

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



## Monitoring and Detecting Online Hate Speech

### **D2.3 - Legal and ethical compliance of the MANDOLA research**

**Abstract:** The MANDOLA consortium ensured the legal and ethical compliance of its research thorough the life of the MANDOLA project. This report presents the actions that have been taken in this regard.

Contractual Date of Delivery	31 December 2016
Actual Date of Delivery	29 September 2017
Deliverable Security Class	Public
Editor	Estelle De Marco
Quality Assurance	Tatiana Synodiou

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
UMO	Principal Contractor	France

## Document Revisions & Quality Assurance

### Internal Reviewer:

Tatiana Synodiou, Associate Professor, Law Department University of Cyprus, Chair of the Ethics Committee of Mandola.

### Revisions

Version	Date	By	Overview
v.2.3.0	20/09/2017	Inthemis (FR) Estelle De Marco as editor with the contribution of Célie Zamora	Preparation of the report including comments from UCY.
v2.3.1	26/09/2017	Inthemis (FR) Estelle De Marco	Slight modifications taking into account comments from Álvaro Ortigosa (UAM) and Ioannis Inglezakis (Aristotle University / FORTH).
v2.3.2	29/09/2017	Inthemis (FR) Estelle De Marco	Final version taking into account comments from Tatiana Synodiou as quality reviewer.



## Table of Contents

<b>DOCUMENT REVISIONS &amp; QUALITY ASSURANCE .....</b>	<b>3</b>
<b>TABLE OF CONTENTS .....</b>	<b>5</b>
<b>1 EXECUTIVE SUMMARY .....</b>	<b>6</b>
<b>2 INTRODUCTION .....</b>	<b>7</b>
2.1 BACKGROUND TO THE MANDOLA PROJECT .....	7
2.1.1 MANDOLA objectives .....	7
2.1.2 MANDOLA activities .....	8
2.2 PURPOSE AND SCOPE OF THE REPORT .....	8
2.3 DOCUMENT STRUCTURE .....	9
<b>3 COMPLIANCE OF THE MANDOLA RESEARCH WITH THE EUROPEAN CONVENTION ON HUMAN RIGHTS REQUIREMENTS .....</b>	<b>10</b>
3.1 LEGAL BASIS .....	10
3.2 LEGITIMATE PURPOSE .....	10
3.3 NECESSITY .....	10
3.4 PROPORTIONALITY .....	13
3.4.1 <i>Strict necessity of the interference</i> .....	13
3.4.1.1 Strict necessity in relation to the context of the research .....	14
3.4.1.2 Strict necessity in relation to the scope of interferences .....	14
3.4.1.3 Strictly necessary in relation to the nature of interferences .....	16
3.4.1.4 Limited by appropriate safeguards (summary) .....	16
<b>4 COMPLIANCE OF THE MANDOLA RESEARCH WITH PERSONAL DATA PROTECTION REQUIREMENTS .....</b>	<b>18</b>
4.1.1 <i>Legal basis</i> .....	18
4.1.2 <i>Legitimate, explicit and specified purpose</i> .....	18
4.1.3 <i>Data quality</i> .....	21
4.1.4 <i>Data minimisation</i> .....	21
4.1.5 <i>Time limitation</i> .....	22
4.1.6 <i>Appropriate legal ground</i> .....	22
4.1.7 <i>Data subject information</i> .....	23
4.1.8 <i>Data subjects' rights of access, communication, rectification and erasure</i> .....	23
4.1.9 <i>Prohibition of automated decisions</i> .....	23
4.1.10 <i>Enhanced protection of some sensitive data</i> .....	24
4.1.11 <i>Security and confidentiality of the processing</i> .....	24
4.1.12 <i>Data protection authority supervision</i> .....	24
4.1.13 <i>Liability and accountability of the data controller</i> .....	25
4.1.14 <i>Adequate level of protection in some case of data transfers</i> .....	25
<b>5 CONCLUSION .....</b>	<b>26</b>

## 1 Executive summary

The MANDOLA research did not intend to collect personal data. However, data have been collected from the Internet and one of the focuses of the research was to provide information to Internet users, policy makers and the Internet industry. As a result, the MANDOLA research was likely to limit or threaten some fundamental rights and freedoms, including the right to privacy, the right to personal data protection, and the right to freedom of expression.

In this context, the MANDOLA consortium has taken several initiatives aiming to make the MANDOLA research compliant with the European Convention on Human Rights and relevant data protection legislations, based on the identification of legal and ethical requirements that has been performed in the MANDOLA Deliverable D2.2<sup>1</sup>.

---

<sup>1</sup> MANDOLA deliverable D2.2 - *Identification and analysis of the legal and ethical framework*, version 2.2.4 of 12 July 2017, MANDOLA project (Monitoring ANd Detecting OnLine hAte speech) - GA n° JUST/2014/RRAC/AG/HATE/6652, <http://mandola-project.eu/publications/>.

## 2 Introduction

### 2.1 Background to the MANDOLA project

MANDOLA (Monitoring ANd Detecting OnLine hAte speech) is a 24-months project co-funded by the Rights, Equality and Citizenship (REC) Programme of the European Commission, 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.

#### 2.1.1 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.

### **2.1.2 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.

## **2.2 Purpose and scope of the report**

The purpose of the current report is to present the initiatives that have been taken in order to ensure the legal and ethical compliance of the MANDOLA research, under the supervision of the MANDOLA Ethics Committee.

These actions have been determined on the basis of the requirements of the European Convention of Human rights requirements (studied in the MANDOLA deliverable D2.2<sup>2</sup>) and of applicable data protection legislations.

---

<sup>2</sup> MANDOLA deliverable D2.2 - *Identification and analysis of the legal and ethical framework*, version 2.2.4 of 12 July 2017, MANDOLA project (Monitoring ANd Detecting OnLine hAte speech) - GA n° JUST/2014/RRAC/AG/HATE/6652, <http://mandola-project.eu/publications/>.



## **2.3 Document structure**

The document is structured as follows.

Section 3 provides a presentation of the actions that have been taken in order to ensure the compliance of the MANDOLA research with the European convention on Human rights requirements.

Section 4 provides a presentation of the actions that have been taken in order to ensure the compliance of the MANDOLA research with applicable data protection legislations.

Section 5 provides a conclusion.

