Code of Conduct on countering illegal hate speech online: **Results of the 3rd monitoring**

Fact sheet | January 2018

Věra Jourová
Commissioner for Justice, Consumers and Gender Equality

Directorate-General for Justice and Consumers

The *Code of Conduct on Countering Illegal Hate Speech Online's* third evaluation reveals continuous progress on the removal of illegal hate speech.

On average, IT companies **removed 70 % of the illegal hate speech notified** to them. Compared with the removal rate of 59 % in the second monitoring exercise (May 2017) and with the 28 % in the first monitoring (2016), we are observing a clear and **steady increase in the removal of hate speech content**.

Today, **all IT Companies meet the target of reviewing the majority of the notifications within 24 hours**, reaching an average of more than 81 %. This represents a significant improvement compared to the shares of 40 % and 51 % of notifications assessed within 24 hours recorded in the previous monitoring rounds.

Since the adoption of the Code in May 2016, IT Companies have strengthened their reporting systems, making it easier to report hate speech, and have improved their transparency vis-à-vis notifiers and users in general. They have increased their staff of reviewers and the resources allocated to content management.

They have strengthened their cooperation with civil society organisations through dedicated partnerships and programmes, as well as through regular trainings, to ensure a better understanding of reporting systems, national context and legal specificities related to hate speech.

In terms of transparency and feed-back to users sending notifications, there is is still scope for progress. Despite notable differences among the companies, on average almost **a third of the notifications do not receive a feedback**.

The Commission will continue to monitor regularly the implementation of the Code by the IT Companies with the help of civil society organisations.

Results of the 3rd monitoring exercise of the implementation of the Code of Conduct

1. Notifications of illegal hate speech

> In the 3rd monitoring exercise, **2 982 notifications were submitted** to the IT companies taking part in the Code of Conduct. This represents a further increase compared to the previous two monitoring exercises.

> It covered **27 Member States** (all except Luxembourg). **33 civil society organisations and 2 national authorities** sent notifications relating to hate speech deemed illegal to the IT companies during a period of 6 weeks (6 November to 15 December 2017). In order to establish trends, this exercise used the same methodology as the previous monitoring rounds (see Annex).

> 1 802 notifications were submitted through the reporting channels available to general users, while 1 180 were submitted through specific channels available only to trusted flaggers/reporters.

> **Facebook received the largest amount of notifications (1 408), followed by Twitter (794) and YouTube (780)**. These shares are comparable with the ones in the December 2016 and May 2017 exercises. Microsoft did not receive any notification.

> In addition to flagging the content to IT Companies, the organisations taking part in the monitoring exercise **submitted 511 cases of hate speech to the police, public prosecutor's bodies or other national authorities**.

2. Time of assessment of notifications

> In **81.7 % of the cases** the IT companies assessed the notifications **in less than 24 hours**, in 10 % in less than 48 hours, in 4.8 % in less than a week and in 3.5 % it took more than a week.

> Facebook assessed the notifications in less than 24 hours in 89.3 % of the cases and 9.7 % in less than 48 hours. The corresponding figures for YouTube are 62.7 % and 10.6 % and for Twitter 80.2 % and 10.4 %, respectively.

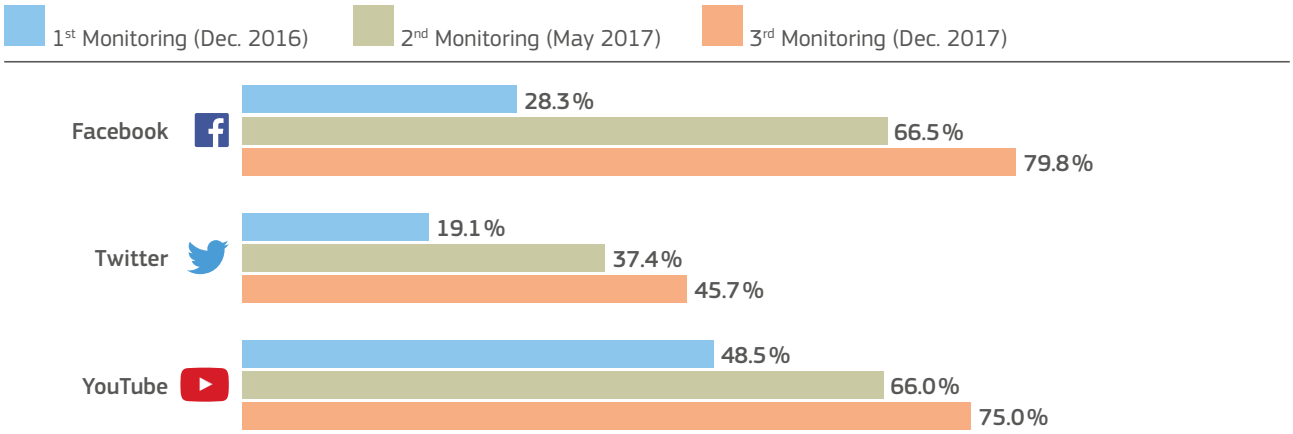
> The target of reviewing notifications within one day is now met by all IT Companies and there has been a steady progress compared to the previous monitoring exercises in May 2017 and December 2016, where respectively 51.4 % and 40 % of all responses were received within 24 hours while another 20.7 % and 43 % arrived after 48 hours. Twitter made the biggest improvement: in May 2017 only 39 % of the cases were reviewed within the day.

