Through These Walls

Impartial dispute resolution of online harm during a global pandemic

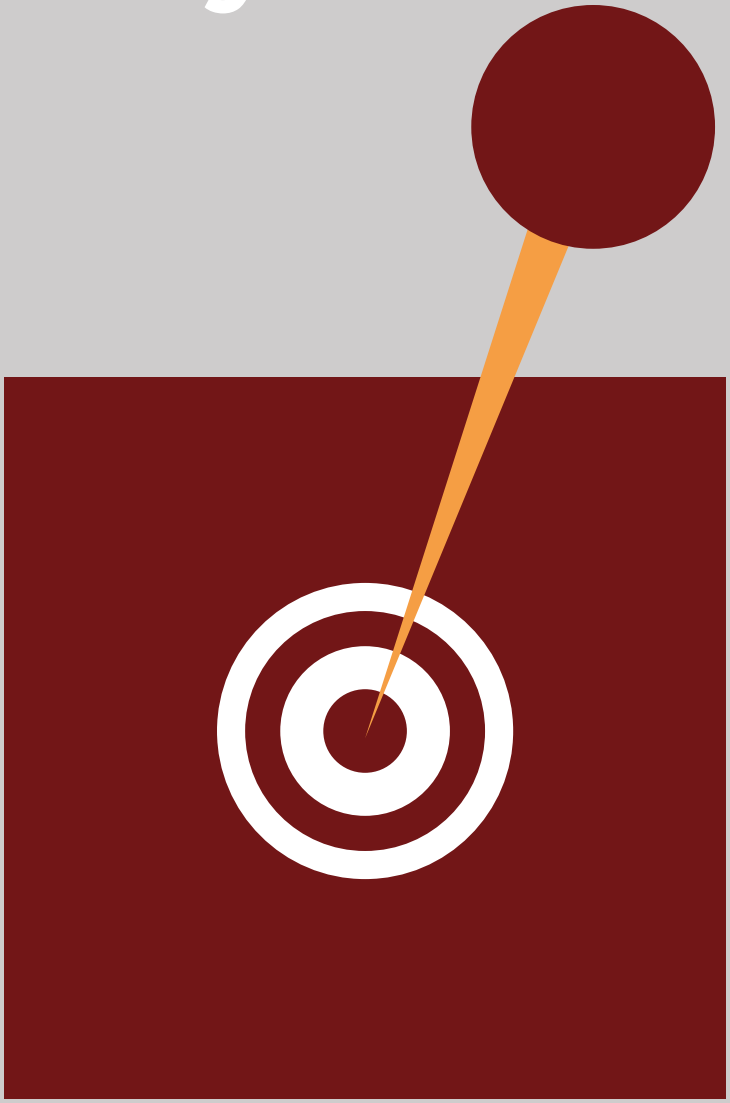
Prepared by
Francesca Farmer and Kathryn Tremlett

REPORT HARMFUL CONTENT
reportharmfulcontent.com

Contents

Table of figures

Executive Summary



The extent and significant impact of harmful content online on individuals and groups is something that society are reminded of on a daily basis. Whilst much of this relates to content that is considered illegal, it is clear that legal content can be equally harmful. Reporting and removing illegal content is relatively straightforward relying on legislation to direct actions. For users, reporting harmful but legal content is more complex, often requiring navigation of community standards and reporting routes and relying on industry platforms to take action.

Report Harmful Content (RHC) is an impartial dispute resolution service that supports users and platforms in reporting legal but harmful online content.

RHC considers eight types of legal but harmful online content spanning; abuse, bullying and harassment, threats, impersonation, unwanted sexual advances, violent content, self-harm/suicide content, and pornographic content.

It does this by providing up-to-date information on community standards and direct links to the correct reporting facilities across engaged platforms. Additionally RHC extends an impartial dispute resolution role for users who have already submitted a report to a platform and would like their outcome reviewed or escalated.

This unique dispute resolution procedure extends users with redress whilst at the same time building trust and confidence in platforms.