### 3 Compliance of the MANDOLA research with the European convention on Human rights requirements

The MANDOLA research had firstly to comply with the European Convention on Human Rights's (ECHR) requirements, as explained in the MANDOLA Deliverable D2.2<sup>3</sup>. This means that four principles, namely the principle of legal basis, of legitimate aim, of necessity and of proportionality<sup>4</sup> had to be respected.

#### 3.1 Legal basis

The legal basis justifying potential personal data processing operations performed by the MANDOLA partners within the framework of the MANDOLA project is the national legislation on the protection of personal data of each data processing controllers or processors, which is supposed to be in line with the provisions of the E.U. Directive 95/46/EC. After May 2018, every processing operation will need to comply with the General Data Protection Regulation and the changes that will be brought as a consequence to domestic legislations<sup>5</sup>.

The legal basis justifying research relating to the information to be provided to Internet users, to the Internet Industry and to policy makers, including the definition of illegal hate speech, was the right to freedom of expression, notably ensured in Article 10 of the European Convention on Human Rights (ECHR).

#### 3.2 Legitimate purpose

The purpose of the MANDOLA project is to perform a research in order to contribute to the combat against online hate speech. In this regards, the purpose is legitimate according to the provisions of the ECHR that protect fundamental rights at stake, including articles 8, 10 and 11 of the ECHR.

#### 3.3 Necessity

Regarding the question of whether the MANDOLA research is "necessary", as required by the ECHR, in the light of the questions to be answered in this regard according to the Article 29 Data Protection Working Party<sup>6</sup>, we can note the following:

---

<sup>3</sup> MANDOLA deliverable D2.2 - *Identification and analysis of the legal and ethical framework*, version 2.2.4 of 12 July 2017, MANDOLA project (Monitoring ANd Detecting OnLine hAte speech) - GA n° JUST/2014/RRAC/AG/HATE/6652, <http://mandola-project.eu/publications/>, especially Sections 3.2 and 4.2.3.3.

<sup>4</sup> For the further description of each of these principles, please refer to the MANDOLA Deliverable D2.2, *op. cit.*, Section 4.1.3.2.

<sup>5</sup> MANDOLA Deliverable D2.2, *op. cit.*, Section 4.2.1

<sup>6</sup> Article 29 Data Protection Working Party in its *Opinion 01/2014 on the application of necessity and proportionality concepts and data protection within the law enforcement sector* (WP 211). See the MANDOLA Deliverable D2.2 Section 4.1.3.2, n°3. See also, British Institute of Human Rights, *Mapping study on projects against hate speech online*, Council of Europe editing 2012, p. 9, 2.1.1, 2, <https://rm.coe.int/16807023b4> (last accessed on 25 July 2017): "*The boundaries of what is regarded as hate speech under [the definition of hate speech proposed by the Committee of Ministers in its Appendix to Recommendation No. R (97) 20 of on "Hate Speech", and retained as a basis of work by several entities] [...] are likely to fall outside the boundaries of speech which is criminalised under national legislation. They are also likely to fall outside the boundaries of speech which should not be restricted under freedom of expression [...]. These are important points because*

- The MANDOLA research is seeking to address an issue which, if left unaddressed, may result in harm to or have some detrimental effect on society or a section of society. Indeed, a lot of information is already available in relation to hate speech, as well as several initiatives and reporting mechanisms. However, it is still very difficult to understand what is exactly online hate speech, which kind of hate speech is illegal and which actions against hate speech appear to be the most appropriate. This confusion affects both the combat against hate speech (that remains difficult) as well as the protection of fundamental rights (which might be impacted by disproportionate or non-appropriate actions against online hate speech).
- There is at this stage no evidence that the MANDOLA research will improve the combat against online hate speech, particularly if at the end of the project the delivered outcomes do unexpectedly not significantly contribute to prevent or reduce online hate speech, or if these results are ignored, not used or not taken into account. However, the MANDOLA consortium does believe that amongst the intended results lie two very interesting instruments, namely a monitoring system and a mobile reporting system, which will both aim to protect to the utmost extent internet users' personal information while making it easier to understand what kind of contents must be reported and to report these contents. In the same line, the MANDOLA consortium intends to do its best effort to provide policy makers, the Industry and Internet users with a valuable information, as objective as possible, in order to assist them in targeting hate speech more efficiently while avoiding adverse impacts on fundamental rights at stake. In addition, the MANDOLA consortium planned to avoid and has avoided, during research, the collection of users' personal information to the most possible extent.
- Existing measures against online hate speech include penal offences<sup>7</sup>, reporting mechanisms<sup>8</sup>, governmental initiatives<sup>9</sup>, private initiative of technical nature<sup>10</sup>, victim assistance<sup>11</sup>, and awareness pages and portals<sup>12</sup>. However, as already outlined above,

---

*the most common strategy of organisations working in this area appears to be to campaign for greater restrictions on content, or to campaign for content to be taken offline”.*

<sup>7</sup> See for example the Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, 28 January 2003, <https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/189> (last accessed on 12 September 2017).

<sup>8</sup> See for example the “Hate speech” page of the INHOPE network at <http://www.inhope.org/gns/internet-concerns/overview-of-the-problem/hate-speech.aspx> (last accessed on 12 September 2017).

<sup>9</sup> See for example the French action plan against racism and antisemitism 2015-2017 (Premier ministre, *La République mobilisée contre le racisme et l'antisémitisme, Plan national 2015-2017*, [http://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2015/09/racisme\\_antisemitisme-dilcra.pdf](http://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2015/09/racisme_antisemitisme-dilcra.pdf)) and the French 2016 mobilisation plan against hate and discriminations targeting LGBT people (DILCRAH, *Plan de mobilisation contre la haine et les discriminations anti-LGBT*, 21 December 2016, <http://www.gouvernement.fr/plan-de-mobilisation-contre-la-haine-et-les-discriminations-anti-lgbt>). URLs last accessed on 12 September 2017.

<sup>10</sup> See for example Council of Europe, *Filtering, blocking and take-down of illegal content on the Internet*, study commissioned to the Swiss Institute of Comparative Law, <http://www.coe.int/en/web/freedom-expression/study-filtering-blocking-and-take-down-of-illegal-content-on-the-internet> (last accessed on 12 September 2017).