3. Removal rates

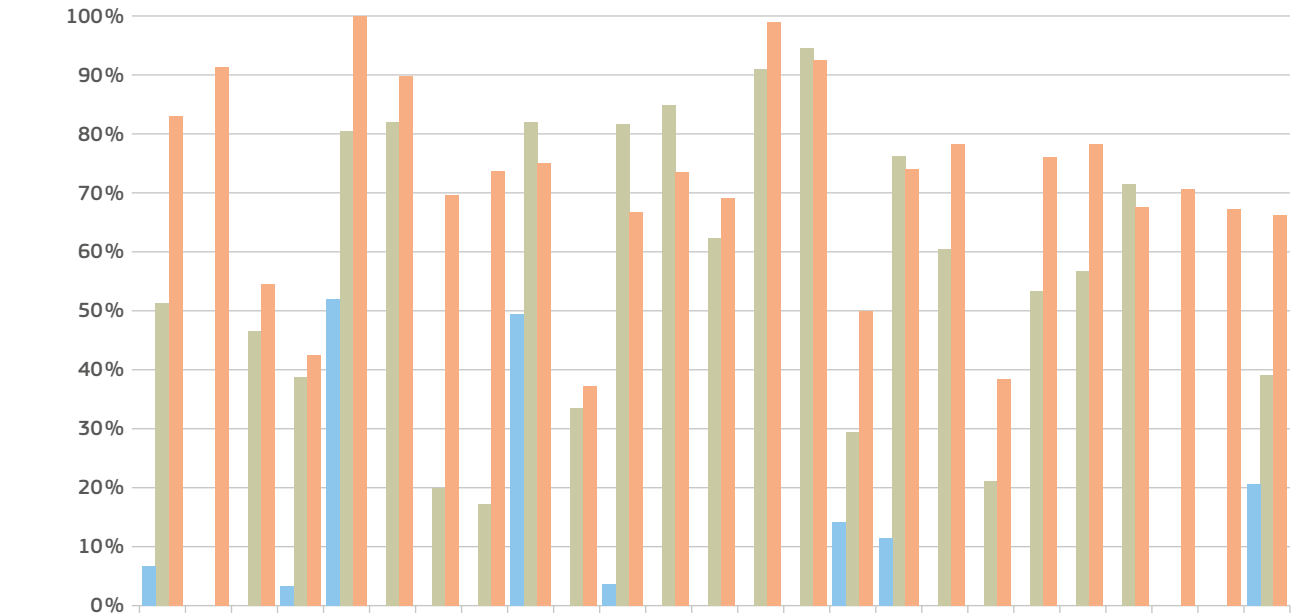
> Overall, **IT Companies removed 70 % of the content** notified to them, while 30 % remained online. This represents a significant improvement with respect to the removal rate of 59 % and 28 % recorded in May 2017 and December 2016 respectively.

> Facebook removed 79.8 % of the content, YouTube 75 % and Twitter 45.7 %. There has been **substantial progress** by all three companies compared to the results presented in May 2017 and December 2016.