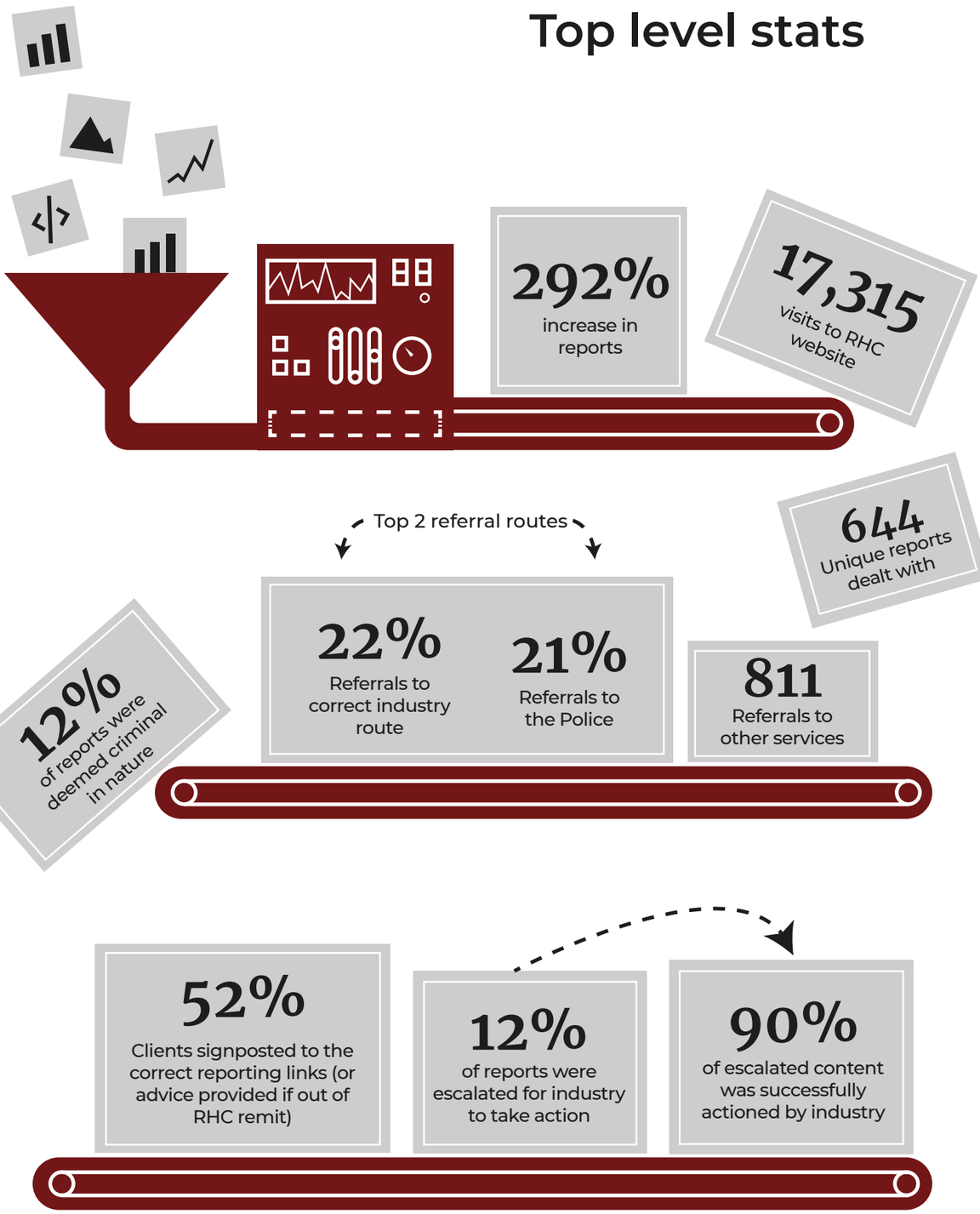
RHC combines a deep understanding of community standards alongside trusted flagger relationships enabling it to effect dispute resolution. RHC also offers advice on additional issues faced online, signposting to other support services and the police when necessary.

RHC was formally launched in December 2019 and the first annual report was published in May 2020.

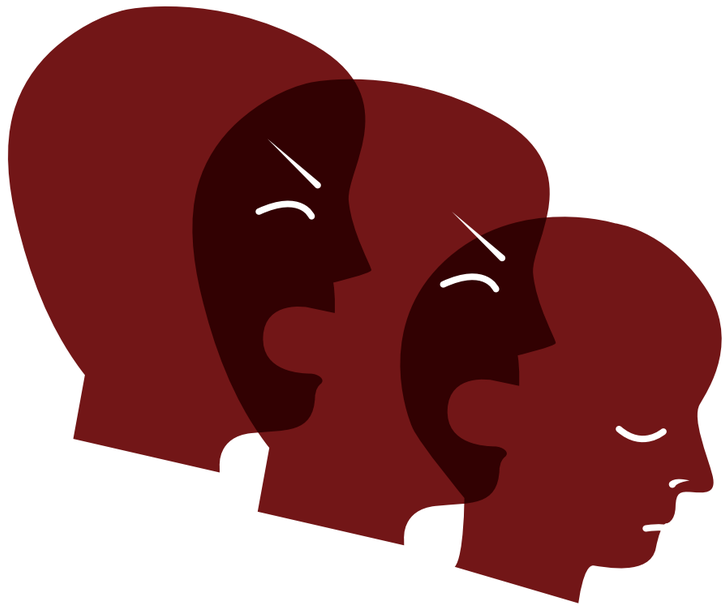
One year on, the second annual report 'Through These Walls' presents results of mixed-methods research carried out on all reports RHC managed within the first full year of public operation (January 2020–December 2020).

