





20-22 NOVEMBER 2017 Sir Ian & Nancy Turbott Auditorium, Parramatta South Campus WESTERN SYDNEY UNIVERSITY

Building resilience and social cohesion by tackling the toxic online environment

Dr Andre Oboler

CEO, Online Hate Prevention Institute Lecture, La Trobe University Law School © Andre Oboler, 2017

The solution requires all of us

Some quick facts:

- 79% of Australian adults use social media, 59% of them every day [1]
- 74% of all Australian adults use Facebook, on average for 10 hrs a week [1]

My key message today:

A toxic online environment is not an "online problem" that can be divorced from the "real world" and left to "those technology people". Nor is it a problem that can be tackled without technological expertise. It is a societal problem needed an integrated response. **We must solve this together.** 0

THE PROBLEM

"Hate 2.0"

- First identified in 2008 through "antisemitism 2.0"[2]
- Normalizes hate speech as an accepted part of the fabric of the social media world
- Not creating bigots so much as bystanders willing to accept & defend the bigotry of others in their group
- It fosters rejection of the ideas of "opposing bigotry", "supporting social cohesion" and "supporting multiculturalism" as values defining our identity / group membership

Problem case study: MangoGate [3]

- This years NSW HSC English Exam with a poem by an Indigenous author
- A Facebook group of 70,000 HSC students student
- 1,000 posted or "engaged" with racist memes
- Others largely accepted the posts, as they had done with homophobic posts during the postal survey
- What they wouldn't accept was criticism of their group, or even those in the group engaging in racism, by in the media
- Racism and bigotry is acceptable, criticism of people for engaging in racism or bigotry is not

The implications of this toxic environment

- This value shift is not contained online and impacts discussion with family, in the work place, in education etc.
- Messages of hate become the environment of the online world [4] and are also embedded in fabric of real world society [5]
- Jeremy Waldron explains these messages as either:[5]
 - To targets: '[d]on't be fooled into thinking you are welcome here'
 - Or to the rest of society: '[w]e know some of you agree that these people are not wanted here... know that you are not alone... there are enough of us around to make sure these people are not welcome... [and] to draw attention to what these people are really like'.
- These messages, and their embedding in the environment, destroy resilience and community cohesion

TACKLING THE PROBLEM

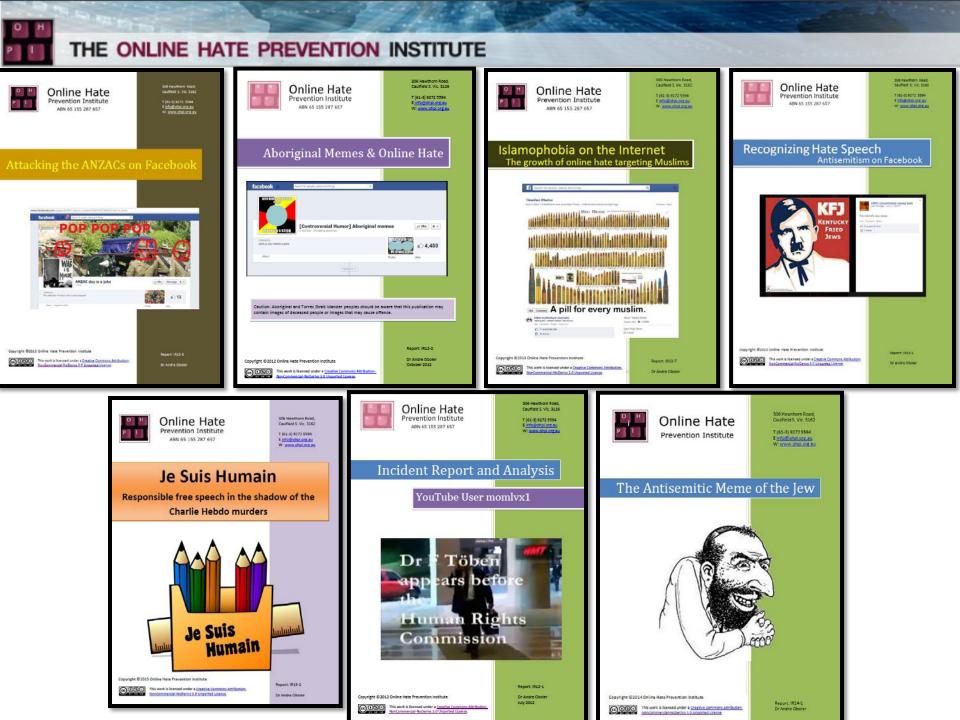
Traditional approach

- Geographically constrained
- Tackled without technical expertise
- Left to technology companies despite their fundamental conflict of interests
- Progress blocked when technology companies speak as experts saying nothing can be done