Removals per IT company (in %)



Rate of removals per EU country (in %) ([1])



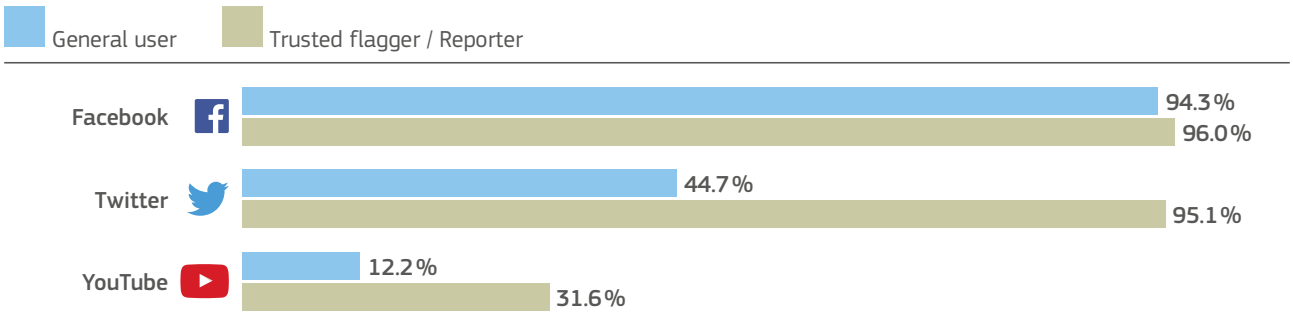
	BE	BG	CZ	DK	DE	EE	IE	ES	FR	HR	IT	CY	LV	LT	HU	NL	AT	PL	PT	RO	SI	SK	FI	SE	UK
1st monitoring (Dec. 2016)	6.9			3.4	52.0			0	49.5		3.6					14.3	11.4				0				20.5
2nd monitoring (May 2017)	51.2		46.5	38.9	80.1	81.8	20.0	17.2	82.0	33.6	81.7	84.8	63.3	90.9	94.5	29.5	76.1	60.3	21.0	53.3	56.8	71.4			39.3
3rd monitoring (Dec. 2017)	83.0	91.3	54.5	42.5	100.0	89.8	69.6	73.8	75.0	37.1	66.9	73.8	69.0	99.0	92.6	50	74.0	78.4	38.6	76.2	78.3	67.7	70.6	67.3	66.3

([1]) The table does not reflect the global issue on illegal hate speech online in a specific country and it is based on the number of notifications sent by each individual organisation. Malta and Greece are not included given the too low number of notifications made to companies (<20). For Luxembourg, no organization participated to this exercise.

4. Feedback to users and transparency

> On average, the **IT Companies responded with a feedback to 68.9 %** of the notifications received. Data show a certain disparity between IT companies when giving feedback to notifications. While Facebook sent feedback in response to 94.8 % of the notifications and Twitter to 70.4 % of cases, YouTube did so only in response to 20.8 % of the notifications. The corresponding figures in May 2017 were 93.7 %, 32.8 %, and 20.7 % respectively. The trend is positive, and Twitter made the most remarkable progress in relative terms.

Feedback provided to different types of user (in %)



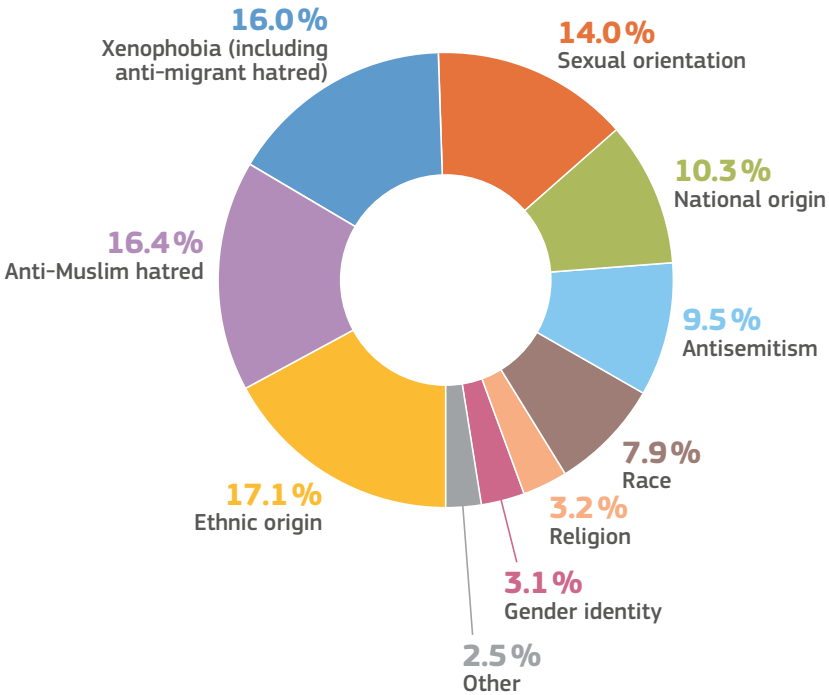
> Twitter and YouTube provide feedback more frequently when notifications come from trusted flaggers. Twitter provided feedback in response to 95.1 % of notifications made using the trusted flaggers' channel, but only gave feedback in response to 44.7 % of those made by general users. For YouTube, the corresponding figures were 31.6 % and 12.2 % respectively. Facebook provides systematic feedback to nearly all notifications (96 % and 94.3 %).