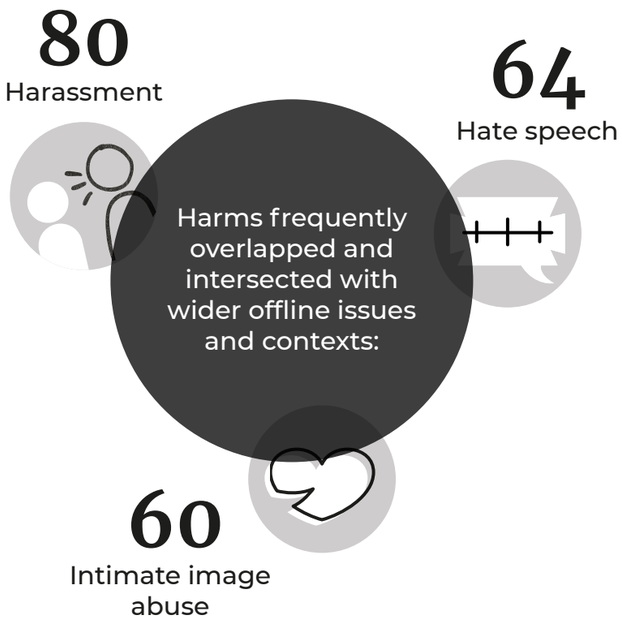
In the year analysed...

Top level stats



Main online harms reported:





Three common trends:

Cluster of domestic abuse, coercive control and harassment issues:

75%
of victims knew the perpetrator

41%
reports raised by those experiencing the abuse

75%
Reported by (or on behalf of) women

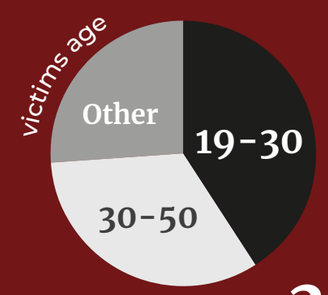
59%
reports raised by loved one or friend

91%
reports had additional harm associated

reports commonly took the form of a privacy violation with an ex-partner or an ex-partner's current partner hacking an old social media account and posting private images

15%
of reports led to mental health issues

33%
of reports had an element of cultural and/or religious sensitivity



35%
of reports RHC referred to the Police

25%
of reports Intimate Image Abuse was an additional issue

Rise in hate speech:

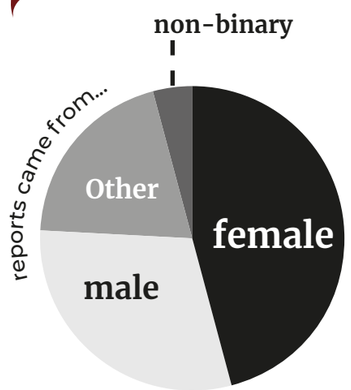
255%
rise in reports

Reports commonly intersected with a primary report of:

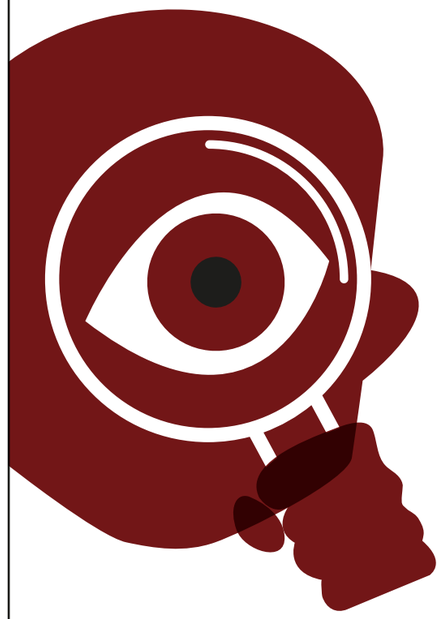
23 harassment

19 abuse

75%
of those reporting were bystanders or observers



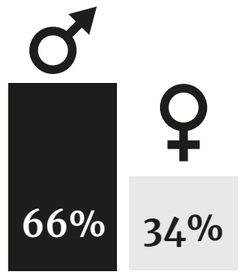
Young males actively searching for harmful content and reporting it:



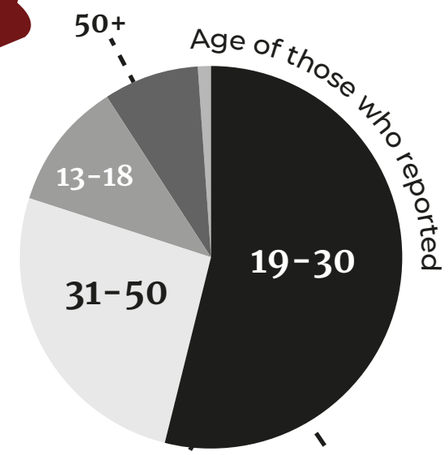
Pornography was the only harm that was predominantly reported by males

Of all of the 'other sites' not within the remit of RHC, the majority were reported for pornographic content

(25%)



Men were twice as likely to be searching for harmful pornographic content than women



This age group was also the most likely to be actively searching for harmful content

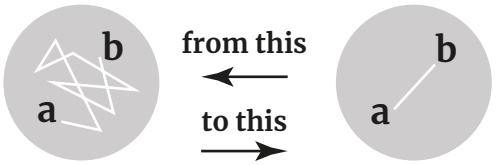
(62%)

67%
of this age group were male

Key recommendations

Continued development of RHC:

- Sustainable funding of the service
- Out of hours support
- Increased capacity to provide more tailored support for clients
- New partnerships developed.