A Decade of Progress with a new approach

Year	Event
2007	Work started on antisemitism 2.0
2008	GFCA and publication of Antisemitism 2.0
2009	GFCA Working Group formed and set creating metrics as a key future challenge
2010	Created the Community Internet Engagement Project to meet the challenge
2011	"Fight Against Hate" solution to the metrics challenge presented to GFCA WG
2012	CIE Project spun off to become OHPI - an independent charity preventing harm from all forms of hate; YouTube report; Aboriginal Memes Report
2013	Antisemitism on Facebook Report, Islamphobia on Facebook report
2014	Fight Against Hate reporting tool launched; focus on briefings
2015	Measuring Antisemitism interim report (Facebook, YouTube Twitter); Responsible free speech report (private funding runs out)
2016	Measuring anti-Muslim hate interim report (mostly Facebook, but also YouTube and Twitter); Measuring Antisemitism final report (Facebook, YouTube Twitter); Creation of CSI-CHAT analysis tool. (Government funding runs out)
2017	Fight Against Hate redesigned by students (Running at minimum operating level)

CASE STUDY 1: THEMATIC REPORTS



CASE STUDY 2: FIGHT AGAINST HATE

0 9(1 0 0 0 9 Transparency xperts countability 0



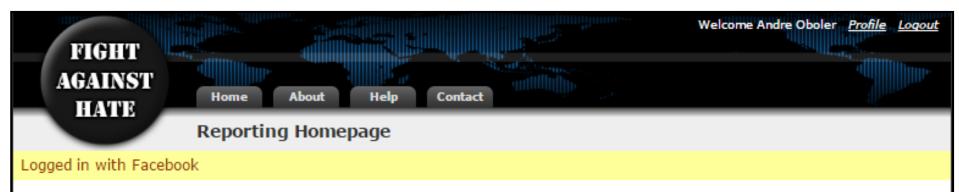
Fight Against Hate is a project of the Online Hate Prevention Institute. It's aim is to bring greater transparency to the way social media companies

respond when users report hate speech to them.

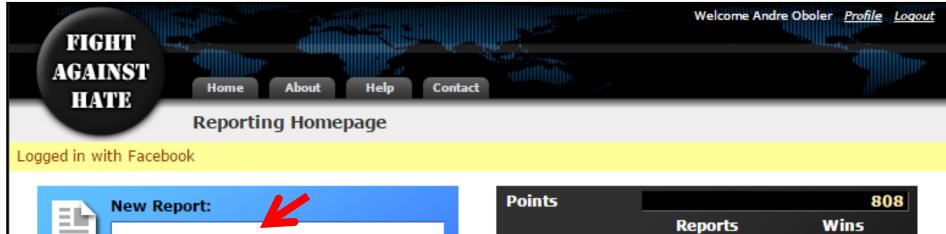
Fight Against Hate aims to empower the public, inform decision makers, and hold social media companies accountable for responding to breaches of their terms of service or when community standards fall below what the public expect.







Currently s	b address of offensive	content) Submit Report	Points Self Team	Report	ts 32 29	808 Wins 0 0
View Reports	[Unavailable]	Review Reports	Sys 5607 Reports	tem Wid 423	2	atus 867 _{Users}
			List by name	• 1	View Te	am Members



 (insert web address of offensive content)

 Currently supports:

 • Facebook • Youtube • Twitter

Submit Report

View ReportsImage: Constraint of the second sec

Points		808		
	Reports	Wins		
Self	32	0		
Team	29	0		
Svst	em Wide St	atus		
5607	4232	867		
Reports	Items	Users		
List by name	• View 1	eam Members		



Classification

- 🖲 Antisemitism
- Ocyberbullying
- Anti-Muslim hate
- Homophobia
- Hate targeting Indigenous Australians
- Hate targeting people with disabilities
- Promoting imminent violence / terrorism / threats to national security
- Misogyny / prejudice against women
- Other hate type not listed
- I don't wish to classify it

Confidence

How confident are you with this classification?

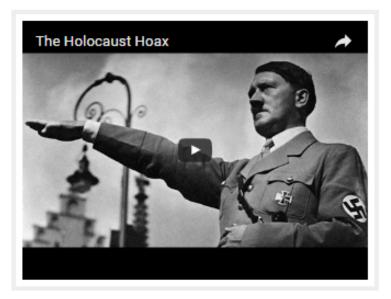
- Certain
- Very confident
- Confident
- Not very confident

Next

👚 НОМЕ

Item Preview

Item detected as: 🛗 YouTube 🛛 Video



ВАСК ТО ТОР 🛧



Antisemitism

- Promoting violence against Jews
- Holocaust denial
- Classic antisemitism (not Israel-related)
- New antisemitism (Israel-related)
- I don't wish to classify it

Confidence

How confident are you with this classification?

