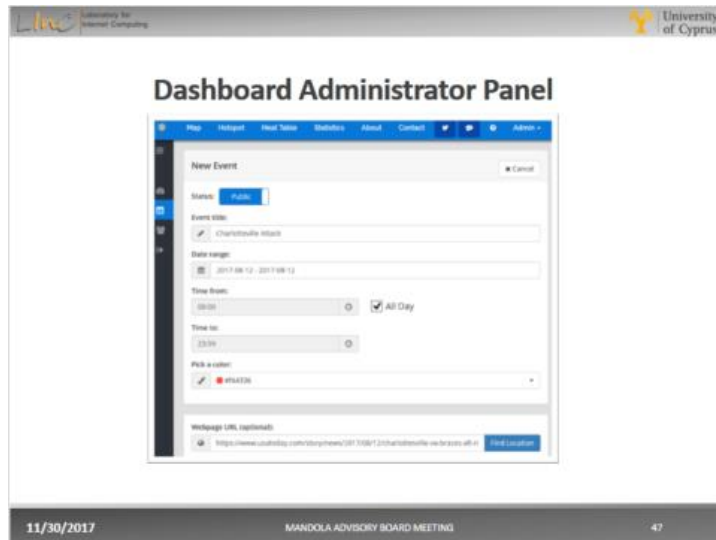


43

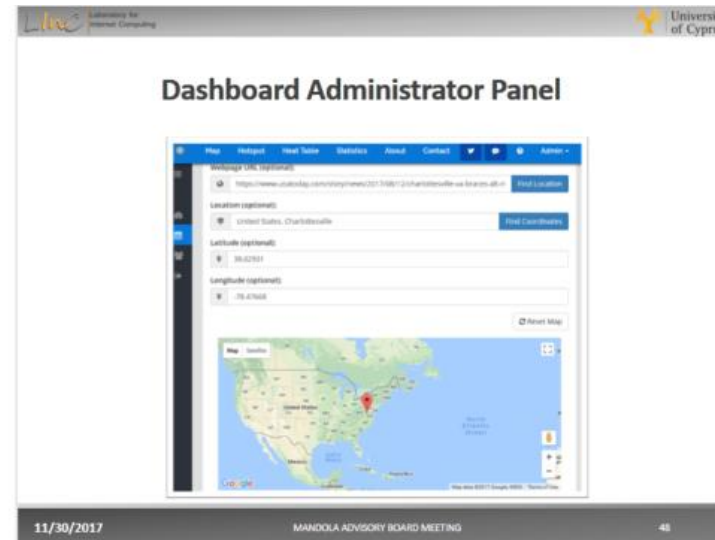


44

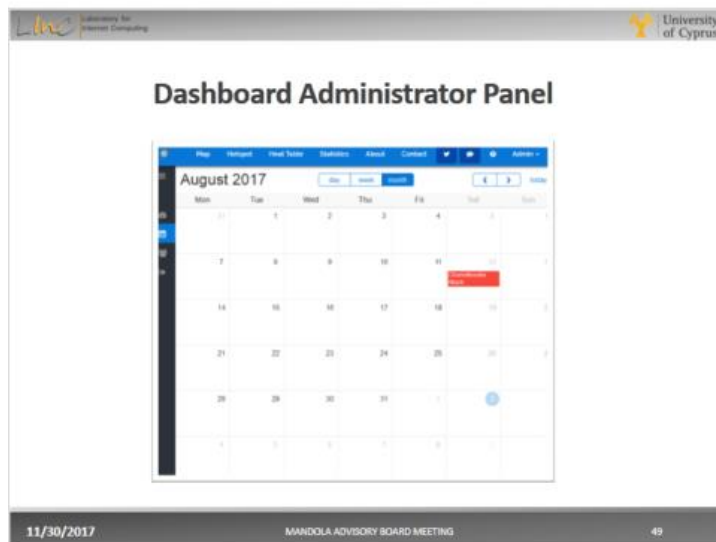




47



48



Mobile Application

- Application's main objectives are:
 - Support anonymous reporting.
 - Raise online hate-speech awareness by viewing statistical analysis and FAQs.
 - Be compatible for Android and IOS devices.
 - Provide the best possible user experience based on each platform's limitations.
- Development technologies:
 - Cordova which enables you to develop a mobile application with web technologies.
 - HTML, CSS and JavaScript.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 50

Mobile Application

The screenshots show the following views:

- Awareness View:** Displays 'HATE-SPEECH ENCOUNTERS' with a bar chart showing a 40% increase in 2016. It also shows 'GLOBAL HATE-SPEECH' with a bar chart showing a 10% increase in 2016.
- FAQs View:** Displays 'FAQs FREQUENTLY ASKED QUESTIONS' with a search bar and a list of questions.
- Report View:** Displays a list of tweets related to hate speech.
- Settings View:** Displays settings for the application, including 'Default OCR language', 'Manage languages for OCR', 'Photograph analysis', 'Keep images only', and 'Enable MANDOLA bubble'.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 51

51

Mobile Application

- Awareness:**
 - Present several facts about hate-speech.
 - Analysis from the dashboard's data. Analysis on data collected from certain events:
 - Olympic games at Rio in August 2016
 - U.S. Presidential Elections 2016

The screenshots show the following views:

- Awareness View:** Displays 'HATE-SPEECH ENCOUNTERS' with a bar chart showing a 40% increase in 2016.
- Extended Hate Speech View:** Displays 'EXTENDED HATE SPEECH' with a bar chart showing a 10% increase in 2016.
- Global Hate Speech View:** Displays 'GLOBAL HATE SPEECH' with a bar chart showing a 10% increase in 2016.
- Important Events View:** Displays 'IMPORTANT EVENTS HATE SPEECH' with a bar chart showing a 10% increase in 2016.
- Settings View:** Displays settings for the application, including 'Default OCR language', 'Manage languages for OCR', 'Photograph analysis', 'Keep images only', and 'Enable MANDOLA bubble'.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 52

52

Mobile Application

- FAQs:**
 - Embedded the FAQs from D4.1.
 - Answers are smaller in order to be more user-friendly.
 - Also added a keyword search mechanism to filter out questions.