Industry action

- Employ impartial dispute resolution provider
- Address capacity issues with moderation resulting from Covid-19
- Streamline reporting flows
- Faster responses to escalations
- Decrease disparity in reports for non-users and users.

Further research

- Actively searching for harmful content
- Impact of online harm during a global pandemic.

Q&A

Introduction & background

Report Harmful Content (RHC) is a national impartial dispute resolution service that has been designed to assist everyone with reporting harmful content online. RHC is provided by UK Safer Internet Centre and operated by SWGfL. The service grew out of SWGfL's previous experience running the Professionals Online Safety Helpline and the Revenge Porn Helpline. Whilst these services offer essential support to members of the children's workforce and adults experiencing intimate image abuse, respectively, certain elements of online safety provision were identified, with which neither of these helplines could assist.

RHC was designed to fill that gap. It empowers anyone who has come across harmful, but not necessarily criminal, content online to report it by providing up-to-date information on community standards and direct links to the correct reporting facilities across multiple platforms. The service also provides further support to clients based in the UK, over the age of 13, who have already submitted a report to industry and would like outcomes reviewed. RHC is able to act in this mediatory dispute resolution role with a number of industry platforms, with whom it has a trusted flagger partnership and their reporting flows integrated into the RHC website. These platforms include: Facebook, Instagram, Snapchat, Twitter, Roblox, TikTok, Discord, Twitch, Match Group (which includes Match, OK Cupid, Ourtime, Tinder, PoF and Twoo), Microsoft (which includes LinkedIn, Bing, Xbox, Skype and Minecraft) and Google (which includes YouTube, YouTube Kids, Google Search and Blogger). All dispute resolution offered by RHC is provided to clients via email contact.

The term 'harmful content' can be very subjective. In order to remove ambiguity, specialist online safety practitioners studied the community guidelines of several different industry platforms. They found that eight areas of content are likely to violate platform terms: abuse, bullying and harassment, threats, impersonation, unwanted sexual advances, violent content, self-harm/suicide content, and pornographic content. RHC practitioners offer impartial dispute resolution associated with these eight types of online harm. They also offer advice on further issues faced online and signpost to support services and the police when necessary.

RHC launched publicly in December 2019 after operating for 12 months in a beta-phase. In order to gain a greater understanding of harmful content online and continue to improve the service, mixed-methods research was carried out on all cases dealt with in the first full public year of operation (January 2020–December 2020). This research builds on the RHC Annual Report 2020 (reviewing data from the beta phase: January 2019–December 2019) launched in May 2020.

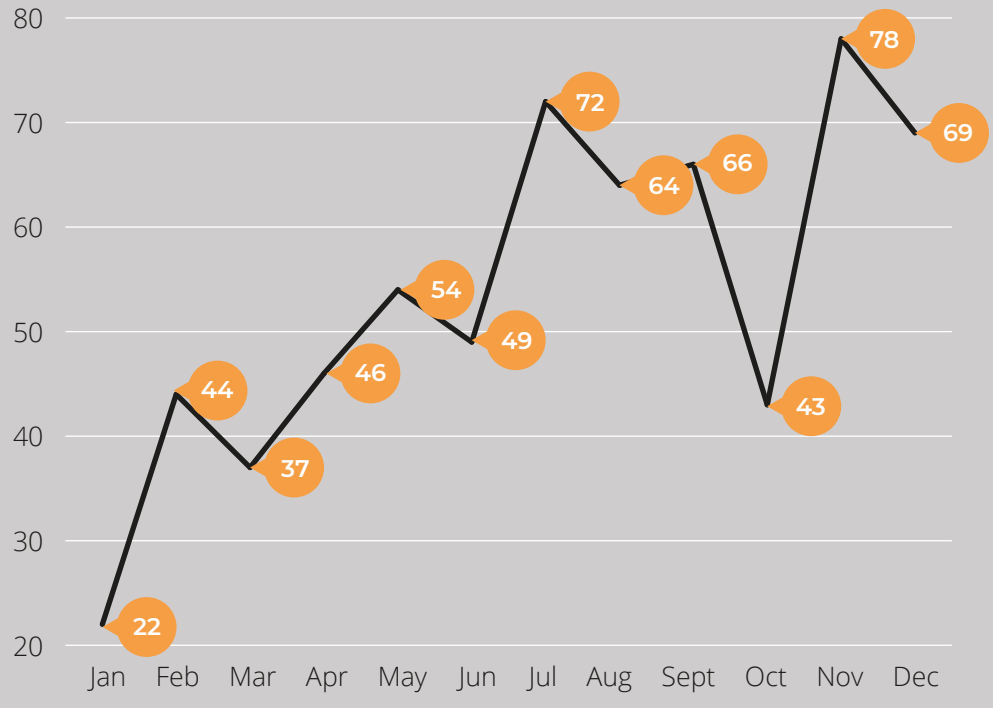
This report begins by presenting top-level statistics, it then moves on to discuss cases in more depth, outlining emerging trends and issues and concludes by outlining recommendations for the future development and growth of the service. As mentioned above, RHC works in trusted flagger partnerships with a number of industry platforms. It also works closely with government departments, both in terms of designing the service and providing consultation on new policy.
Due to the complex nature of online harms and their impacts, the service also maintains relationships with, and makes referrals to, other support agencies, charities, the police and social services. This report has been designed with all of these parties in mind, in the interests of information sharing for best practice. More broadly, this report will also be of interest to academics, researchers, journalists and others with an occupational interest in online safety.