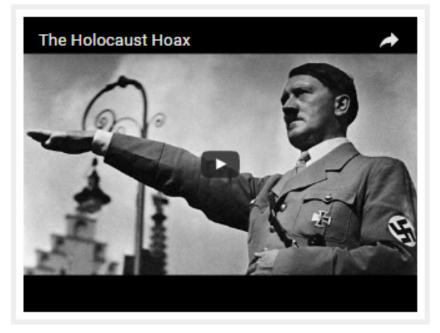
- Certain
- 📧 Very confident
- Confident
- Not very confident

Next

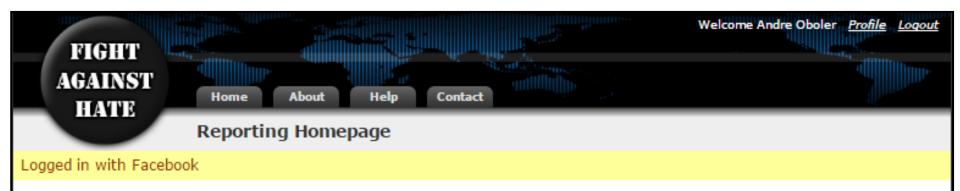
A HOME

Item Preview

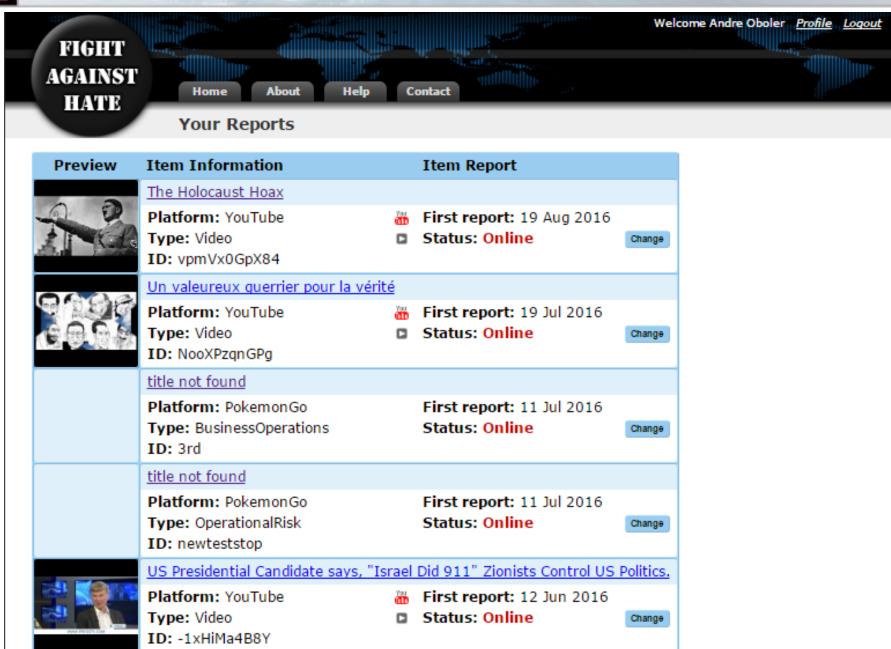
Item detected as: WouTube 🛛 Video

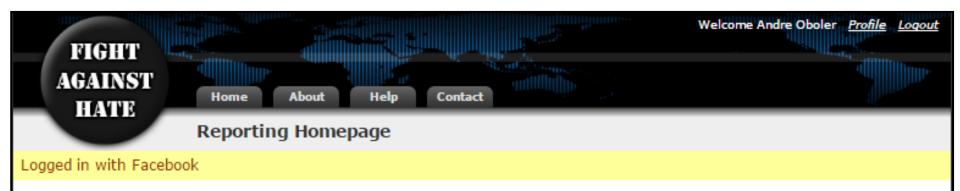


ВАСК ТО ТОР 🛧

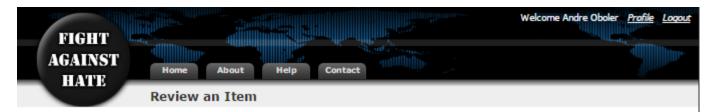


- New Report:		Points		808		
				Reports	Wins	
	(insert web address of offensive content)			32	0	
NEW Currently su • Facebook	Jpports: • Youtube • Twitter	Submit Report	Team	29	0	
			System Wide Status			
		Q	5607	4232	867	
Mary Deserts	f the second sector 1	Desidence Descentes	Reports	Items	Users	
View Reports	[Unavailable]	Review Reports				
			List by name	• View	Team Members	





New Report:		Points		808			
				Reports	Wins		
(insert web address of offensive content)			Self	32	0		
	NEW Currently supports: • Facebook • Youtube • Twitter Submit Report			29	0		
	Di di	EA	Syst	em Wide St	ide Status		
1 I I I I I I I I I I I I I I I I I I I		5607	4232	867			
View Reports	[Unavailable]	Review Reports	Reports	Items	Users		
view keports	[Unavailable]	Review Reports					
			List by name	• View	Team Members		



Thank you for offering to review items others have reported.