The screenshots show the 'FAQs FREQUENTLY ASKED QUESTIONS' view with a search bar and a list of questions.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 53

Mobile Application

- MANDOLA Bubble:**
 - To make the user experience more smooth we implemented the **MANDOLA Bubble**, which is a background process facilitating the automation of the form filling.
 - When the application is running, a floating Bubble appears on the mobile phone's screen.
 - We require from the user to copy and paste the URL and the text from the source.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 54

The Mandola "Bubble"

1 Step sequence to enable the MANDOLA Bubble for the MANDOLA Reporting Application

11/30/2017 MANDOLA ADVISORY BOARD MEETING 55

55

Mobile Application

- **Settings:**
 - Select the default OCR Language
 - Download another OCR Language.
 - Enable or disable storage of cropped images.
 - Enable or disable the **MANDOLA Bubble**.
 - Enable or disable automatic analysis.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 56

56

Mobile Application

- **Reporting:**
 - The user is required to fill the source **URL**, the **hate-speech text** and the **categories** to which it belongs. Optionally can fill a **title** input for more details.
 - The user must be able to report hate-speech encounters from both **public** and **private** sources.
 - Thus we provide two main methods for reporting:
 - Using MANDOLA Proxy server and custom browser to report public hate-speech encounters.
 - Using Optical Character Recognition – OCR to report private hate-speech encounters.
 - The user can see the reports after they are completed and view them.

11/30/2017 MANDOLA ADVISORY BOARD MEETING 57

Mobile Application

- **Reporting:**

Reporting Method

Report Preview

11/30/2017 MANDOLA ADVISORY BOARD MEETING 58

2 Encountering hate-speech publicly in mobile browser and reporting via in-app-browser method

11/30/2017 MANDOLA ADVISORY BOARD MEETING 59

59

3 Encountering hate-speech publicly in twitter native application and reporting via in-app-browser method

11/30/2017 MANDOLA ADVISORY BOARD MEETING 60

60

4 Encountering hate-speech in private in facebook native application and reporting via the screenshot method

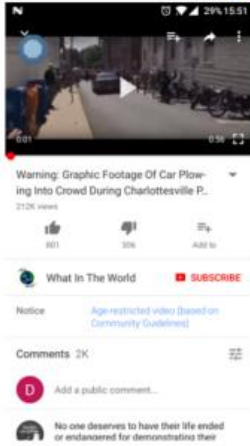
11/30/2017 MANDOLA ADVISORY BOARD MEETING 61

5 Encountering hate-speech in public in youtube native application and reporting via combining the screenshot and url method

11/30/2017 MANDOLA ADVISORY BOARD MEETING 62

Reporting while browsing YouTube Mobile Application

- Enabled the MANDOLA Bubble.
- Found the Charlottesville Protest YouTube video.
- Contains sensitive and graphic footage.




11/30/2017 MANDOLA ADVISORY BOARD MEETING 63

63

Reporting while browsing YouTube Mobile Application

- Scroll the comments and spot a hateful one.
- Contains political and racist hate-speech.
- Also has four up-votes.




11/30/2017 MANDOLA ADVISORY BOARD MEETING 64

64

Reporting while browsing YouTube Mobile Application


- Take a screenshot of the comment.
- The MANDOLA Bubble will receive it.
- The screenshot flag will be shown.



11/30/2017 MANDOLA ADVISORY BOARD MEETING 65

Reporting while browsing YouTube Mobile Application

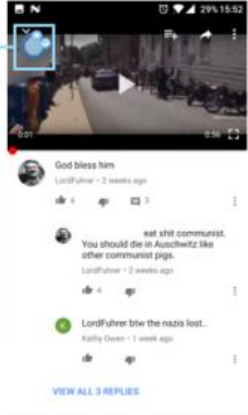
- Can report with only image without any connection to this video.
- Can also copy the link to this video in order to connect the comment.



11/30/2017 MANDOLA ADVISORY BOARD MEETING 66

Reporting while browsing YouTube Mobile Application

- Now the MANDOLA Bubble received the copied link and connected the previous screenshot.
- Just press the MANDOLA Bubble to proceed with the reporting.

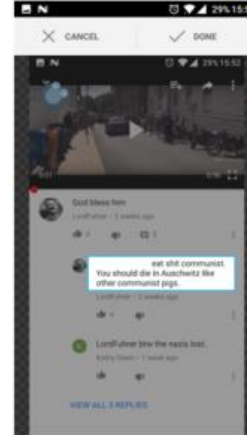


11/30/2017 MANDOLA ADVISORY BOARD MEETING 67

67

Reporting while browsing YouTube Mobile Application

- Place the selection box on-top of the hateful comment.
- Press "DONE" and the OCR – Optical Character Recognition will convert it to message.

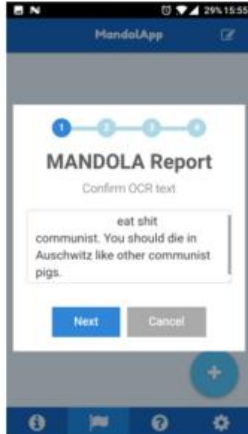


11/30/2017 MANDOLA ADVISORY BOARD MEETING 68

68

Reporting while browsing YouTube Mobile Application

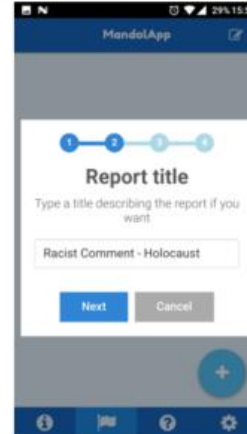
- Now there are four steps.
- First you can correct any spelling mistakes done by the OCR – Optical Character Recognition Module.



11/30/2017 MANDOLA ADVISORY BOARD MEETING 69

Reporting while browsing YouTube Mobile Application


- It is optional to assign a title to the report for either your or the moderators help.



11/30/2017 MANDOLA ADVISORY BOARD MEETING 70

Reporting while browsing YouTube Mobile Application

- The YouTube video URL is automatically placed in the report.
- This is because the MANDOLA Bubble received the copied link.

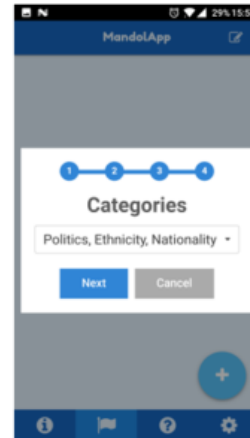


11/30/2017 MANDOLA ADVISORY BOARD MEETING 71

71

Reporting while browsing YouTube Mobile Application

- The last step is to assign hate categories describing the comment.
- For the specific one it is Politics, Ethnicity and Nationality.
- By pressing "Next", the report is sent to the reporting service.

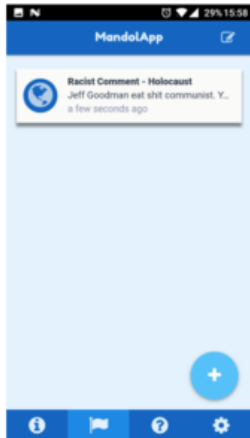


11/30/2017 MANDOLA ADVISORY BOARD MEETING 72

72

Reporting while browsing YouTube Mobile Application

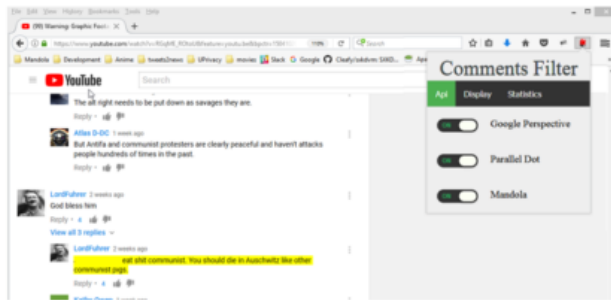
- It is also stored on your device so you are able to view the history of your reports.