<sup>11</sup> See for example the page dedicated to victims of the Ligue internationale contre le racisme et l'antisémitisme (International league against racism and anti-Semitism - LICRA): <http://www.licra.org/victimes/?statut=victime> (last accessed on 12 September 2017).

<sup>12</sup> See for example the No Hate Speech Movement, *“a youth campaign of the Council of Europe for human rights online, to reduce the levels of acceptance of hate speech and develop online youth participation and citizenship, including in Internet*

online hate speech and the understanding of this phenomenon seem to be still an issue<sup>13</sup>, as well as undue limitation of fundamental rights at the occasion of the combat against it<sup>14</sup>. The objective and added value of the MANDOLA research is to favour a best identification of potential illegal speeches, a quicker report of these speeches, a better understanding of the phenomenon and to favour best practices in terms of private initiatives, among those that are respectful for other fundamental rights at stake. This last issue is of importance since online hate speech seems to be accompanied with some actions belonging to private justice<sup>15</sup>, which constitute a threat for several fundamental rights and freedoms<sup>16</sup>.

- In relation to hate speech, opinions that favour criminalisation<sup>17</sup> and preventive actions taken by Internet stakeholders to prevent and delete illegal hate speech on internet servers<sup>18</sup> coexist with opinions that favour de-criminalisation<sup>19</sup>, education and adapted public policies<sup>20</sup>, and/or call to punish crimes rather than hide them<sup>21</sup>. In addition, several rights such as the rights to privacy and to free speech might be

---

*governance processes*": <http://www.nohatespeechmovement.org/>. This campaign is for example relayed by the French Ministry for Youth, Education and Sports (<http://mic.jeunesse.gouv.fr/no-hate-presentation/mouvement-contre-le-discours-de-haine/>). URLs last accessed on 12 September 2017.

<sup>13</sup> See for example Asja Rokša Zubčević, Stanislav Bender, Jadranka Vojvodić, *Media regulation authorities and hate speech*, June 2017, Council of Europe publishing, <https://rm.coe.int/media-regulatory-authorities-and-hate-speech/16807338f5>, p.10: "there is no universal definition of hate speech, which is a matter of large debate on many international fora. Some critics have argued that the term hate speech is used to silence critics of social policies that have been poorly implemented. On the other hand, there are many issues surrounding hate speech and global definition might not be the best tool to dealing with them".

<sup>14</sup> Young People Combating Hate Speech On-line, *Mapping study on projects against hate speech online*, prepared by the British Institute of Human Rights, 15 April 2012, Council of Europe publishing 2012 (DDCP-YD/CHS (2012)), <https://rm.coe.int/16807023b4> (last accessed on 12 December 2017), for example Section 2.1.1, 2, p. 9.

<sup>15</sup> See inter alia Council of Europe, *Recommendation CM/Rec(2012)4 of the Committee of Ministers to Member States on the protection of human rights with regard to social networking services*, [https://search.coe.int/cm/Pages/result\\_details.aspx?ObjectID=09000016805caa9b](https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa9b) (last accessed on 12 September 2017), especially §3 ("Threats (to Human Rights) may, in particular, arise from lack of legal, and procedural, safeguards surrounding processes that can lead to the exclusion of users"), §5 and §10.

<sup>16</sup> See for ex. Council of Europe, *Recommendation CM/Rec(2016)5 of the Committee of Ministers to Member States on Internet freedom*, [https://search.coe.int/cm/Pages/result\\_details.aspx?ObjectID=09000016806415fa](https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016806415fa)

<sup>17</sup> See for example Peter Moore, *Half of Democrats support a ban on hate speech*, 20 May 2015, <https://today.yougov.com/news/2015/05/20/hate-speech/> (last accessed on 25 July 2017).

<sup>18</sup> See for example Eric Silverberg, Carrie Charpentier, Adam Goldman, Karina Luevano, and Jeffrey Petit, *ISP Censorship*, especially in "History of ISP Censorship", <https://cs.stanford.edu/people/eroberts/cs181/projects/1998-99/nuremberg-files/censorship.html>; Joe Mc Namee, "Self-regulation of content by the online industry", in *The Online Media Self-Regulation Guidebook*, Ed. by A. Hulin and M. Stone, OSCE Representative on Freedom of the Media, 2013, pp. 44 et seq., <http://www.osce.org/fom/99560?download=true> (URLs last accessed on 25 July 2017).

<sup>19</sup> See for example Sandy Starr, "Understanding Hate Speech"; in *Hate Speech on the Internet*, pp. 10 et seq., <http://www.osce.org/fom/13846?download=true>; Eric Heinze, *The case against hate speech bans*, 9 April 2014, <http://www.eurozine.com/the-case-against-hate-speech-bans/> (also published in *Nineteen arguments for hate speech bans – and against them*, <http://freespeechdebate.com/discuss/nineteen-arguments-for-hate-speech-bans-and-against-them/>; In defence of hate speech, 15 December 2016, <https://www.economist.com/news/leaders/21711914-criminalising-offensive-language-only-empowers-bigots-defence-hate-speech>; Peter Tatchell, *Argument – Should hate speech be a crime?*, 1<sup>st</sup> December 2012, <https://newint.org/sections/argument/2012/12/01/is-hate-speech-crime-argument> (URLs last accessed on 25 July 2017).

<sup>20</sup> See for example Iginio Gagliardone, Danit Gal, Thiago Alves, Gabriela Martinez, *Countering online hate speech*, UNESCO, 2015, especially pp. 46 and s., <http://unesdoc.unesco.org/images/0023/002332/233231e.pdf>.

<sup>21</sup> See European Digital Rights, *Internet blocking - crimes should be punished and not hidden*, [https://edri.org/wpcontent/uploads/2013/12/blocking\\_booklet.pdf](https://edri.org/wpcontent/uploads/2013/12/blocking_booklet.pdf).

directly impacted by unfettered proactive monitoring<sup>22</sup>. As a consequence, it was of utmost importance, during the MANDOLA research, to make sure that technical instruments to be developed in order to combat hate speech as well as MANDOLA recommendations and advices are created or prepared taking into account all these schools of thought, in order to continuously question the relevance and opportunity of the products under development in relation to their specific purposes. This way, the MANDOLA research was in a position to fulfil the necessity requirement through the provision of useful tools and of constructive information to the current debates.