Your assistance with reviews is important as it helps to improve the quality of the data this system collects.

Content to Review

This item is a YouTube Video
View it below, or via <u>this direct link</u>



👚 номе

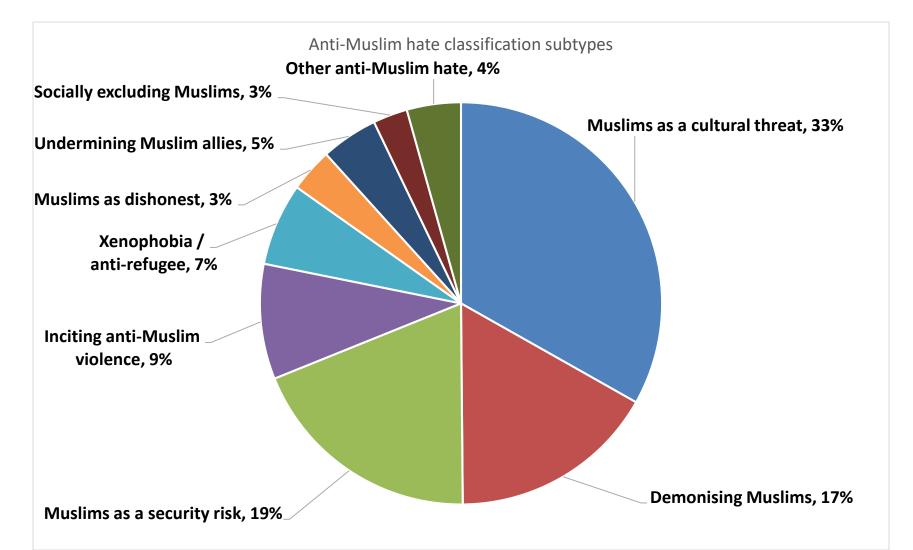
Examples of statistics that become available

- Items reported over time
- Increase in items by social media platform
- Trends over time by hate type
- Time taken to remove items
- Average time to remove item by platform
- Trends in reports by state/country
- Responsiveness by platform

CASE STUDY 3: EMPIRICAL REPORTS

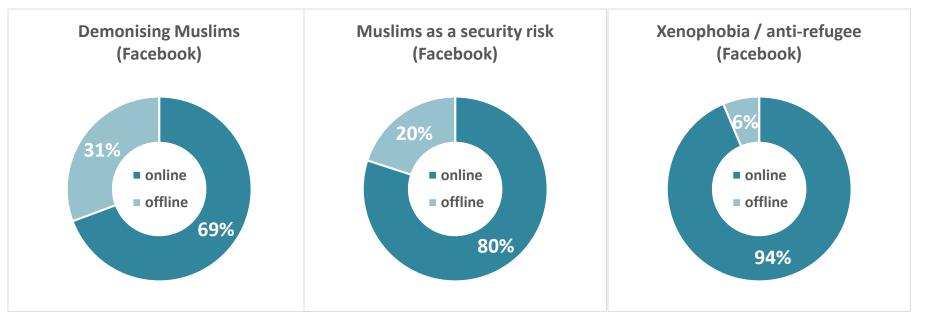
Spotlight on Anti-Muslim Hate Report

Based on a sample of 1111 Items of Anti-Muslim Hate Speech



Spotlight on Anti-Muslim Hate Report

Take down rates so far

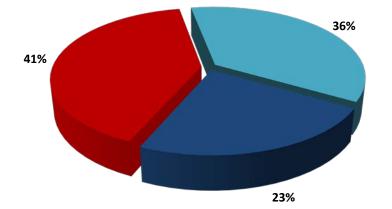


These items have been reported to the platforms through the usual reporting mechanisms. We will be offering senior management the list we are using, and allow them time to review the items, before publishing the final report.

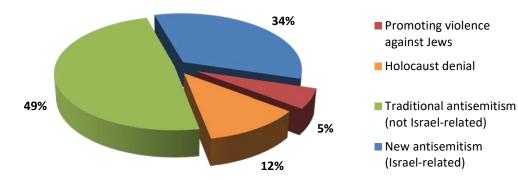
Sample size: 2024 items

Facebook
 YouTube
 Twitter

Antisemitism by social media platform











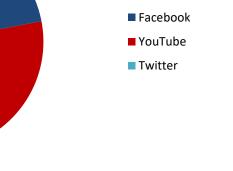


We see that different kinds of Antisemitism are more prevalent on different platforms. Prevalence is a combination of what users upload, and what action the platform is taking to remove such content.