11/30/2017 MANDOLA ADVISORY BOARD MEETING 73

Ongoing Development

- Cross Browser plugin for annotating toxic and hateful social commentary.



11/30/2017 Workshop: "From Gendola to MANDOLA: ADVISORY BOARD MEETING", 8 February 2017 74


Laboratory for Internet Computing


University of Cyprus

Acknowledgments




MANDOLA





This work has been produced with the financial support of the Rights, Equality and Citizenship (REC) Programme of the European Union under Grant Agreement JUST/2014/RRAC/AG/HATE/6652

Demetris Paschalides, Dimos Stephanides, Haris Efstathiades, Tatiani Synodinou, Alvaro Ortigosa, Marcos Hernando, George Pallis, Marios Dikaiakos

11/30/2017
Workshop: "From GREEK JOURNALISM TO DIGITAL JOURNALISM", 5 February 2017
75


Laboratory for Internet Computing




University of Cyprus



11/30/2017
Workshop: "From GREEK JOURNALISM TO DIGITAL JOURNALISM", 5 February 2017
76



9 Appendix D: Privacy Impact Assessment of the MANDOLA outcomes

 <p>Privacy Impact Assessment of the MANDOLA outcomes</p> <p><i>Presentation of the PIA and of the Advisory Board consultation results</i></p> <p>Estelle De Marco <i>Inthemis</i></p> <p><small>2nd AB meeting - Brussels – 7 September 2017</small></p> <p><small>markatos@ica.forth.gr www.mandola-project.eu 1</small></p>	 <p>I- Presentation of WS2 works</p> <p>Several deliverables focussing on different aspects</p> <ul style="list-style-type: none">▪ D2.2: Identification and analysis of the legal and ethical framework▪ D2.1a and b: Definition of illegal hate speech and implications▪ D2.4a: PIA – methodology▪ D2.4b: PIA – final report <p><small>markatos@ica.forth.gr www.mandola-project.eu 2</small></p>
--	---




II- Presentation of the PIA

Assessment of

Definition

Understood in a broad sense:

Assessment of risks posed by a project to the right to private life and to personal data protection, and more widely to the other rights and freedoms either exercised by individuals in their respective personal spheres, or restricted by extension because of a privacy limitation or a personal data processing.



Monitoring and Detecting
Online Hate Speech

markatos@ics.forth.gr
www.mandola-project.eu
3

II- Presentation of the PIA

Method

Created on the basis of existing methods, refined in order to ensure an extended protection of rights and freedoms.

- Methods designed in several projects (ePOOLICE, PIAF, VIRTUOSO)
- Guidelines on risk management (ENISA, EBIOS)
- The Article 29 Data Protection Working Party Guidelines on DPIA and opinion on the DPIA template for Smart Grid and Smart Metering Systems developed by the Expert Group 2 of the European Commission
- Article 35 of the GDPR / 26 of the Directive 2016;
- One of the first books on Privacy Impact Assessment edited by David Wright and Paul De Hert;
- Guidelines published by the French Data Protection Authority;
- The UK information Commissioner's Office (ICO) PIA code of practice.




Monitoring and Detecting
Online Hate Speech

markatos@ics.forth.gr
www.mandola-project.eu
4

II- Presentation of the PIA

Steps of the PIA

1. Determining the necessity of a PIA and its scale
2. Determining the assessment team and its objectivity
3. Description of the scope and framework of the study
4. Assessment of the risks to fundamental rights / freedoms
5. Risk treatment
6. Stakeholders consultation
7. Monitoring and review




Monitoring and Detecting
Online Hate Speech

markatos@ics.forth.gr
www.mandola-project.eu
5

II- Presentation of the PIA





MANDOLA outcomes, subject to the PIA

- A monitoring dashboard
- A smartphone app
- A reporting portal
- Information dedicated to policy makers and the Internet Industry
- Information dedicated to Internet users



Monitoring and Detecting
Online Hate Speech

markatos@ics.forth.gr
www.mandola-project.eu
6

<div data-bbox="963 204 1131 279">  MANDOLA Monitoring and Detecting Online Hate Speech </div> <h2 data-bbox="369 236 833 279">II- Presentation of the PIA</h2> <p data-bbox="369 306 1079 343">PIA 1st conclusions (subject to AB consultation)</p> <ul data-bbox="369 354 1131 726" style="list-style-type: none"> ▪ Recommendations resulting from the analysis of legal and ethical requirements, addressing: <ul style="list-style-type: none"> ▪ The MANDOLA partners, ▪ Future developers of the monitoring dashboard & smartphone app ▪ System or data controllers , ▪ LEA, policy makers and States. ▪ Recommendations resulting from the risk treatment analysis, addressing: <ul style="list-style-type: none"> ▪ Same stakeholders + ▪ Future broadcasters of MANDOLA products, ▪ All stakeholders. <div data-bbox="414 769 548 785">markatos@ics.forth.gr</div> <div data-bbox="869 769 1019 785">www.mandola-project.eu</div> <div data-bbox="1102 769 1120 785">7</div>	<div data-bbox="1765 204 1933 279">  MANDOLA Monitoring and Detecting Online Hate Speech </div> <h2 data-bbox="1176 236 1697 279">III- AB consultation - content</h2> <p data-bbox="1176 306 1422 343">Questions asked</p> <ul data-bbox="1176 354 1870 550" style="list-style-type: none"> ▪ General or focussed opinion on the PIA ▪ General or focussed opinion on the recommendations ▪ Request for any other comment on D2.4 and D2.2 <p data-bbox="1176 571 1288 608">Results</p> <ul data-bbox="1176 619 1691 694" style="list-style-type: none"> ▪ Four answers so far ▪ Six (all very valuable) comments <div data-bbox="1220 769 1355 785">markatos@ics.forth.gr</div> <div data-bbox="1675 769 1825 785">www.mandola-project.eu</div> <div data-bbox="1908 769 1926 785">8</div>
<div data-bbox="963 817 1131 892">  MANDOLA Monitoring and Detecting Online Hate Speech </div> <h2 data-bbox="369 849 929 892">III- AB consultation - outcomes</h2> <p data-bbox="369 922 862 959">Comment n°1 (3/4 – 1 remains silent):</p> <p data-bbox="369 962 1097 1026"><i>Most recommendations are fine to very complete work, clearly explained... Thank you!</i></p> <p data-bbox="369 1050 638 1086">Comment n°2 (1/4):</p> <p data-bbox="369 1090 1131 1217"><i>Issues linked with the collection of personal data relating to victims to be also taken into account, esp. in Section 3.1 (Step 1 – Determination of the necessity of a PIA and its scale).</i></p> <p data-bbox="403 1233 515 1270">Answer:</p> <ul data-bbox="403 1276 1131 1369" style="list-style-type: none"> ✓ It will be added; ✓ the whole study will be checked in order to ensure it does not lead to other modifications <div data-bbox="414 1385 548 1401">markatos@ics.forth.gr</div> <div data-bbox="869 1385 1019 1401">www.mandola-project.eu</div> <div data-bbox="1102 1385 1120 1401">9</div>	<div data-bbox="1765 817 1933 892">  MANDOLA Monitoring and Detecting Online Hate Speech </div> <h2 data-bbox="1176 849 1736 892">III- AB consultation - outcomes</h2> <p data-bbox="1176 922 1444 959">Comment n°3 (2/4):</p> <p data-bbox="1176 962 1892 1058"><i>The definition of hate speech that is used / the reason why some offences and not others are included in the definition is not clear</i></p> <p data-bbox="1209 1090 1321 1126">Answer:</p> <ul data-bbox="1209 1133 1892 1356" style="list-style-type: none"> ✓ This definition has been determined in D2.1 (it was actually the subject of D2.1) ✓ Following this comment it will be further clarified in D2.1b and will be included in D2.4b (the PIA) ✓ We will in addition make sure it is clear to people who access MANDOLA outcomes <div data-bbox="1220 1385 1355 1401">markatos@ics.forth.gr</div> <div data-bbox="1675 1385 1825 1401">www.mandola-project.eu</div> <div data-bbox="1908 1385 1926 1401">10</div>