- The opportunity and relevance of the MANDOLA technical developments, taking into account oppositions expressed by society in relation to the monitoring of Internet content, the difficulty to define illegal hate speech and the sanctioning of illegal hate speech, have been *inter alia* ensured, during the life of the project, through the strict application of the ECHR and of the personal data legislation requirements. This has especially led to avoid to the utmost extent the processing of personal data, which have been strongly protected where avoidance was impossible with certainty (for example where personal data are hidden in a text or behind an URL).

To conclude, it seems that the MANDOLA research complies with the principle of necessity, provided that it succeeds to bring clarity to the current context, if it correctly takes into account divergent views and if it provides for sufficient safeguards to make technical developments as less intrusive as possible into people's freedoms. These requirements seem to have been fulfilled during the life of the project, as it will be further demonstrated during the assessment of the compliance of the MANDOLA research with personal data protection requirements<sup>23</sup>.

In addition, to be acceptable, the MANDOLA research must be of a proportionate nature.

### 3.4 Proportionality

Regarding the question of whether the MANDOLA research is "proportionate", as required by the ECHR, in the light of the questions to be answered in this regard according to the Article 29 Data Protection Working Party<sup>24</sup>, we can note that it appears strictly necessary provided that some particular safeguards are in place.

#### 3.4.1 Strict necessity of the interference

The MANDOLA research may be considered as "strictly necessary", as long as data protection laws are respected, including the requirement of personal data minimisation.

---

<sup>22</sup> See for ex. Council of Europe, *Recommendation CM/Rec(2016)5 of the Committee of Ministers to Member States on Internet freedom*, *op. cit.*

<sup>23</sup> See Section 4 of the current report.

<sup>24</sup> Article 29 Data Protection Working Party in its *Opinion 01/2014 on the application of necessity and proportionality concepts and data protection within the law enforcement sector* (WP 211). See the MANDOLA deliverable D2.2 - *Identification and analysis of the legal and ethical framework*, version 2.2.4 of 12 July 2017, MANDOLA project (Monitoring ANd Detecting OnLine hAte speech) - GA n° JUST/2014/RRAC/AG/HATE/6652, <http://mandola-project.eu/>, Section 4.1.3.2, 4.

### **3.4.1.1 Strict necessity in relation to the context of the research**

- The MANDOLA research aims at shedding light on the hate speech definition, context and issues; at helping the assessment of hate speech spread online; at answering Internet users' questions and as easing Internet users' reports. Provided that personal data collected or otherwise processed are reduced to the minimum required to serve the afore-mentioned purposes, these actions seem to be adapted to the severity of the social need, which is to clarify the context, the phenomenon, and to identify the appropriate means to combat hate while considering all the opinions in this field and fundamental rights preservation requirements.
- The behaviour that is intended to be restricted (through the combat against online hate speech) is to hurt people in their dignity and right to non-discrimination, which is not a legitimate behaviour. Moreover, the MANDOLA research is adapted to the new challenge which is the increasing use of the Internet in order to spread hate speech<sup>25</sup>. All these elements show that the MANDOLA research seems to be adapted to its context.

### **3.4.1.2 Strict necessity in relation to the scope of interferences**

- **In relation to the volume of processed information**, the MANDOLA consortium has ensured that no direct personal data were collected and that the collection and processing of indirect personal information were reduced to the minimum. Only some texts and some URLs have been collected in order to train the hate speech classifier with the support of social scientists, and a function ensured that user-names and Twitter mentions were removed from these texts. Other texts coming from Twitter and Google streams were dynamically analysed in real-time, in order to extract from them only a small number of data elements that were namely the qualification of the content on a legal point of view (hate speech or non-hate speech), the type of hate speech, and the geolocation (timestamp of the tweets where the user has enabled the location service), reducing the decimals of the coordinates to 3 in order to minimise the possibility of any identification of a particular individual. These elements were stored in a secured database and visualised in the dashboard.

As a result, no personal data have been processed during research, unless accidentally, because of the nature and content of the sources that were scanned (and which might contain signs which are likely to indirectly refer to one single person and which cannot be automatically removed). In any case, identifying authors of potentially illegal content was not an objective of neither the MANDOLA research, nor of the system under development.

However, in a full ethical approach, the relevant data protection legislations have been respected, as explained below in this report. In particular, the relevant data protection authorities have been notified, processed data have been stored on a secure server, transmitted securely to this server and accessed by relevant MANDOLA partners only on a need-to-know basis, and social analysts have signed agreements of confidentiality and of non-misuse. In addition, in order to avoid Intellectual property rights issues, all the online

---

<sup>25</sup> See for example Young People Combating Hate Speech On-line, *Mapping study on projects against hate speech online*, prepared by the British Institute of Human Rights, 15 April 2012, Council of Europe publishing 2012 (DDCP-YD/CHS (2012)), available at <https://rm.coe.int/16807023b4> (last accessed on 21 August 2017), Section 4, p. 15; Steve Lohr, *Online Hate Sites Grow With Social Networks*, 16 March 2010, <https://bits.blogs.nytimes.com/2010/03/16/online-hate-sites-grow-with-social-networks/> (last accessed on 25 July 2017).



newspapers that have been automatically scanned have authorised the relating MANDOLA partner (namely UAM) to do it, for the purposes of the MANDOLA research.

**In relation to the information prepared for Internet users, policy makers and the industry**, the provision of objective and exhaustive information has been a constant preoccupation MANDOLA consortium, in order to ensure the usefulness and the proportionality of the research outcomes.

**In relation to the information intended to be provided by the monitoring dashboard and the smartphone app**, a particular care has targeted the issue of the display by geographic areas of hate speech statistics, in order to prevent stigmatisation, and the issue of providing information that could not mislead smartphone users, policy makers and the Internet industry. Results of these precautions can be found in the MANDOLA Deliverable D2.4b<sup>26</sup> as regards the MANDOLA outcomes. In relation to the research, all the persons working on the project were aware that the obtained results did not enable to draw conclusions in relation to the behaviour of inhabitants of considered geographical areas, all demonstrations given to people who were external to the MANDOLA consortium did blur results in order to prevent any ethical incident, and information dedicated to Internet users, policy makers and the Internet industry has been carefully written in order to reflect the MANDOLA results and their significance as objectively as possible.