5. Grounds for reporting hatred

> Ethnic origin (17.1 %), anti-Muslim hatred (16.4 %) and xenophobia (16 %) were the most commonly reported grounds of hate speech.

> The results, which are in line with the trends in May 2017, confirm the predominance of racist hatred against ethnic minorities, migrants and refugees. Data on grounds of hatred are an indication of trends and may be influenced by the field of activity of the organisations participating to the monitoring.

Notifications per ground of hate speech (in %)



ANNEX

Methodology of the exercise

- The third exercise was carried out for a period of 6 weeks, from 6 November to 15 December 2017, using the same methodology as the first monitoring exercise.
- 33 organisations and 2 public bodies (in France and Spain) reported on the outcomes of a total sample of 2 982 notifications from all the Member States except for Luxembourg. An additional 9 cases were reported to other social platforms.
- The figures do not intend to be statistically representative of the prevalence and types of illegal hate speech in absolute terms, and are based on the total number of notifications sent by the organisations.
- The organisations only notified the IT companies about content deemed to be "illegal hate speech" under national laws transposing the EU Council Framework Decision 2008/913/JHA ([2]) on combating certain forms and expressions of racism and xenophobia by means of criminal law.
- Notifications were submitted either through reporting channels available to all users, or via dedicated channels only accessible to trusted flaggers/reporters.
- The organisations having the status of trusted flagger/reporter often used the dedicated channels to report content which they previously notified anonymously (using the channels for all users) to check if the outcomes could diverge. Typically, this happened in cases when the IT companies did not send feedback to a first notification and content was kept online.
- The organisations participating in the second monitoring exercise are the following:

COUNTRY	N° OF CASES
BELGIUM (BE)	
CEJI - A Jewish contribution to an inclusive Europe	13
Centre interfédéral pour l'égalité des chances (UNIA)	45
BULGARIA (BG)	
Integro association	23
CZECH REPUBLIC (CZ)	
In Iustitia	101
DENMARK (DK)	
Anmeldhad.dk / Reporthate.dk	80
GERMANY (DE)	
Freiwillige Selbstkontrolle Multimedia-Diensteanbieter e.V. (FSM e.V.)	45
ESTONIA (EE)	
Estonian Human Rights Centre	98
IRELAND (IE)	
ENAR Ireland	46
GREECE (EL)	
SafeLine / Forth	9
SPAIN (ES)	
Fundación Secretariado Gitano	116
Federación Estatal de Lesbianas, Gais, Transexuales y Bisexuales (FELGTB)	35
Spanish Observatory on Racism and Xenophobia (OBERAXE)	86
FRANCE (FR)	
Ligue Internationale Contre le Racisme et l'Antisémitisme (LICRA)	122
Platforme PHAROS	26
CROATIA (HR)	
Centre for Peace Studies	124
ITALY (IT)	
Ufficio Nazionale Antidiscriminazioni Razziali (UNAR)	269
CYPRUS (CY)	
Aequitas	103

COUNTRY	N° OF CASES
LATVIA (LV)	
Mozaika	20
Latvian Centre for Human Rights	106
LITHUANIA (LT)	
National LGBT Rights Oganisation (LGL)	105
HUNGARY (HU)	
Háttér Society	97
MALTA (MT)	
Malta LGBTIQ Right Movement (MGRM)	8
NETHERLANDS (NL)	
Meldpunt Internet Discriminatie (MiND)	1
Magenta Foundation	27
AUSTRIA (AT)	
Zivilcourage und Anti-Rassismus-Arbeit (ZARA)	107
POLAND (PL)	
HejtStop / Projekt: Polska	134
PORTUGAL (PT)	
Associação ILGA Portugal	101
ROMANIA (RO)	
Active Watch	63
SLOVENIA (SI)	
Spletno oko	60
SLOVAKIA (SK)	
digiQ	93
FINLAND (FI)	
Finnish Red Cross	34
SWEDEN (SE)	
Institutet för Juridik och Internet	52
UNITED KINGDOM (UK)	
Galop	100
Community Security Trust	53
Tell Mama/Faith Matters	13

([2]) http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:328:0055:0058:en:PDF