III- AB consultation - outcomes



Overview of the MANDOLA definition of hate speech:

- ✓ In order to compare legislations efficiently, we have firstly searched for all offences and civil or even administrative tort that might be used to sanction online actions that are motivated by hate or at least by the will to offend another person, due to one of his or her particular characteristics.

For ex., has been studied the offence of realisation of a montage of the private images of someone else without his / her consent, if not specified that it is a montage, because it might show a will to mock or particularly offend the victim, whereas the simple publication of private images might pursue several other illegal purposes.

markatos@ics.forth.gr

www.mandola-project.eu

11

III- AB consultation - outcomes



- ✓ This broad definition has been retained because boundaries of hate are very difficult to identify while one of the MANDOLA's aim was to provide for recommendations in relation to the definition of illegal speech. This implied to make a wide mapping of existing provisions
- ✓ Why « hate » is difficult to identify:
 - It is very often the first motive for the commission of an offence
 - criteria such as "origins" or "handicap" are not of help since (1) they vary between States and (2) some States prohibit some actions whatever the specific hate-related motivation
 - This is reinforced by the fact that under several legislations, hatred-related motivations are an aggravating circumstance in relation with all the penal infringements (6 States / 10)

markatos@ics.forth.gr

www.mandola-project.eu

12

III- AB consultation - outcomes



- ✓ From findings, we have been able to identify 4 categories of hatred-related behaviours

- (1) illegal in all or almost all the studied States;
- (2) illegal or partially illegal in a majority of these States;
- (3) illegal in a minority of these States;
- (4) additional behaviours that should be illegal according to European and International instruments.

- ✓ Behaviours have been defined...

1. In their most common definition;
2. where not possible (too wide heterogeneity of legislations), based on the definition provided by European and/or international instruments.
3. Where not possible (found for offences punished in a minority of countries), the retained definition has been the more interesting one in terms of "novelty" compared to other close illegal behaviours already studied.

markatos@ics.forth.gr

www.mandola-project.eu

13

III- AB consultation - outcomes



- ✓ 4 categories of hatred-related behaviours...
 1. Behaviours that are illegal in all or almost all the studied E.U. Member States: (in short)
 - (1) Public incitement to hatred or eventually violence or discrimination on illegal grounds;
 - (2) Making available materials inciting (and eventually promoting) hate and eventually violence or discrimination based on certain grounds through a computer system;
 - (3) (4) Public insult and defamation based on certain victim's characteristics;
 - (5) Threatening a natural person with the commission of a serious offence, eventually motivated by racism and xenophobia;

markatos@ics.forth.gr

www.mandola-project.eu

14

III- AB consultation - outcomes



2. **Behaviours that are illegal or partially illegal in a majority of the studied E.U. Member States:** (in short)
- (1) **Participating / establishing organisations that promote or incite** discrimination, hate/violence based on certain persons' characteristics
 - (2) **Public condoning, denying or grossly trivialising crimes** against peace, crimes of genocide, crimes against humanity and war crimes, eventually subject to conditions relating to the impact of the action or to the perpetrator's motivation.;
 - (3) **Sending of grossly offensive and/or indecent or obscene or menacing content**, mostly for any reason;
 - (4) **Direct public incitement to commit any offence or crime**, for any reason;
 - (5) **Illegal motivations as aggravating circumstance**;
 - (6) **Blasphemy / Insult to religion**

markatos@ics.forth.gr

www.mandola-project.eu

15

III- AB consultation - outcomes



3. **Behaviours that are illegal in a minority of the studied E.U. Member States:** (in short)
- (1) **Sending a message, or whatever content, which can cause** annoyance, harassment and / or needless anxiety to another person, which the sender knows to be false, for any ground
 - (2) **Promotion or public incitement to hostility or violence** between communities
 - (3) **Recording of images of the commission of a crime or misdemeanour** against a person, for any ground and by any means
 - (4) **Realising a montage with the talk or the images** of a third party without his or her consent, if it is not obvious that it is a montage or if it is not specified that it is a montage, for any ground
 - (5) **To misuse / usurp someone else's identity**, for any ground

markatos@ics.forth.gr

www.mandola-project.eu

16

III- AB consultation - outcomes



4. **Additional behaviours that should be illegal according to European and International instruments:** (in short)
- (1) **Dissemination of ideas based on racial superiority or hatred** against any race or group of persons of another colour or ethnic origin
Covered by only 2 legislations (taken into consideration under the offence of incitement to / broadcast of hatred)
 - (2) **Provision of any assistance to racist activities**, including the financing thereof.