- **Persons concerned by the processing operations** that feed the monitoring dashboard under development have been numerous, since automatic analyses of web contents and of social networks have been necessary. However, as already mentioned, the monitoring dashboard and the associated crawler did not collect personal data (beyond some texts and URLs stored separately at UCY in order to train the hate classifier, which are not supposed to be personal data but which might include such kind of data accidentally). On the opposite no Internet user has been concerned by the processing operations that feed the smartphone app, since it has been developed with data provided by the MANDOLA developers and has not been used during research by external smartphone users. In this context, the number of people concerned did not appear to be an issue as long as relevant data protection legislations were respected in relation with information that might be of personal nature.

**On the opposite, the MANDOLA research consisting of providing information did not impact any individual**, since results have not been published before finalisation. In addition, they have been the subject of a privacy impact assessment, along with the technical research. This privacy impact assessment, which includes a privacy and personal data protection compliance check, is presented in the MANDOLA Deliverable D2.4b.

- Finally, on the question of whether the MANDOLA outcomes leave some scope for the potentially limited freedoms<sup>27</sup>, including privacy, the answer is clearly yes, taking into account the elements analysed above.

---

<sup>26</sup> See the MANDOLA Deliverable *D2.4b (final report) - Privacy Impact Assessment of the MANDOLA outcomes*, September 2017, <http://mandola-project.eu/publications/>, especially Section 3.3.1.5, Section 3.4.2, Section 4.1.1.2; Section 4.1.2.1; Section 4.2.3.

<sup>27</sup> In other words, if they do not extinguish the possibility to exercise these freedoms. See the MANDOLA Deliverable D2.2 - *Identification and analysis of the legal and ethical framework*, version 2.2.4 of 12 July 2017, MANDOLA project (Monitoring ANd Detecting OnLine hAte speech) - GA n° JUST/2014/RRAC/AG/HATE/6652, <http://mandola-project.eu/>, Section 4.1.3.

### **3.4.1.3 Strictly necessary in relation to the nature of interferences**

The last question regarding proportionality is to know if other measures, of a less intrusive nature, could be considered, and, if yes, why they have been rejected.

The answer seems to be negative in relation to the provision of information, since basic information does already exist in the field of the combat against hate speech and since the benefits of the MANDOLA project are expected to partly lie in broadness of its legal and field research.

The question is more difficult to answer in relation to the monitoring dashboard and the smartphone app. Indeed, the content of the project has been defined before it started, and its purpose was not to investigate the possibility to create other systems than the ones that have been promised, but only to ensure the relevance of these systems and to investigate the possibility of building them in the best possible way.

However, it appears that producing statistics on online hate speech is necessary<sup>28</sup>, and that it inevitably involves the scan of web content. It also appears that an easy-to-use reporting system is a victims' need<sup>29</sup>, and is particularly beneficial on smartphones, which are widely used today to access the Internet. In addition, only a smartphone app enables to provide the functionalities proposed by the MANDOLA consortium, knowing that smartphone users will stay free to use other reporting mechanisms and first the MANDOLA reporting portal instead of the app, and to send their report directly to the relevant assistance service, using the web. In this sense, it does not appear that less intrusive techniques of another nature could have been proposed.

### **3.4.2 Limited by appropriate safeguards (summary)**

For clarity reasons, preceding analyses already highlighted the safeguards proposed by the MANDOLA consortium in order to palliate weaknesses of the previous steps of the assessment. These recommended safeguards can be summarised as follows:

- Personal data collection and processing have been reduced to the minimum required to serve the MANDOLA research purposes, namely to those accidentally included in the Twitter and web texts that have been collected, in addition to the collection of URLs (which might be indirectly personal) and locations (timestamp of the tweets where the user had enabled the location service).
- No direct personal data have been voluntarily collected and those that could be found in collected contents have been removed. More precisely, user-names and Twitter mentions have been removed where detectable and decimals of geographical coordinates have been reduced to 3 in order to minimise the possibility of any identification of a particular individual.
- Relevant data protection legislations have been respected, as explained below in this report. In particular, relevant data protection authorities have been notified, processed data have been securely transmitted and stored on a secured server and accessed by

---

<sup>28</sup> See for example the European Union Agency for fundamental rights, *Hate crime*, <http://fra.europa.eu/en/theme/hate-crime>; Ester Strømme, *Hate Crime and Hate Speech in the European Union*, 6 Nov. 2014, *Foreign Affairs Review*, <http://foreignaffairsreview.co.uk/2014/11/hate-crime/> (URLs last accessed on 12 September 2017).

<sup>29</sup> See for example the European Union Agency for fundamental rights, *Hate crime*, *op. cit.*



relevant MANDOLA partners only on a need-to-know basis, and social analysts have signed agreements of confidentiality and of non-misuse.

- The provision of objective and exhaustive information has been a constant preoccupation MANDOLA consortium, in order to ensure the usefulness and the proportionality of the research outcomes consisting in providing information to Internet users, policy makers and the industry. This information has not been shared with concerned stakeholders before finalisation and a complete privacy impact assessment targeting these outcomes has been performed<sup>30</sup>.
- Particular care has targeted the issue of the display by geographic areas of hate speech statistics in the monitoring dashboard, in order to prevent stigmatisation, and the issue of providing information that could not mislead smartphone users, policy makers and the Internet industry. Especially, all the persons working on the project were aware that the obtained results did not enable to draw conclusions in relation to the behaviour of inhabitants of considered geographical areas, all demonstrations given to people who were external to the MANDOLA consortium did blur results in order to prevent any ethical incident, and information dedicated to Internet users, policy makers and the Internet industry has been carefully written in order to reflect the MANDOLA results and their significance as objectively as possible.

---

<sup>30</sup> See the MANDOLA Deliverable D2.4b (final report) - *Privacy Impact Assessment of the MANDOLA outcomes*, 30 September 2017, <http://mandola-project.eu/publications/>.

## 4 Compliance of the MANDOLA research with personal data protection requirements

In addition to the requirements of the ECHR, the MANDOLA research had to comply with data protection legal requirements. This means that the following principles, which have been identified in Deliverable D2.2<sup>31</sup>, had to be respected:

### 4.1.1 Legal basis

Taking into account that the MANDOLA research did not appear to involve important risk for privacy, and that research is explicitly targeted in the current data protection legislation, the MANDOLA works did not seem to need an additional legal basis than current legislation on personal data protection, which will be replaced in 2018 by the E.U. General Data Protection Regulation and future domestic legislations implementing the E.U. Directive on personal data protection for the police and criminal justice sector.