Promoting violence against Jews 16 42 **44** Facebook 27 YouTube 72 Twitter 105 New antisemitism 137 167 120 Facebook 214 YouTube 433 Twitter 253

Holocaust denial Facebook YouTube Twitter

Traditional antisemitism





Facebook

YouTube

Twitter



Measuring the Hate The State of Antisemitism in So



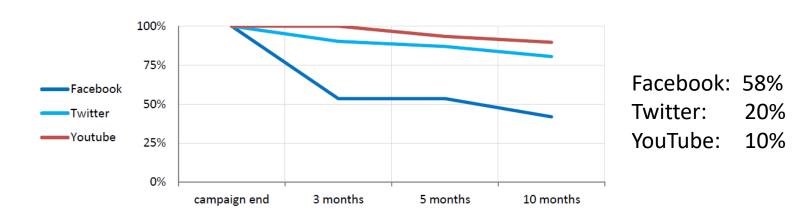
	Global Forum for Combating Antis

Measuring the Hate



	Traditional	New Antisemitism	Holocaust Denial	Violence	Plat. Avg.
Facebook	42%	27%	58%	75%	37%
Twitter	25%	20%	20%	14%	22%
YouTube	9%	4%	10%	30%	8%
Category Avg.	21%	16%	22%	26%	

e.g. Removal of Holocaust Denial



CASE STUDY 4: COMMUNITY AND CAMPAIGNS

The Facebook Page

- Over 24,000 supporters
- A "No Platform" policy banning people from problematic groups and creating a safe space
- Used to encourage engagement in campaigns (not just sharing of information)
 - Briefings
 - Major campaigns

Briefings

Cohesbullying of Mexicon Veisceleb.

Muslim Pacebook page.

Repe Jokes on Pacebook

Reporting "Jew Watch"

romoved.

OHPI published a briefing highlighting a cyberbullying campaign against the Muslim lawyer and community leader Marjam Upiazadob instigated by an anti-

OHPI has called on Pacebook to remove a Public Group on Facebook called "Up, Not Rape If You Hollor Surprise". Repollates have been benned from Pacebook since May 2013, hence, this Group should

OH71 confronted antisemitism on Google4. In a briefing, we discussed a Google Group, its antisomitie character and content and called on our supporters to report this community to Google.

Odagg Roof Support Page removed On June 21, OHPI published a briefing calling for the removal of a white suprematist facebook page set up to support Oxigon Roof, the resist killer of 9 black people in Charleston, US. The page has since been

Freedom of Speech on Facebook In this briefing, OHPI pointed out the hypotrisy of Peobook's claim that it stands for freedom of speech when it explicitly disallows promotion of content on its platform that criticises Pacebook. Such limitations on speech in a virtual monopoly are dangerous.

Antisemitic Pacebook Page removed OHPI reported and get removed an antisemitic Necebook page "Yorkshire: Anti Communista". As we discuss in our briefing, the page was using communism as an excuse to attack Jows in general.

of life

lesed by recebeek.

tatefully presenting Muslims as a threat to our way

This briefing examines the way anti-Muslim groups ercsont local Muslims as a "threat to our way of life". this threat is presented in three ways: the first says that Muslims want Shatia law to replace the law of the land; the second presents. Muslims as a "cultural threat" for not fitting in; and the third presents Muslims as an economic threat, be that as a drain on the welfare system or a threat to jobs. A Mosque at Currumbin

On September 14, OHPI published its latest briefing on a facebook page protesting against approval for a mosque in Currumbin (Gold Coast, Queensland). The page not only incited hatred against Muslims, it was as accking to intimidate the local council and was

Bolden Dawn, Holocaust Denial & Facebool As Grocer, outlawed Holocaust denial, OH71

published a briefing drawing attention to recebook's refusal to accept Holocaust denial as hate speech and to the Greek nee-Nazi party Goldon Dawn's ontinued efforts to gain a feethold in Australia.