Top level statistics

1. Helpline use and growth

During 2020, RHC practitioners dealt with 644 unique reports. This is a 292% increase on the previous year which had 164 reports. The average amount of contacts per report in 2020 was 2.26 with 274 pieces of content being escalated with industry and 246 pieces of content actioned. Figure 1 shows the volume of reports per month. The service increases steadily in popularity throughout the year with a slight dip in September. The increases correspond to certain promotional activities. The increase in January 2020 corresponds to the launch of RHC in Dec 2019. The increase in June corresponds to the launch of the 2019 RHC annual report. Finally, the steep increase in October can be explained by the launch of a university campaign in September and October.

Figure 1: Volume of reports per month



2. Dispute resolution: RHC response

RHC practitioners respond to reports in 12 ways which increased from the four ways reports were responded to in 2019. These responses include the reported online harm not being in remit but with the reporter being signposted to relevant advice, criminal matters, content escalated to industry and the reporter not being based in the UK. These responses are not mutually exclusive and are dependent upon various factors including nature of harm, location of content, age of client, whether the client is based in the UK, potential criminality of content and previous reporting channels pursued.

1. Escalated content with industry:

Where content is (1) deemed to fall under the definition of an online harm, (2) located on one of the industry platforms with which RHC works in partnership and (3) has already been reported unsuccessfully to the platform by the client, practitioners escalate reports for review directly with industry platform contacts via trusted flagger routes. The client must be over the age of 13 and UK based for this route to be pursued.

2. Online Harm not in remit – signposted to relevant advice:

Some reports fall outside the remit of the RHC project such as content hosted on non-partnership platforms. In these instances, practitioners provide clients with clarification as to the correct nature of their issue and direct them to more appropriate sources of support. This includes the subheadings:
Not considered harmful: This response

- Not considered harmful: This response was used when the report was not considered harmful per the definition of harmful content used by RHC.
- Offline harm – signposted to relevant advice: This response was used when the harm occurred offline rather than online. In these cases, the client was given details of relevant organisations and advice.
- Not based in UK: This response was used for clients located outside of the UK.
- Agreed with industry: This response was used when RHC agreed with the initial response from the platform.
- Content already actioned: This response was used if the content reported had already been actioned by industry.
- Redirected to industry: This response was used when the conditions of RHC were met with the exception that the client has not already made a report to the industry platform. In these cases, practitioners directed clients to the correct reporting links and encouraged them to re-report to RHC should industry reports be unsuccessful.

3. Repeat reporter:

Where an individual reports multiple cases to RHC

- Action taken – still escalated to industry: Where the same client continues to report any issues that are similar in nature and in the project scope. The client continues to adhere to our processes; submitting the correct reports to industry in the first instance and therefore we can continue to help escalate content.
- No further contact: Where the same client continues to report the same issue but becomes hostile/abusive/harassing in nature, refusing to cooperate with the process we have outlined. In these cases, practitioners follow our zero-tolerance policy.
- Client not cooperative: Where the client does not adhere to requests to follow our processes, in particular, where we require them to report the content to the platform concerned in the correct way before acting in a mediatory capacity.