Might in several countries be sanctioned under the prohibition of the complicity / aiding and abetting offences introduced into the domestic law in the field of racist activities, but has not been noticed labelled as above (except regarding the financing of the organisations that promote or incite discrimination).

markatos@ics.forth.gr

www.mandola-project.eu

17

III- AB consultation - outcomes



- ✓ **From this extensive definition of illegal hate speech, we have issued a short definition**
- **Not used during the MANDOLA research** (because too large to tackle illegal hate speech only)
 - **Content:**
 - Incitement, propagation or support to hatred, violence, discrimination, segregation, or hostility; incitement or threat to commit harm or violence or a crime or a misdemeanour; humiliation, offence to dignity, insult, defamation, discrimination or harassment; the action to force or to prevent or to commit threat in order to compel someone to do something against his/her will, committed against a person, a group of person and even a community, on grounds of some of their particular characteristics.
 - The outrage, insult, defamation or blaspheme directed against religion, ideology, the Divine, or the offence of believers' religious feelings.

markatos@ics.forth.gr

www.mandola-project.eu

18

III- AB consultation - outcomes



Comment n°4 (1/4):

Impacts on fundamental rights of the Dashboard results are correctly assessed but safeguards to be brought must be complemented

- *Proportionality of inhabitants AND users must be considered*
- *Countries must not be considered to present a « dangerous » state of hate.. To be reworded*
- *Cultural aspects must be taken into account (hate speech can be culturally trivialised without intent of inciting hate)*
- *Visible clarifications on the way subjectivity and polarity have been assessed is necessary (the use of keywords is a limitative methodological shortcut, hate-speech words can be used for other purposes than hate speech and hate speech can exist through metaphors and words shared by some people only).*

markatos@ics.forth.gr

www.mandola-project.eu

19

III- AB consultation - outcomes



Comment n°4

Answer:

- Some of these recommendations were already done but expressed less clearly or comprehensively; especially, we let to further developers the duty to perform research in order to identify all the ways that enable to reach results' accuracy to the utmost extent
- We will make sure these recommendations are explicitly included in D2.4b, and are either implemented in the prototype or are the subject of recommendations of further development.

markatos@ics.forth.gr

www.mandola-project.eu

20

III- AB consultation - outcomes



Comment n°5 (1/4):

Basic awareness of all the judiciary on cybercrime and electronic evidence to be ensured (including in rel. to the existence of specialised teams)

Answer:

- D2.4b only recommends to favour initial /professional LEA training (inter alia to ensure their knowledge about the possible falsehood of reports, content and digital identities),
- This recommendation will be added.

markatos@ics.forth.gr

www.mandola-project.eu

21

III- AB consultation - outcomes



Comment n°6 (1/4):

Summarised recommendations might be difficult to understand for non-legal persons. One solution could be to make links between these recommendation and their justification in previous sections.


Answer:

We will create these links.

markatos@ics.forth.gr

www.mandola-project.eu

22


Monitoring and Detecting
Online Hate Speech

IV- Discussion

Any other comments?

- *On the method used*
- *On the PIA content (such as the identification of risks)*
- *On Section 4 (recommendations)*
- *On any other issue you would like to raise*

Thank you very much to all for your involvement!

markatou@ics.forth.grwww.mandola-project.eu23



10 Appendix E: A short review of the Landscape analysis and introduction to Mandola Stakeholder Survey

 <p>MANDOLA: Monitoring and Detecting on-line Hate Speech https://www.surveymonkey.com/r/BKQ7VRF</p> <p>Cormac Callanan Leader WS4 - Aconite, Dublin</p>	 <p>WS4 activities</p> <ul style="list-style-type: none">• D4.1 FAQ Book• D4.2 Best Practice Guide• D4.3 Stakeholders Workshop• D4.4 Landscape Analysis• D4.5 Stakeholders Survey<ul style="list-style-type: none">• Of current and future activities• D4.6 Network Liaison Officers Meeting (19th Sep 17)  <p>mandola-contact@mandola-project.eu www.mandola-project.eu 4</p>
--	--



LANDSCAPE DOCUMENT

mandola-contact@mandola-project.eu

www.mandola-project.eu

5

Landscape Document

- Deliverable 4 from WS4 for the Mandola project.
- Focuses on the ongoing initiatives
- Current activities in Europe
- Brief Gap Analysis

mandola-contact@mandola-project.eu

www.mandola-project.eu

6

- Many countries are already supported by hotlines taking reports about hate speech
 - INHOPE network
 - INACH network
- Some countries do not have a structured response to hate speech or a method to process complaints or reports about hate speech.
 - Example: some members of INHOPE network are from area of children's rights and have no mandate to respond to hate speech except as it affects children.

mandola-contact@mandola-project.eu

www.mandola-project.eu

7

Objective

- Highlight best practice in this field
- Determine areas which need focus
- Several forms of hate speech are illegal in the European Union (EU), but not all Member States punish the exact same behaviours (Mandola Legal WS2)

mandola-contact@mandola-project.eu

www.mandola-project.eu

8

Countries



- Bulgaria
- France
- Greece
- Ireland
- Spain

mandola-contact@mandola-project.eu

www.mandola-project.eu

9

Bulgaria



- Hate speech has become part of the curriculum of almost every Bulgarian institution.
- All have demonstrated clear position on policy of "zero tolerance for hate speech"
- In 2015 Council for Electronic Media and Bulgarian Central Election Commission initiated the **MoU for non-use of hate speech during the municipal elections campaign**
- Agreement was signed by political parties and parliamentarians, media representatives and NGOs.

mandola-contact@mandola-project.eu

www.mandola-project.eu

10

France



- In 2012, government created an inter-ministerial delegation to combat racism and antisemitism,
 - Extended in 2016 to focus on LGBT hate (DILCRAH)
- End of 2014 xenophobic acts increasing
- Government launched action plan against racism and antisemitism 2015-2017
- In 2016 a specific mobilisation plan against hate and discriminations targeting LGBT people

mandola-contact@mandola-project.eu

www.mandola-project.eu

11

Greece-1



- Greece has always had great interest and sensitivity for the prevention of discrimination
- Albanian migrants were the first victims of racist attacks 30 years ago in Greece. Migrants and refugees from Middle East, Pakistan, Afghanistan increased the racist behaviour by citizens.
- Recent wave of refugees and immigrants in last 2 years from Syria motivated the Government and national LEA to take initiatives against hate crime and hate speech prevention.

mandola-contact@mandola-project.eu

www.mandola-project.eu

12

Greece-2



- In 2014, Ministry of Justice, Transparency and Human Rights **completed transposition** of Council Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia into criminal law.
- Greek NGOs, institutes and research centres have participated in many research projects for combating hate speech and hate crime

mandola-contact@mandola-project.eu

www.mandola-project.eu

13

Ireland



- Irish national broadsheets have shown increasing reporting of 'hate speech' over the past 5 years
- Irish **tabloid** and **regional** papers have not shown much interest in 'hate speech'.
- Interest shown by broadsheets is on **international events** and has **strong emphasis on events relating to the social media companies** (with bases in Ireland.)
- Online bullying, a related concept to 'hate speech' is reported frequently as a matter of Irish interest.