				U				
	Holocaust Danial on Facebook In a briding published on October 3, 2014, OH4 draw startsdam i dome thelexast domial page of facebook c_ftelecaust. The Oreat type/bage Daposof, "type/page - baseling the reliceaust" "reliceaust is able list"-which permits compile theories about the Holocaust. Anti-Muslim Hate on Facebook On October 2, 2014, OHP Jublished a detailed	n md		Denmark Attacked In response to the torer estacks in Donmark, OHP released a short statement reminding people of the importance of responsible free speech. It also link to the report we published in the weke of the Cha typkin, estacks.	e ed	v2 taking about th		Cobarth OHPI p campai loador (Muslim
	briding calling stands to a popular Australian page on facebook called "Take Sack Australia". And been monitoring the page for a while but a recent spike in its popularity led us draw public attention to its hateful content.	hate HPT	The rise of Antisemitism	The why of fixing entisemitian OHPI looks at why Jows in particular are largeted binding. It debunks some of the popular theories the subject, and suggests that the reason could be that the Jows are the "original other", at least in the Western culture.		Public Bela	THIS PAGE!	Rape Jo OHPI h Group (Surprise Receber go.
REMOVED	Squadron 88 features Squadron 88 is a recordly formed noo-Nasi group which has been distributing anti-immigrant and antisemitic flices across Sydney. On September 2 OMP published a briefing on its online and offlin neist activities.	7,		Aboriginal Hete on Pacebook CHPI has published a new briefing calling for the removal of a new anti-Aboriginal Facebook Page "displicitle depolycycycop ". The page is promotif dispetited white supermanist theories and ugly storestypes: about Australia's Indigenous communists.	4	····	THIS PAGE A	Report OH71 or briefing characte to repo
EIKERAGE	Hete against cyclists On January J, 2015, OHPI published a briefing commining filterablem of hete against cyclists o social media, as well as the cycychychycag Sector by the cyclist lobby to draw attention to the prob	and Dec.	The grouters of ordere Missogyny (Intel against worker)	Pighting Online Misagary on IWO To mark the informational Womon's Day on March OHP published a binding by the forminat campaig Califon Report about her experience of misaganisti harasament and exployed without no Witten Harmony Day campaign	nor	III. PAGE		Or Jun romoval up to s people romove
Holocaust Memorial Day 2015	Helocaust Mamorial Day To mark the Helocaust Momorial Day on January OHP) published a beeing calling on social moti comparise to add Helocaust denial to their list to banned hats speech. Our briefing discussed how Helocaust denial is accepted as antisomitic hat append by many different genoment agrees around the work. It has been discussional to the		Harmony Day	namony bay campaign To mark Hennoy Day, OHP campaigned to get a entosemile, and-Atualian and and-Abenignal imag removed from Acabeack. The and-Abenignal imag was qpg(d)y removed, but the and somitie and and Atualian images are still online. OHPI gats hate removed				Freedo In this Facebox when it its plat on spec
MODERN	guipagge, sadomia and motia. Unfortunatdy, social motia has provided self-styled toleosust revisionista a platform to take such therein to the public without any sheaks or filters. Social motia platforms should put a top to this. Note of Noteen Antisamilian Our breding highlighted numerous comments m	e efe	Removed 3 Twitter Users 4 Accebok Pages context included monory, watersmaan and anti-indigenous join us at facebook, com/uninehale	In late March we had four facebook pages and the Twitter handles removed that tagefler premoted anti-aboriginal hate, misogony and antisomitism. Online Rants & Free Search		UIII PAGE	REMØVED III	Antiaen OHPI ro Necebor discuss commu
ANTISEMITISM	on OHM's (ppg)rapp age following the publicate a briefing on Holeaust double, the commons is how the small-alexane conflicts is being used to shudlown any legitimatic concerns arised by Jow with respect to aning antisomation. To conflicts actions in Alabains to the Nais government poli- and Holeaust is a part of Hammi's oxial modia a Mategy to turn public opinion against Issuel.	ow nacli		What should from the basis of judgement when deading whether an online and constitutes a minimal directive discuss the the same the perspective of both US law and Australian law.			A MERCARD COMMUNIC	
	poning Momophobic Azerbook page 91 is curred y vanning a campaign to report a moghobic haddosk page "Mexistance Ageinst famite (ageiga)".	Calls to Kill on Facebook	Lens to Auri on response OHPI successfully campaigned for the re- hatebook hyperger, "Kill badge Goldon", called for the mundler of the captured isn I twent on the calibratic his death and po- antisemitic content.	The page ARENT GAY.	"Stop the Straight Hate" (104 "Stop the Straight Hate" (104 The page was started on July genered 425 likes. It was sta Supreme Court ruling on Jun samoos marriage in the US. of the page are Australian.	nophobic facebook page 716487785164679). 4, 2015 and has rediscon after the US 26, 2015 that legalised Many of the supporters	Insert Home of	Internet
REPORT THIS PAGE A A REPORT THIS PAGE A De De De De De De De De De De	poning "Guya Should Ba Fit & Muscular Apagh" 91 is also calling its supported to report another montholis, looky having page "Gass should be fit i muscular Apaghg". Anad more here and report page	Comparison with the Nazis	Israel-Vasi Compation This OHY's briefing optians why it is be and offonistic compare the small gov the fact party. This comparison has been popularity is solid motifs, depice being recognised as a form of antisomiliam.	mmont to paining	The dange of the Austellian "Austalia officially now has modellad on the fa-right wir surger". That's the opening: yidney Moming Headle last Austalia, however, has a twi- nethania, and the mevemen- right groups from oversas li bewn, but there is also a con- multiculuraism in parts of A	a political party og movements in of an article in the unday. The favright in st. There are the overt is supported by fav- ke Greece's Goldon cession to	G 05	ED .
Alamantalia on the Antonia	ling out popular anti-Maalim myths a briefing, we tackle the popular anti-Mualim mont that somehav their presence in Australia "Onest to our way of life". We also share a video which an Austrian talmic schlere oplans the toevenial concepts ξξησία and ζημά.	Antisemitism: The Jews Control the Media Canard	Javes Control the Media Canad OHP1 debunks the myth that "Jows Conto Media". Vecylain the origins of this or theory and provide references on the top optimation net time someone you know the "Jows control the media" line	rapiney ie. Use our	The same "And any Grand as for the	e flog of the Year" omc of the arguments lem goggig, pages to tent and initialised ney Swam player. We context within which the ght of this, many	Disli	ike