4. Criminal matter:

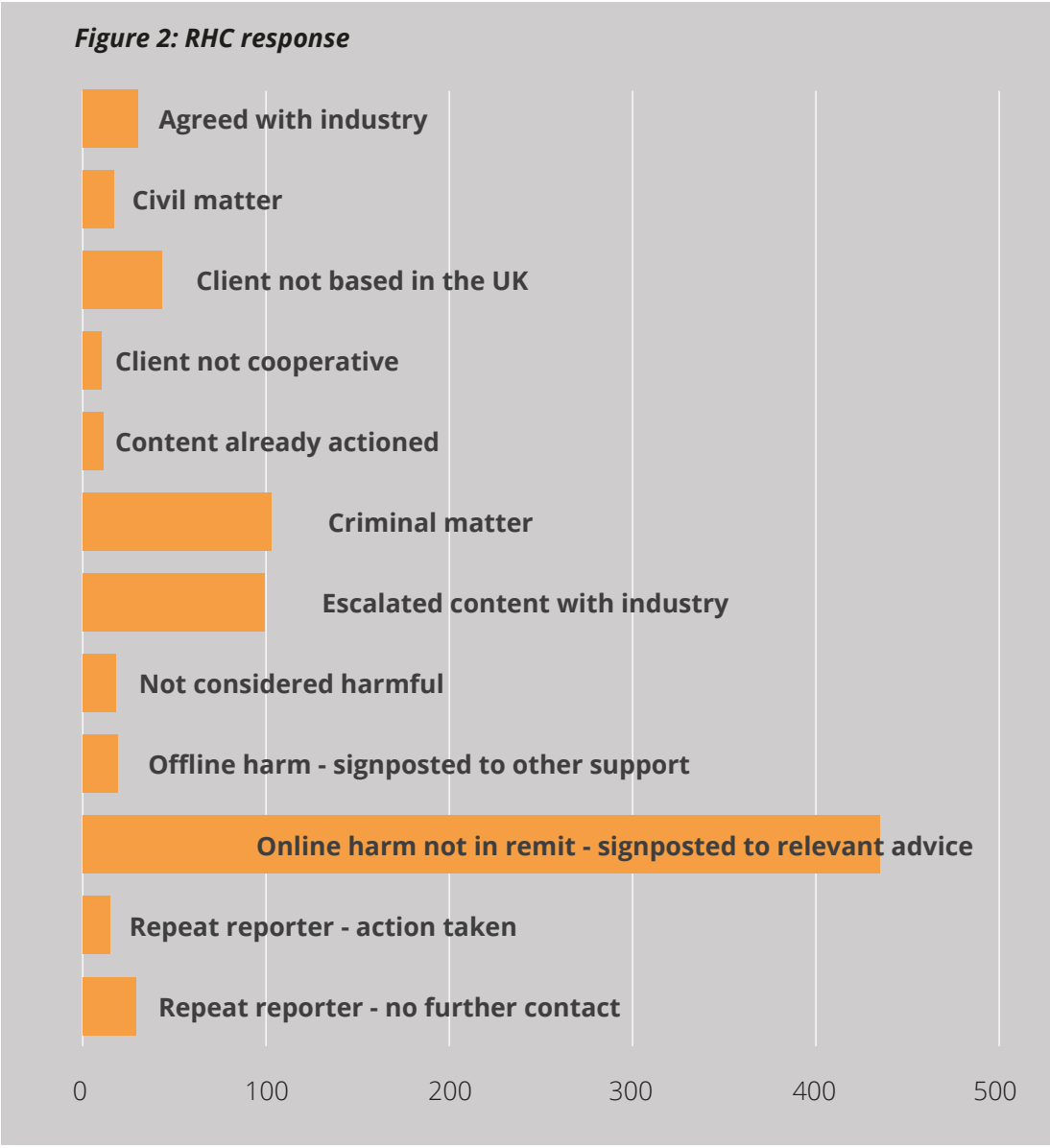
- Where the issue reported is clearly criminal in nature, practitioners direct clients to the appropriate law enforcement bodies (e.g. the police, True Vision, the NCA Child Exploitation and Online Protection command (CEOP) and the Internet Watch Foundation (IWF)).

5. Civil matter:

- Where the issue reported may or may not be criminal in nature and legal advice may be beneficial, practitioners signpost clients to legal information (e.g. rights of women and the SPITE project, etc.)

Figure 2 shows the breakdown in service response with 'online harm not in remit, signposted to relevant response' being the response given most to reporters (52%), followed by 'criminal matter' (12%) and 'Escalated content with industry' (12%). Of the 435 offered this response, 30% of these (131 reports) offered at least one additional response. The response that overlapped with this most frequently was 'criminal matter' (52%) and as a result, the reporter was directed to contact the Police.