### 4.1.2 Legitimate, explicit and specified purpose

The purpose of assisting in the combat against online hate speech by shedding light on the hate speech definition, context and issues, by helping the assessment of its spread online, and by answering Internet users' questions and easing their reports, is legitimate, since it is in accordance with the law in the broadest sense (under the reserve of the necessity and proportionality of the measure).

Such a purpose is also specified, in other words sufficiently defined prior the time of the data collection, provided that the MANDOLA partners did not use processed data for other purposes, which has been ensured through confidentiality and non-misuse agreements.

Processing operations that have been included in this specification have been identified and detailed in several MANDOLA deliverables.

Such a purpose is moreover explicit, and appears to be understandable by anyone.

Regarding the question to know whether the potential collection or processing of personal data by (1) the scanning component of the monitoring dashboard and by (2) the module that trains the hate speech classifier are compatible with the first processing that led to the publication of the information on the public Internet<sup>32</sup>, we can note the following, in the light of the compatibility test proposed by the Article 29 Data Protection Working Party<sup>33</sup>:

- On the first hand, the intended purpose of the publication of an article on a webpage, or of a comment on such a webpage, for the internet user, is the publicity of the information to the entire internet. The intended purpose of a publication on the public part of a social network, for the Internet user, is the publicity of this information to the

---

<sup>31</sup> See the MANDOLA deliverable D2.2 - *Identification and analysis of the legal and ethical framework*, version 2.2.4 of 12 July 2017, MANDOLA project (Monitoring ANd Detecting OnLine hAte speech) - GA n° JUST/2014/RRAC/AG/HATE/6652, <http://mandola-project.eu/>, Section 4.2.3.3.

<sup>32</sup> We do not consider the collection of data on Internet's private areas, which is not projected within the framework of both the monitoring dashboard and the smartphone app, and which would imply a higher interference with privacy and other rights, and should be covered by national rules relating to private electronic communications / correspondences interception.

<sup>33</sup> Article 29 Data Protection Working Party, *Opinion 03/2013 on purpose limitation*, 2 April 2013, WP203, III.2.2, p.23 *et seq.*

other users of this social network, and eventually to all Internet users if the author chooses to be visible from search engines. However, some social networks impose to Internet users to make publicly available certain identified information, such as their name. The purpose of the publicity of such information will be, in these cases, the wish to be present on this social network.

On the other hand, the intended purpose of the MANDOLA scanning component of the dashboard was during research to analyse the afore-mentioned information (in case the afore-mentioned information sources were scanned) in order to (1) detect its potential illegal nature, (2) classify it into a Subcategory of illegal contents, (3) collect some non-personal data namely geolocation (timestamp of the tweets where the user had enabled the location service) reduced to 3 decimals in order to avoid any identification of a particular individual.

Regarding the intended purpose of the MANDOLA module that trained the hate speech classifier, it was to (1) collect potentially illegal texts that have been afterward analysed by social scientists, to (2) store them in a database separated from the rest of the system, and to (3) train the hate speech classifier thanks to the data they contained<sup>34</sup>.

- Regarding the content of the relation between these two groups of purposes, it consists in the possibility given to the public (of one given social network or more generally of the Internet) to access personal information, and, as a consequence, to use this information according to its context of publication. Therefore, potential personal data processing operated by the MANDOLA research might have partly gone beyond this relation, since (1) they were including data coming from persons who would have liked to restrict the publication of these data to the natural persons using the concerned social network only, and since (2) they aimed at using information out its context of publication by (2a) extracting from it non-personal information or (2b) extracting from it some data, further analysed by non-human means, in order to train a hate classifier. However, in the first case the use of the information could not impact the data subject since the latter could not be recognised, and in the second case the indirect personal data that might potentially have been included in collected texts were not supposed to be used, neither by the system nor by humans, and has not been used.
- Regarding reasonable expectations of Internet users or of the users of potentially scanned social networks, in terms of private life protection, the possibility that a third party will access and therefore use their information is at least known by users, as soon as they do accept this publicity, either in the case this publicity has been freely chosen or in the case it has been partly imposed by the general conditions of use of the concerned social network (publications that have been partly "imposed" by a social network - usually at least name and photographs - may lead to a higher expectation

---

<sup>34</sup> The assessment of the illegality of a content is based on a) the classification algorithm and b) the ground truth data set. Specifically, UCY has used the Stochastic Gradient Descent classifier from the following library: <http://scikit-learn.org/stable/modules/sgd.html#classification>. The SGD Classifier has a number of parameters (see [http://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.SGDClassifier.html#sklearn.linear\\_model.SGDClassifier](http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html#sklearn.linear_model.SGDClassifier)). Using the Randomised search method (see [http://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.RandomizedSearchCV.html](http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html)), the research team has searched for the best parameters for the classifiers (based on their data set).

of privacy, in terms of non-reuse of this information, but have not been processed during the MANDOLA research). For the rest, all the Internet users who publish information on the web or on social networks are not all the time conscious of the existing possibility of analysing deeply this information in order to correlate it with other data, and to deduce from the whole other information. As a consequence, the MANDOLA research might have gone beyond reasonable expectations of Internet users in terms of private life respect, where they were accidentally related to information that could be have been linked to the direct or indirect identity of a given internet user. However, such kind of link has never been done.