Major campaigns

- Other organisations invited to partner on these
- The aim is to activate the public and collect data
- The end result is an empirical report



SAMIH: Spotlight on Anti-Muslim Internet Hate

About the Campaign

Current Partners

Australian Federal Police



Current Supporters

Multicultural Communities Council of Illawarra



Faith Matters

Cyberspace Law and Policy Community

Federation of Ethnic Communities' Councils of Australia (FECCA)

All Together Now

The Agency La Trobe University



CYBERSPACE LAW & POLICY COMMUNITY

Telimama Measuring Anti-Muslim Attacks





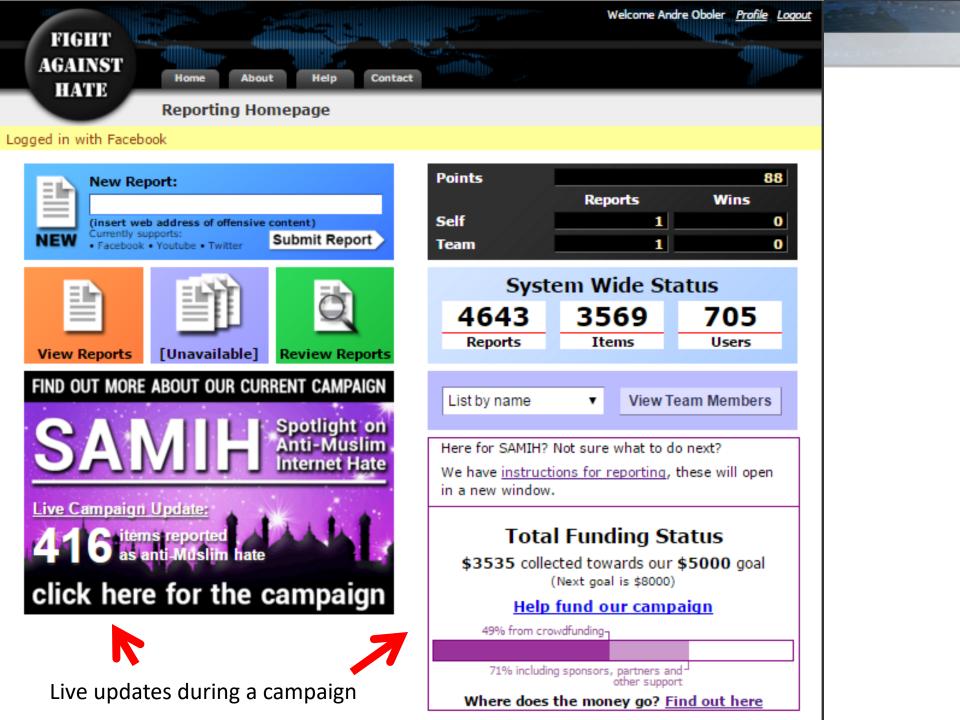




Islamic Council of Victoria



ISLAMIC COUNCIL OF VICTORIA



CASE STUDY 4: LIVE REPORTING

Live reporting

- We often attend events and share pictures and video live via our page
- This allows more people to feel a part of these positive activities
- It directly undermines the messages of hate and helps to improve resilience
- It leverages both our online capability and other organisations real world ability to convene in person
- Facebook: <u>www.facebook.com/onlinehate</u>
- Twitter: @onlinehate

MEASURING IMPACT

- Fight Against Hate
 - Launched by Paul Fletcher MP on behalf of then federal Communications Minister Malcolm Turnbull MP
 - Cited in 2 UNESCO reports as an innovative tool to tackle hate speech
 - Presented at the UN in NY, to international bodies, endorsed by the GFCA
 - Successfully used to create the two major empirical reports (antisemitism and anti-Muslim hate)

• Facebook group

- 24,000 is a large Facebook group, about 25% the size of the Australian Human Rights Commission's page (despite huge differences in budget, staffing & government status)
- Some posts have very high levels of engagement
- Reports
 - Cited in parliamentary reports, UNESCO reports, a recent Special Rapporteur's report to the UN Human Rights Commission
 - Changes to software of social media companies based on recommendations in our reports

• Briefings

- A significant number of the items we focus on come down
- Good engagement numbers