mandola-contact@mandola-project.eu

www.mandola-project.eu

14

Spain



- Racism on the Internet is "alarmingly increasing"
(European Commission Against Racism and Intolerance, 2011, p. 2231; via Ben-Devid & Matamoros-Fernandez, 2016, page 1168).
- Yearly increase in research projects regarding hate speech and in initiatives from Spanish Government and LEA on this phenomenon.
- Examples:
 - Legislative changes in 2015
 - Creation and implementation of a Police Protocol to respond to hate crime incidents
 - Improvements of the data gathering system.
- Spanish prosecutors and Ministry of Interior are active agents assessing the actual impact and penetration of hate speech in society, by promoting initiatives and formative actions.

mandola-contact@mandola-project.eu

www.mandola-project.eu

15

Spain Statistics



MOTIVATION	Nº REPORTED INCIDENTS
Antisemitism	4
Aporophobia	1
Religion	7
Disability	14
Sexuality	15
Racism/Xenophobia	16
Ideology	56
Gender	4
Total	117

CATEGORIZATION	Nº REPORTED INCIDENTS
Slander	37
Threats	29
Threats to religious groups	6
Vexation	6
Degrading Treatment	5
Others	34
Total	117

mandola-contact@mandola-project.eu

www.mandola-project.eu

16

KEY THEMES

mandola-contact@mandola-project.eu

www.mandola-project.eu

17

Countries

- Bulgaria
- France
- Greece
- Ireland
- Spain

mandola-contact@mandola-project.eu

www.mandola-project.eu

18

Bulgaria

- Under-reporting of hate crime incidents and especially of hate crime online
- Due to the low level of public understanding of human rights prevents reliable statistics of hate speech online cases.
- Although hate speech spread is obvious to all and all politicians acknowledge this it cannot be supported with concrete data.

mandola-contact@mandola-project.eu

www.mandola-project.eu

19

France-1

- Racist and anti-Semitic acts have decreased in France in 2016 (-44,69% - 1125; 2034 in 2015) (French Minister of Interiors)
- This is attributed to the 2015-2017 Action plan against racism and antisemitism 2015-2017 and the coordinated action of the government and of the inter-ministerial delegation for the combat against racism, antisemitism and LGBT hate (DILCRAH).
- 182 acts against Muslims were identified (429 in 2015)
- 335 anti-Semitic acts (808 in 2015).
- 608 acts were racist acts not targeting Muslims or Jews (797 in 2015)

mandola-contact@mandola-project.eu

www.mandola-project.eu

20

France-2



- Increased tolerance for all community groups
 - After decrease stopped in 2014. (French National Consultative Commission of Human Rights (CNCDH))
- Tolerance is less linked to facts (such as terrorist attacks) than to the context and to the way politicians and media talk about immigration and diversity.
- Liability of politicians and media is of particular importance
- Commission *"is convinced that the fight against racism lies before everything on the deconstruction of prejudices and preconceived ideas"*

mandola-contact@mandola-project.eu

www.mandola-project.eu

21

Greece



- Official statistical data for hate crime or hate speech were not found
- Very usual to encounter hate speech offline and online people are not keen on reporting those incidents.
- Often victims do not realize they are victims of racist behaviour and do not know ways of supporting their human rights.
- Do not report the incidents to the police
 - feel that they are not going to receive real protection from them
 - cannot afford the costs of litigation.
- Lots of hate crime or hate speech incidents remain underreporting.

mandola-contact@mandola-project.eu

www.mandola-project.eu

22

Ireland



- Dublin is home to significant European headquarters and offices for social media companies including Facebook, Twitter, Google, LinkedIn, Microsoft.
- Presence is reflected in the number of 'hate speech' articles in the broadsheets related to these companies over the 12 months to March 1st 2017

mandola-contact@mandola-project.eu

www.mandola-project.eu

23

Recurring focus on Social Media companies in Irish Broadsheets



Table 4: Articles on Social Media Companies in Irish broadsheets

Paper	Facebook	Twitter	Google	Microsoft
Irish Times	17	5	10	3
Independent	10	3	2	1
Examiner	9	5	5	1

mandola-contact@mandola-project.eu

www.mandola-project.eu

24

Ireland



- Two other characters feature prominently with respect to hate speech in the broadsheets...

- ...Geert Wilders and Donald Trump.

mandola-contact@mandola-project.eu

www.mandola-project.eu

25

Ireland



Categories within Hate Speech

MANDOLA divides "hate speech" into 10 categories. Irish media can be examined in relation to reporting on these categories.

Table 5: Number of articles in Irish broadsheet media on "hate speech" with reference to MANDOLA categories for the 12 month period ended 1.03.17

Category	Irish Times	Independent	Examiner
Ethnicity	2	3	1
Nationality	1	2	1
Sexual	1	4	1
Gender	1	1	
Politics	15	9	6
Sport			
Religious	6	1	
Disability			
Personal	1		

mandola-contact@mandola-project.eu

www.mandola-project.eu

26

Spain



- Number of incidents of hate speech unknown
- Civil society plays a crucial role in the fight against hate speech,
- it has access and information from victims that do not report to authorities for various reasons
- do not feel that reluctant to seek the assistance and support of NGOs.
- Policy makers and LEA could benefit from increasing and improving the communication with these organisations, and by listening to the recommendations that these organisations can provide them with.

mandola-contact@mandola-project.eu

www.mandola-project.eu

27

GAPS



mandola-contact@mandola-project.eu

www.mandola-project.eu

28

Countries

- Bulgaria
- Cyprus
- France
- Greece
- Ireland
- Spain

mandola-contact@mandola-project.eu

www.mandola-project.eu

29

Bulgaria-1

- Need a common approach uniting all efforts from stakeholders.
- Interaction among various public institutions has improved
 - by establishing inter institutional working groups / initiatives
- Public institutions are **not** fully aware of what civil society and academia are doing
- Number of online hate speech criminal cases is very low because:
 - Misconception within general public on what is legal and what is illegal hate speech - which leads to underreporting.
 - Lack of knowledge on what hate speech crime is and how to investigate cases among LEA - especially Regional LEA.

mandola-contact@mandola-project.eu

www.mandola-project.eu

30

Bulgaria-2

- Civil society and academia are deeply engaged in the problem
- Most projects are based on implementing campaigns
 - Not on analysing the governmental policy, legal framework, investigation and procedure of hate crime.
- Awareness campaigns are effective for improving public understanding of the problem
- Needs to be supported by analytical and research projects in order to reach the decision makers, public institutions, the judiciary and law enforcement.
- In terms of NGO and academia projects and initiatives - still little attention is given to intolerance and discrimination based on sexual orientation, gender, age, health (illnesses), disability, political beliefs.