- Regarding the sensitivity of concerned data and the impact of a further processing of these data, especially in emotional terms, scanned data involved in the MANDOLA research could be of all types, due to the nature of the sources that have been scanned. Therefore, these data might have been sensitive within the framework of the training of the hate speech classifier (no potentially sensitive data have been collected outside this component), particularly because social networks have been involved, since Internet users use often such networks to publish very personal information (such as philosophical or religious views, health...). Mentions of criminal infringements (linked or not to an individual) might also have been collected and processed, since it is highly possible that they appeared in searched documents, given the subject of the research which was hate speech. The impact of a further processing might have been high, if such a processing had driven to take decisions against individuals or to enhance the publicity of the data involved. The existence of the processing itself, if Internet users had been informed of the processing but not of the safeguards put in place to protect rights and freedoms, might have driven to restrict other freedoms than the right to private life, for example the freedom of expression, the freedom to communicate and the freedom to develop relationships with other human beings, if this information generates self-censorship behaviours.
- In order to compensate for the weaknesses found out during the previous steps of the compatibility evaluation, the MANDOLA consortium has implemented the following safeguards:
  - ✓ The MANDOLA consortium has made its best efforts in order to ensure that no individual was identifiable in the MANDOLA hate database. This implied to remove all detectable names and other visible signs that might have led to or that might have been personal data.
  - ✓ Since simple texts might accidentally lead to identify an individual, even if most visible personal data are removed, security measures have been applied against undue internal or external access, in order to ensure that access to a text or to a URL stored in the hate speech database pursued the only aim of verifying the illegal nature of one given content, in order to enhance the performances of the dashboard.

To this end, access to the information contained in the database and to originating URLs has been restricted to identified persons accredited to do it on a "need to know" or "need to use" basis, and social analysts have signed agreements of confidentiality and of non-misuse. The information was stored on a MANDOLA partner's server, namely a UCY server. At the end of the research, this information

will be kept at UCY as part of its data set under development which aims at training hate speech classifiers (and which has been and will be further developed independently from the MANDOLA project). It will be covered by the protection granted to the other information included in this data set, implemented by UCY.

- ✓ During the MANDOLA research relating to the monitoring dashboard, the MANDOLA consortium has taken great care of avoiding any decision that could restrict some individual's rights.
- ✓ Transparency has been ensured to the utmost extent, especially by publishing technical deliverables on the MANDOLA website as early as possible in the lifecycle of the project.

#### 4.1.3 Data quality

- ✓ Fairness of the processing is an obligation for all entities, at least where they do not belong to the police and criminal justice sector<sup>35</sup>. Within the context of the MANDOLA system, the issue of the fairness of the processing appears to be connected to the issue of the compatibility of the further processing, regarding the reasons that lead the Internet user to publish information on the Internet, since the conditions to be verified are in both case the same. This question has already been analysed in the previous point of the current study.
- ✓ Lawfulness of the processing implies processing operations that comply with the requirements of relevant data protection legislations, which the MANDOLA partners took the commitment to respect, and which have been implemented as a precautionary principle during the research performed by the consortium.
- ✓ Accuracy, reliability, completeness and up-to-dateness of personal data that may be processed during the MANDOLA technical research are principles that may pose more difficulties, since personal data that are potentially processed are not processed voluntarily in order to identify natural persons and attribute them some other kind of information, but accidentally, because of the nature of the scanned information sources. Therefore, it is neither really possible, nor a task attributed to the system, to verify if data that may be of a personal nature are accurate, reliable, complete or up-to-date. It would even be not desirable that the system performs such verifications, since it would enable to identify more precisely individuals, whereas the objectives are elsewhere. On the opposite, data potential unreliability operates in this situation to the benefit of natural persons, since it weakens the potential link that could be made between a given data and a real identifiable person. The MANDOLA researchers who could access these data were fully aware of the potential non reliability of the data that were processed.

#### 4.1.4 Data minimisation

Personal data that are processed must be adequate, relevant and not excessive (i.e. limited to the minimum necessary) in relation to the purposes for which they are processed. It

---

<sup>35</sup> See the MANDOLA deliverable D2.2 - *Identification and analysis of the legal and ethical framework*, version 2.2.4 of 12 July 2017, MANDOLA project (Monitoring ANd Detecting OnLine hAte speech) - GA n° JUST/2014/RRAC/AG/HATE/6652, <http://mandola-project.eu/publications/>.

seems that this requirement has been fully applied during the MANDOLA research, since (1) no personal data has been collected voluntarily, (2) only a mandatory number of texts has been stored in the MANDOLA hate speech database in order to train the hate classifier, (3) these texts have been anonymised to the utmost possible extent (taking into account the technologies available to perform this task and the aim of the research), and (4) for the rest, only non-personal information have been extracted from the web (such as the nature of the speech), beyond geolocation data which have been reduced to three decimals in order to prevent the possibility to identify a single person as being at the origin of a given Internet content.

#### 4.1.5 Time limitation

The only data that have been collected during the MANDOLA research and that might accidentally be of a personal nature, namely texts issued from the Internet and their URLs, will be kept by UCY on UCY servers at the end of the MANDOLA research, solely to the extent that they can further serve the purposes of training hate classifiers, under the control of UCY which will take care of the compliance of this further storage with its applicable data protection legislation. No other entity will access to these data, outside authorisations provided by this legislation and the relevant data protection authority.

#### 4.1.6 Appropriate legal ground

The legal ground for (potentially) processing personal data, in relation with processing operations performed by the MANDOLA partners operating the data sets used to train the hate speech classifier and to those who have processed online data in order to extract from sources non-personal information, is "*the legitimate interests pursued by the controller (...)*", which may be evoked "*except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject*" , on the basis of article 7, f of Directive 95/46/EC and of the national provisions that have implemented this article (the GDPR contains equivalent provisions in its article 6 f).

In order to be granted with the benefit of the "*legitimate interest pursued by the controller*" legal basis, several criteria must be met on the basis of the tests which performance is recommended by the Article 29 Data Protection Working Party<sup>36</sup>:

- ✓ Processing operations must be necessary to pursue the system's purposes. On this field, we have already analysed that data collected and processed during research have been limited to the utmost extent;
- ✓ As already discussed, the technical systems' purposes are lawful and present a real and present interest<sup>37</sup>. They have moreover been clearly articulated<sup>38</sup>.
- ✓ As already discussed, processed data might be sensitive, and certain processing operations may impact some individuals' rights and go beyond data subjects' reasonable

---

<sup>36</sup> Article 29 Data Protection Working Party, *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC* (WP 217), 9 April 2014, III.3.1, p. 25, [http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217\\_en.pdf](http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf) (last accessed on 23 May 2014).

<sup>37</sup> See above the sub-section relating to the privacy and personal data protection requirements at the Council of Europe level (necessity and proportionality principles).

<sup>38</sup> See above the sub-section relating to the requirement of legitimate, explicit and specified purpose.

expectations, essentially where data from social networks are processed<sup>39</sup>. In addition, data controllers have a dominant position.