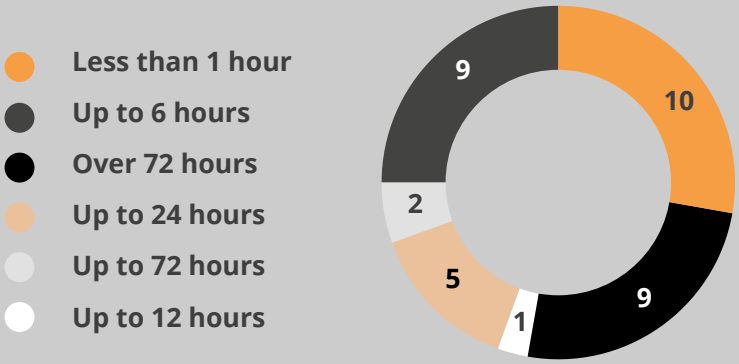
Figure 2: RHC response



3. Dispute Resolution: Industry Response:

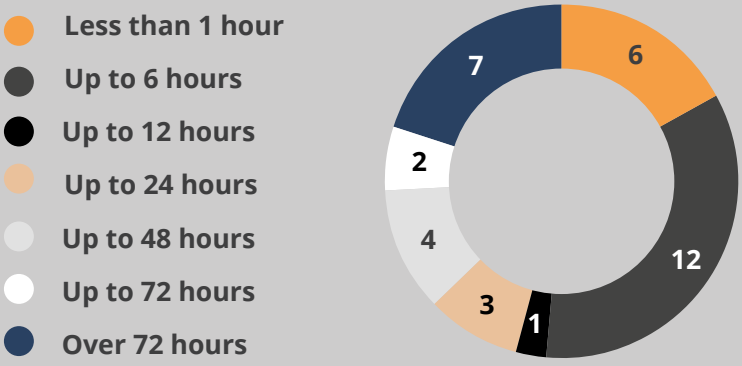
Once RHC identified that the correct reporting route had been used by clients but no action taken by industry, they would escalate content with industry for actioning. In 90% of reports content was actioned successfully (i.e. removed, restricted, regained access to). Twenty-four per cent of reports had content removed within six hours of the content being escalated with partner organisations followed by 23% of reports which took over 72 hours for content to be removed. Only six cases (5%) took up to 72 hours for content to be removed while four cases were not actioned by partner organisations.

Figure 3: Actioned content on Facebook



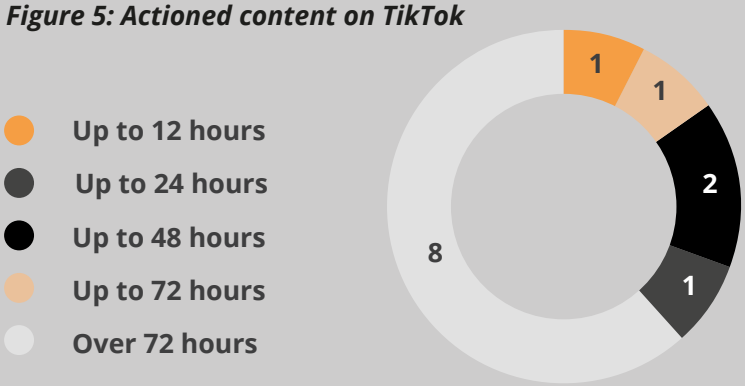
Facebook was the organisation that had the most content reported (36 reports). The platform took down 28% content within an hour of the content being reported (Figure 3). Yet, 25% of content took over 72 hours to be actioned.

Figure 4: Actioned content on Instagram



The second most reported platform, Instagram, had 35 reports where harmful content was identified and actioned (Figure 4). The majority of these reports were actioned within six hours (38%) followed by 19% of content taking over 72 hours to be actioned.

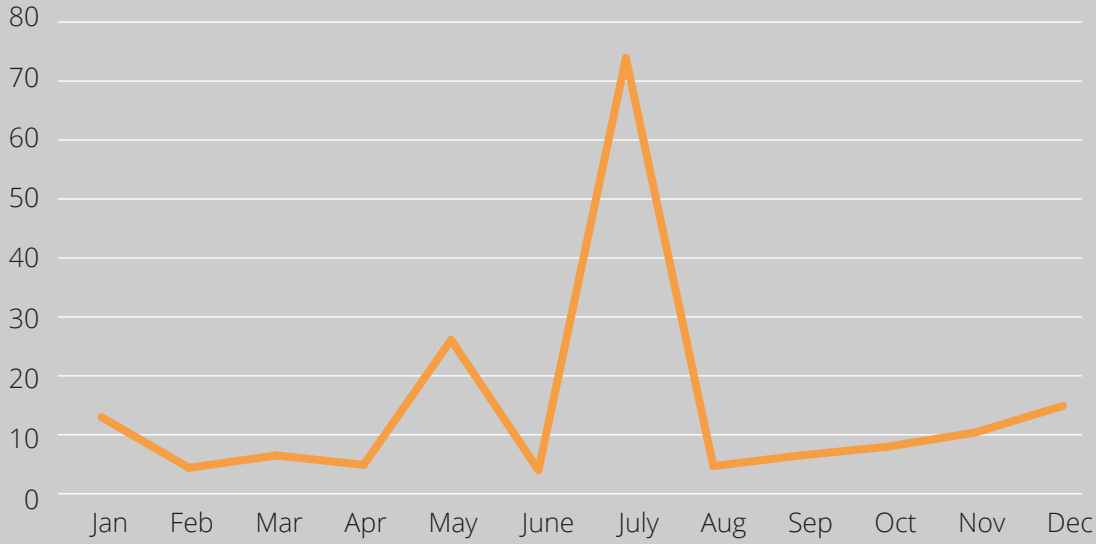
Figure 5: Actioned content on TikTok



Of all of the platforms, TikTok, on average, took over 72 hours the majority of the time (58%). The shortest amount of time for content to be actioned on TikTok was up to 12 hours (Figure 5).

Throughout 2020 it took partner organisations an average of 14–15 days to action content escalated by RHC. Figure 6 shows the average number of days taken for industry to action content on a month-by-month basis. The peaks in number of days taken to action content in May (26 days) and July (74 days) correlate with the first national lockdown: the content was first escalated with industry partners in the month of April. Many industry platforms experienced delays with moderation processes at the beginning of the pandemic as they adapted to comply with government guidance. For many moderators, this meant operating on rota basis and at a reduced capacity, inevitably impacting response times. From August onwards there has been a steady rise in response times.