mandola-contact@mandola-project.eu

www.mandola-project.eu

31

Cyprus-1

- Legislation penalising 'hate speech' on the grounds of sexual orientation or gender identity applies lower fines and punishment than other offences based in racism
- NGOs in Cyprus do not have an active role in the development of State policies and little action arises from their recommendations.
- LGBT organisations have not been treated as important stakeholders in shaping Human Rights issues in particular with regard to sexual orientation and gender identity, since they were **not** invited in any formal consultations by State authorities.
- No special guidance issued to public officials or state representatives on hate speech and discrimination on the grounds of sexual orientation or gender identity.

mandola-contact@mandola-project.eu

www.mandola-project.eu

32

Cyprus-2



- Incidents of hate speech or discrimination by the police in the exercise of their duties still occur
 - Although special guidelines have been issued since 2013 for combating and tackling racist violence, xenophobia and discrimination by the police.
- 2011 Law on Combating Certain Forms and Expressions of Racism and Xenophobia by means of Criminal Law has not yet been applied in any case
- No known conviction where the court took into account homophobic, racist or xenophobic motivation during sentencing.
- Apart from Law 26 (III) 2004 which has implemented the Additional Protocol 189, there is no other special regulatory framework or code of conduct about online hate speech
- Lack of criminalisation of public expression, which expresses an ideology which claims national or ethnic superiority.

mandola-contact@mandola-project.eu

www.mandola-project.eu

33

Cyprus-3



- Under-reporting has been recognised by institutions as a major issue in the realm of both hate speech and hate crime.
- No statistics maintained on the number of cases related to discrimination brought to justice.
- No estimates of the number of discrimination cases brought to justice in any journals or textbooks.

mandola-contact@mandola-project.eu

www.mandola-project.eu

34

France-1



- Difficult to identify gaps since issue is subjective.
- Part of a wider debate relating to the most efficient means to combat hate...
 - while strictly respecting fundamental freedoms and rights in a State governed by the rule of law.
- "opinion on the combat against hate speech on the Internet" has been issued on 12 February 2015 by the French National Consultative Commission of Human Rights

mandola-contact@mandola-project.eu

www.mandola-project.eu

35

France-2



- CNCDH considers that the increase of online hate speeches, which is fed by "social tensions and the citizenship' crisis", "challenges the efficiency of policies and of allocated means, and more generally, the efficiency of existing legal mechanisms, in particular of the repressive arsenal".
- CNCDH believes situation requires a review to identify new control strategies

mandola-contact@mandola-project.eu

www.mandola-project.eu

36

Recommendations of the CNCDH

- To affirm the **digital sovereignty of the State**;
- To **reinforce existing mechanisms** in the area of the combat against hate speech on the Internet;
- To **set up a reactive and innovative institution for web regulation**, which could especially lead to diversify answers brought to online hate speech
 - Noting that *"the involvement of a judge is necessary in order to order and to control the removal of an illicit content and the blocking of an Internet site, where these measures constitute severe interferences with the freedom of expression and to communicate"*.
- To **adopt a national action plan on education and digital citizenship**.

Greece


- Most important gap with regard to racist behaviour is the **lack of education**.
- **Society is not well educated** to respect the equality, the human rights of others and to fight several forms of intolerance related to ethnicity, gender, sexuality, political views, religion, etc.
- Many people are **not aware of reporting tools and mechanisms where victims can request support**
- **Reporting systems and clear avenues to prosecution** could undoubtedly empower the confidence between the society and the LEAs for the limitation of hate speech incidents.

Ireland

- One exception to 'internationalism' of hate speech is in the related concept on '**online bullying**' which almost always refers to bullying between teenagers in Ireland
- Ireland has a high proportion of its population in this age bracket and Irish newspapers devote much space to educational matters
- There have been a number of teenage suicides linked to the phenomenon of online bullying.
- While all draw on international experience and the involvement of international social media companies they very strongly relate this issue to Irish teenagers.

Spain


- Some gaps in Spain between different institutions and parts of society regarding hate speech.
- More collaboration and communication between different organisations and institutions is needed, both at national and international levels.
- Number of online hate speech remains unknown
- this poses several problems in the correct detection and analysis of the phenomenon.
 - Victims lack confidence on the procedures regarding the investigation of hate speech incidents
 - Victims tend to underestimate the importance of reporting those incidents to authorities.



<https://www.surveymonkey.com/r/BKQ7VRF>

STAKEHOLDER SURVEY


mandola-contact@mandola-project.eu www.mandola-project.eu 41



Stakeholders Survey

- 29 questions
- <https://www.surveymonkey.com/r/BKQ7VRF>
- (Survey also available in Spanish)

mandola-contact@mandola-project.eu www.mandola-project.eu 42



Stakeholder Survey

Thank you for your interest.

Si desea contestar a esta encuesta en español, pinche aquí
<https://es.surveymonkey.com/r/VJNDZ9N>


The Mandola project focusses on different stakeholders including witnesses of on-line hate speech incidents, policy makers and citizens who are victims or perpetrators of online hate speech.

Witnesses have the possibility to report hate speech anonymously. Policy makers can use up-to-date on-line hate speech information that can be used to create adequate policy in the field. Member States citizens can gain a better understanding of what on-line hate speech is.

All stakeholders should be able to recognize legal and illegal on-line hate-speech and should know what to do when they encounter illegal on-line hate.

The MANDOLA project addresses the two major difficulties in dealing with on-line hate speech

mandola-contact@mandola-project.eu www.mandola-project.eu 43



Which Stakeholders?

1 Indicate the type(s) of activity that best classifies your organisation

<input type="radio"/> Academic	<input type="radio"/> Law Enforcement
<input type="radio"/> Individual - No affiliation	<input type="radio"/> Legal Expert
<input type="radio"/> Industry Other	<input type="radio"/> NGO
<input type="radio"/> Internet Access Provider	<input type="radio"/> Social Media Services
<input type="radio"/> Internet Hosting Provider	
<input type="radio"/> Other (please specify)	

mandola-contact@mandola-project.eu www.mandola-project.eu 44

Stakeholders Survey

MANDOLA
Monitoring and Detecting
Online Hate Speech

11 Which of the following are the most common motives for hate speech?

	Often	Sometimes	Rarely	Never
Age	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Class	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Elderly People	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ethnicity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Foreigners	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gender	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Immigrants	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nationality/National Minority	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
People with disabilities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Politics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Refugees	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Religion	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sexuality	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Social	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Women	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other (please specify):

mandola-contact@mandola-project.eu

www.mandola-project.eu

45

MANDOLA
Monitoring and Detecting
Online Hate Speech

16 How did you or would you respond to online hate speech?

- ☐ Block user on personal block list
- ☐ Ignore it
- ☐ Remove content
- ☐ Report by scoring system
- ☐ Report to Content Moderator
- ☐ Report to independent hotline
- ☐ Report to legal authorities
- ☐ Report user to hosting organisation
- ☐ Respond with Counter-speech
- ☐ Other (please specify)

mandola-contact@mandola-project.eu

www.mandola-project.eu

46

MANDOLA
Monitoring and Detecting
Online Hate Speech

ROLE OF ADVISORY BOARD

mandola-contact@mandola-project.eu

www.mandola-project.eu

47

MANDOLA
Monitoring and Detecting
Online Hate Speech

Why are we here today?