Year	Number of	Total Likes /
	Briefings	Shares
2014	32	8,800
2015	45	29,800
2016	84	39,217

The real impact

- By tackling all forms of hate through one organisation and in one community it shows that the affect group is not separate from society but part of it
 - Increases targeted groups resilience by undermining the messages of hate
 - Our message is they are welcome and we will stand with them
 - While they may be the targeted minority now, they are invited to be part of the majority in standing up with against hate targeting others
 - Increases individuals resilience by giving them reporting as an empowering and practical action
 - Empowers people to support different groups, strengthening community cohesion when it is needed most

0

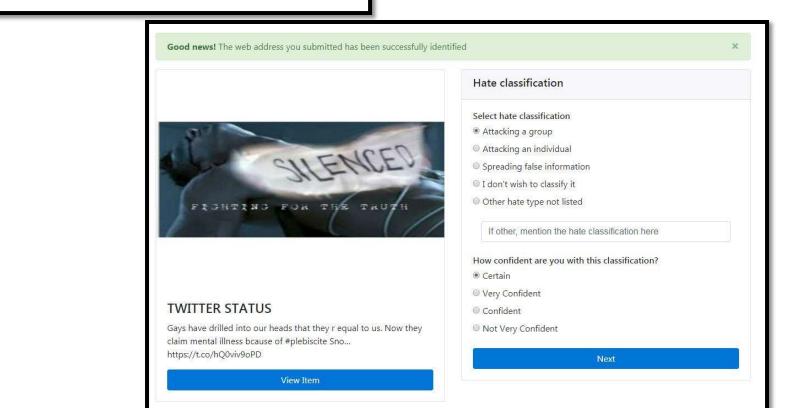
THE NEXT STEP....

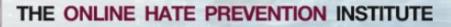
THE ONLINE HATE PREVENTION INSTITUTE

aste the web address of the conten	Submit Web Address	
aste the web address of the conten	Submit Web Addres	

The new software (released early 2018)

- 1. A reporting form embedded on organisation's own websites
- 2. Enabling more people to gather data from the public & see what they collect





CSI-(Crowd Source Cyber Hate	ed Intelligen	се			Q ingenization Logo
Home Hate Items - Incoming Items Filter By:		Administration +		Test2 Us	er Help Log Out
2 sheeled	Eiltor				
3 checked - Folders	Filter	Hate Item	Hate Type(s)	Report Details	Categorize
		Hate Item Image:10157615962925548 f	Hate Type(s) antisemitismDenial (80%)	Report Details Reports: 1 Last Reported: 08 Nov 2016 First Reported: 08 Nov 2016	Categorize

3. An advanced analysis tool connects to this data and enables data sharing as well as working with the data

- (a) Incident Reports
- (b) Trend analysis
- (c) Categorization
- (d) Annotation

human rights agencies, police and community groups	CSI-CHAT Crowd Sourced Intelligence Cyber Hate and Threats		(ingerization L
CSI-CHAT Crowd Sourced Intelligence		inistration -	Test2 User Help Log Ou
Cyber Hate and Threats	Test case file		
Home Hate Items • Reporting • Administration • Incoming breakdown from 01-01-2016 to 01-01-2017	Report Note:	Items: antimuslim antisemitism antimulsim Image:10157615962925548 antisemitismDenial	Add
28.9%	Items Added:	URL: https://www.facebook.com/10157615962925548 Description:	Note:
Title Save Report Title		No Description found.	
© Team=MC ² 2016	© Team=MC ² 2016	Save	

The challenge... working together

- The problem of Hate 2.0 is growing.
- It is destroying people's resilience and that of communities and it is ripping the social cohesion of society apart.
- We need to stand together to tackle this toxic environment.
- We need to overcome the technological barriers.
- We need to unite bringing all of our strengths together, sharing best practice and enabling others.
- Only then will we see toxic tide turn.

References

- [1] <u>Sensis Social Media Report 2017</u>, released 22 June 2017 p 10.
- [2] Andre Oboler, "<u>Online antisemitism 2.0. 'Social antisemitism'</u> <u>on the 'Social Web'</u>", *Jerusalem Center for Public Affairs*, number 67 (1 April 2008).
- [3] Osman Faruqi, "Analysing MangoGate: How A Year 12 Facebook Group Became Ground Zero For Harassment And Abuse", Junkee, 19 October 2017.
- [4] Andre Oboler, "The normalisation of Islamophobia through social media: Facebook" in Islamophobia in Cyberspace: Hate Crimes Go Viral, Routledge, 2016, p 45.
- [5] Jeremy Waldron, The Harm in Hate Speech (Harvard University Press, 2012) 2—3.

Contact details

- Website: ohpi.org.au
- Twitter: @oboler / @onlinehate
- Facebook: facebook.com/onlinehate
- E-mail: <u>ceo@ohpi.org.au</u>