- ✓ However, the scale of the processing is very limited, and precautions have been taken in order to ensure that no personal data is used in order to identify an individual or to take a decision against an individual. Given the fact that the MANDOLA technical systems are susceptible to enhance the combat against hate speech, the balance between this interest and the harm suffered by individuals due to these systems - which appears to be very limited in practice - seems to be in favour of the MANDOLA systems (and therefore the controllers') legitimate interests.

As a conclusion, potential personal data processing operations performed during the MANDOLA research appear to be based on an appropriate legal ground.

#### **4.1.7 Data subject information**

In relation to the data that might be related to individuals, included in the analysed and stored web or social networks contents, data subject's information is not possible, since the latter are not identified. However, transparency of research has been ensured through the publication of technical deliverables on the MANDOLA website, as well as a specific note relating to personal data protection. The MANDOLA partners are moreover clearly identifiable on the MANDOLA website from the beginning of the project, including UCY and UAM which are potential data controllers<sup>40</sup>.

#### **4.1.8 Data subjects' rights of access, communication, rectification and erasure**

During the MANDOLA research, any internet user could address a request to UCY, as well as any other MANDOLA partners, based on information available on the MANDOLA website.

#### **4.1.9 Prohibition of automated decisions**

The MANDOLA research did not aim at taking decisions that produce legal effects concerning natural persons who would have been potentially identifiable in the database.

However, the MANDOLA non-technical developments aiming at providing information to several categories of recipients, even if they do not contain any personal data, might have led to decisions against individuals or groups or individuals, particularly as they relate to the behaviours to adopt when facing a potentially illegal content, to the appropriate definition of hate speech and to the places where policies against hate speech should be ideally implemented. Even if this issue is going slightly beyond the question of the prohibition of automated decisions, since no personal data is involved, it is strongly linked to it<sup>41</sup> and has for this reason been a constant concern. As a result, no information has been shared if not

---

<sup>39</sup> See above the sub-section relating to the requirement of legitimate, explicit and specified purpose.

<sup>40</sup> See Section 4.1.13 of the current report.

<sup>41</sup> Indeed, the data protection legislation is a practical implementation of the ECHR principles, applied to personal data processing. The fact that this legislation does not regulate processing operations other than those that include personal data does not mean that the ECHR principles do not apply to the latter processing operations.

blurred with external parties, and a privacy impact assessment of the MANDOLA products has been performed<sup>42</sup>.

#### **4.1.10 Enhanced protection of some sensitive data**

Sensitive personal data (data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs...), as well as location data (where published in plain text by Internet users on Internet public parts), may be processed by the data set that trains the hate-speech classifier. This is due to the nature of the system (which aims at detecting potential penal offences) and to the nature of the data used by this system (web and social media content).

However, such processing of personal data is supposed to be very rare in the dashboard hate speech database since user-names and Twitter mentions are removed from texts that are stored. In addition, the MANDOLA partners ensured that it was very clear for all persons working on the MANDOLA research that data might be non-reliable, and cannot lead to conclusions in relation with a particular individual.

As a consequence, it seems that all appropriate safeguards have been taken to avoid the processing of sensitive data to the utmost possible extent.

For the rest, no location data identified as such, no communications and no traffic data are processed.

#### **4.1.11 Security and confidentiality of the processing**

The dashboard's hate speech database (which is associated with the data set that may be used to train the hate speech classifier, and which is a module that is independent from the MANDOLA dashboard) is hosted at UCY and benefits from several security measures, including data encryption, password protection and the use of HTTPS, designed to ensure a level of security appropriate to the risk<sup>43</sup>, and consequently to protect the texts and URLs that are stored against alteration, unauthorised disclosure or access, and against all other unlawful forms of processing.

#### **4.1.12 Data protection authority supervision**

Data processing operations that have taken place during the MANDOLA research have been performed under UCY's control, which is the data controller for these operations. As a result, UCY took the commitment to notify the MANDOLA processing operations to the Cypriot Data Protection Authority. On 18 December 2015, the latter Data Protection Authority sent a letter stating that *"the data processing that will happen in the context of the European Research Project MANDOLA (...) appears to be in line with article 7(1)(2) of the Personal Data Protection (Protection of the Human) Laws of 2001 to 2012 (N. 138(I)/2001 as it was modified with N. 37(I)/2003 and N105(I)/2012), hereafter mentioned as "the Law" and as a result it has been registered in the file of Records and Processing that the Data Protection*

---

<sup>42</sup> See the MANDOLA Deliverable D2.4b (final report) - Privacy Impact Assessment of the MANDOLA outcomes, 30 September 2017, <http://mandola-project.eu/publications/>.

<sup>43</sup> This wording is used in the new E.U. data protection regulation. However, this is a traditional requirement in the area of risk management.



*Authority has based on articles 7(4), 23(1)(ι) and 24(1)(α) of the Law. This file is accessible to the public according to article 24(2) of the Law”<sup>44</sup>.*

In addition, UAM collected automatically some texts from online newspapers (in their part dedicated to comments from Internet users), in order to send them to UCY together with their URL, with the aim of feeding the data set that trains the hate speech classifier. To this end, UAM asked the authorisation of concerned newspapers and scanned only those for which an authorisation had been given. It notified the processing to the Spanish Data Protection Authority, which acknowledged receipt of the notification in June 2017.

#### **4.1.13 Liability and accountability of the data controller**

During the MANDOLA research, UCY was the data controller in relation with the MANDOLA processing operations that might accidentally include personal data, since these operations have been performed by its own components or under its control. As a result, UCY was responsible for respecting the Cypriot Data Protection legislation, and has fully assumed this responsibility.

In addition, UAM notified the processing operations it has performed in order to feed the UCY module aiming to train the hate speech classifier, and has also committed to respect the Spanish data protection legislation.

#### **4.1.14 Adequate level of protection in some case of data transfers**

The MANDOLA research did not imply any transfer of personal data to third parties.

---

<sup>44</sup> Translated from Greek Language by UCY.

## **5 Conclusion**

The MANDOLA research appears to comply with ethical and legal requirements in the field of fundamental rights protection, based on the initiatives taken in this purpose by the MANDOLA consortium during the life of the MANDOLA research, and presented in the current report.