- We want your **advice**
 - **Advisory Board**
- You know a lot about this area
 - Can you share some of your **knowledge**?
 - Can you share some of your **experience**?
 - Some of your **wisdom**?



mandola-contact@mandola-project.eu

www.mandola-project.eu

48

MANDOLA
Monitoring and Detecting
Online Hate Speech

How are we going to get this advice?

- Interactive Brain storming sessions
- Post it notes
- Share your advice
 - Even half-baked ideas
 - All ideas!
 - In research all ideas are welcome



mandola-contact@mandola-project.eu www.mandola-project.eu 49


MANDOLA
Monitoring and Detecting
Online Hate Speech

Brainstorming Query I

- What did you learn from the work of Mandola?

3

List up to three ideas




mandola-contact@mandola-project.eu www.mandola-project.eu 50

MANDOLA
Monitoring and Detecting
Online Hate Speech

Brainstorming Query II

- List **1** (or **2**) significant (positive or negative) changes during the time (Oct15-Sep17) of the Mandola project in each of the following areas?
 - Legislation
 - Enforcement
 - Internet industry
 - Victims & Perpetrators



mandola-contact@mandola-project.eu www.mandola-project.eu 51


MANDOLA
Monitoring and Detecting
Online Hate Speech

Brainstorming Query III

- What are the strengths and weaknesses of COUNTER SPEECH strategies?

4

List up to four ideas.



mandola-contact@mandola-project.eu www.mandola-project.eu 52



Thank you

Cormac Callanan
Leader WS4 - Aconite

Thank you!



mandola-contact@mandola-project.eu

www.mandola-project.eu

54



11 Appendix F: Brainstorming Panel / Question 1

MANDOLA
Monitoring and Detecting
Online Hate Speech


Brainstorming Query 1

- What did you learn from the work of Mandola?

3

List up to three ideas

mandola-contact@mandola-project.eu www.mandola-project.eu



Q1-1

① - Cooperation between various disciplines - aspects private and academic sector - is extremely useful

② - Difficult to strive a balance between detection of hate speech and freedom of speech →

43

Q1-2

Q. I

1. It is difficult to measure Hate Speech
2. It is difficult to define Hate Speech
3. It is difficult to counter Hate Speech

Q1-3

i. - Definition will also remain difficult after Mandola

- Mandola offers a platform to act in practice to combat

- Data collection need to be translated into an evolutionary picture - Mandola contributes greatly

44

Q1-4

There is no easy method to identify hate speech

Q1-5

Q1 - Complexity of legal definition

- Possibility to develop innovative apps
- Importance of regular review by people - automation difficult

Q1-6

- Complexity ①

- Variety of stakeholders

→ More work to do

Q1-7

- Is law the answer?

- Interdisciplinary needed

- but each one solves in their own field (only?)

Q1-8


- ① Issue of policy confidentiality
- ② Quick screenshot button to report Hate Speech for citizens.
- ③ Mixing type of hatred issue (Ethnicity, nationality, sexual → gender)

12 Appendix G: Brainstorming Panel / Question 2

Brainstorming Query II

MANDOLA
Monitoring and Detecting
Online Hate Speech

- List 1 (or 2) significant (positive or negative) changes during the time (Oct15-Sep17) of the Mandola project in each of the following areas?
 - Legislation
 - Enforcement
 - Internet industry
 - Victims & Perpetrators



mandola-contact@mandola-project.eu

② ^{Legislation} Net's DG
(negative)
^{Internet Industry} Code of Conduct
2nd monitoring period
really (positive)

Q2-1

47

Q:2
There is much more awareness about Hate Speech.

Q2-2

Q2
Need stronger public concern on illegal content and need to discuss positive measures

Q2-3

48

- rise of anti-migrant hate speech in Europe
+ tendency to extend grounds of hate speech
e.g., anti protection ^{rights} homophobic speech in a number of European countries

Q2-4

- Lots of progress went on to make users report
- Code of conduct for Internet Industry
Q2-5
- Enforcement proven Not to be done

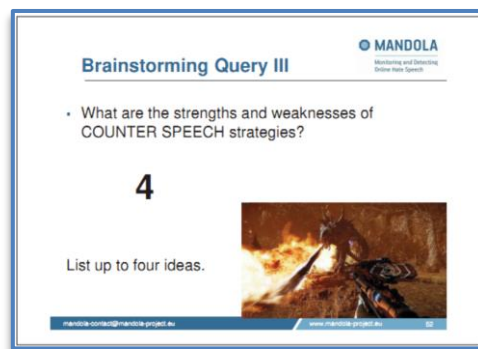
② TRUMP

Q2-6

②
Negative: Victims and Perpetrators
Positive: Enforcement (restricted)

Q2-7

13 Appendix H: Brainstorming Panel / Question 3



3/ Weaknesses

1. Not enough of it
2. Hard to measure success

Q3-1

Strengths

1. Potential to change hearts + minds
2. More effective than delete

Q3

Weakness: it is defense, not offense

Strength: Response of a community

Q3-2

Counter speech

way to

- make active & responsible citizens
- Hard to know what to counter exactly
- to not let Hate Speech not respond. Remove it's went make people think differently

Q3-3

Lauph :)

(Counter-act hate speech with humor and statistics)

Q3-4

51

52

3/ Counter speech. **Q3-5**

- + Probably the best & most effective way of combatting h.s.
- May be difficult to mobilize in countries where problem awareness/educ is low

III

Strengths: seems to be working more than other approaches

Weakness: may lead to confrontation, flame wars. Not easy to implement

Q3-6

Q3

- + Crowdsourcing approaches
- difficult to implement
- + It has to come from people you trust

Q3-7

③ **Q3-8**

- Difficult to use the proper language/argumentation
- Uncertain if it reaches the right public/audience
- Legitimizing hatespeech

No legal enforcement in severe cases.

Q3-9

Q3-10

- Education to respect of others / others' right is fundamental in a multi-cultural society
- people can use hate speech words as metaphors without hate speech intent, without paying attention to it.
- Current initiative are not going far enough
- gov & media share a huge part of liability in spreading

Q3-11

- what is
- best of (has read)
- concise
- to be point (to have impact)