Figure 6: Average number of days taken for industry to action content



4. Nature of reports:

Out of the main eight online harms, reports involving bullying and harassment were most common (193 reports). This was followed by pornographic content (141 reports), abuse (94 reports), impersonation (85 reports), violent content (71 reports), self-harm/suicide (32 reports), direct threats (31 reports), and unwanted sexual advances (21 reports). Figure 7 shows the proportion of each type of harm against the overall harms recorded.

Figure 7: Proportion of harm by type

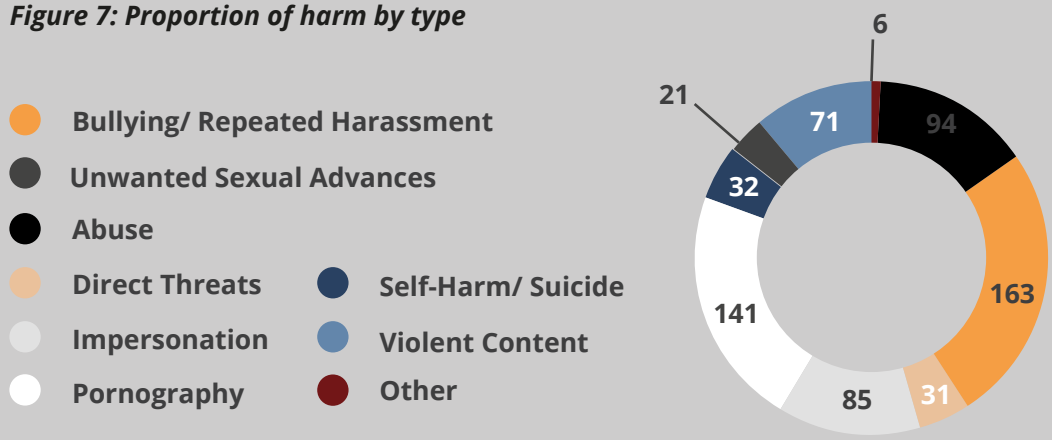


Figure 8 depicts the year-on-year changes in the different harms reflecting the similarities in 2019 and 2020.

Figure 9 shows type of harms reported by month.

Figure 8: Harms by type 2019/2020

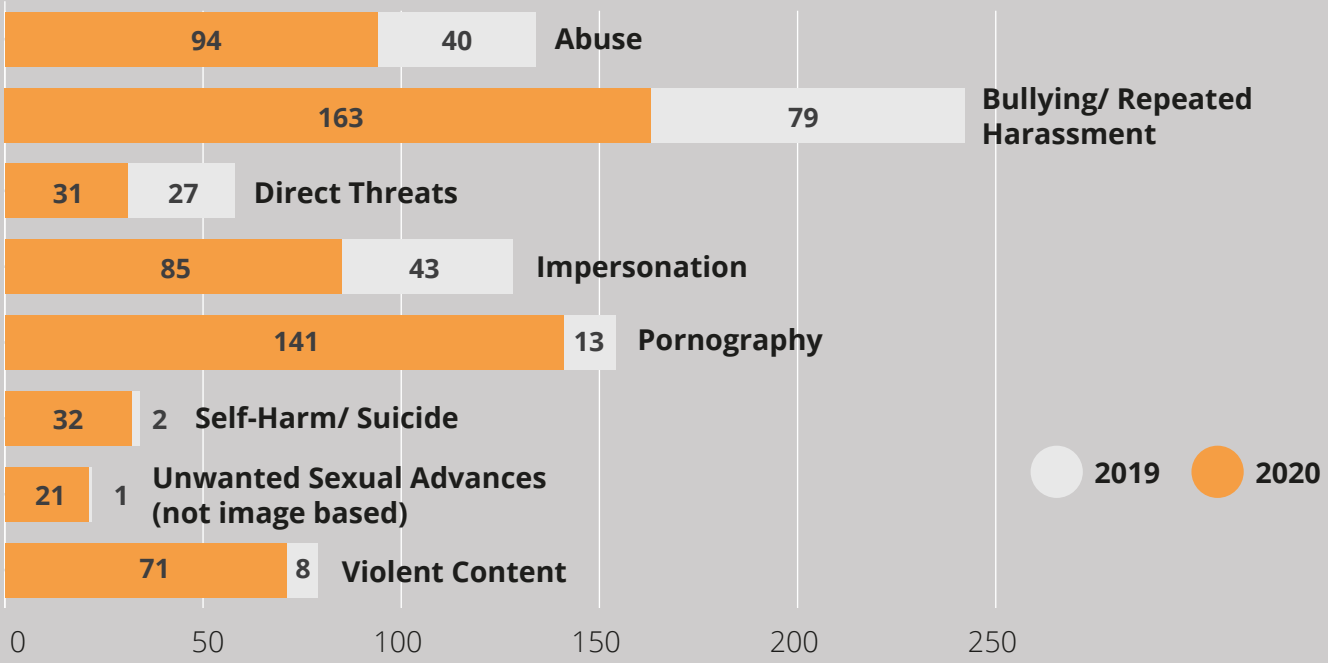
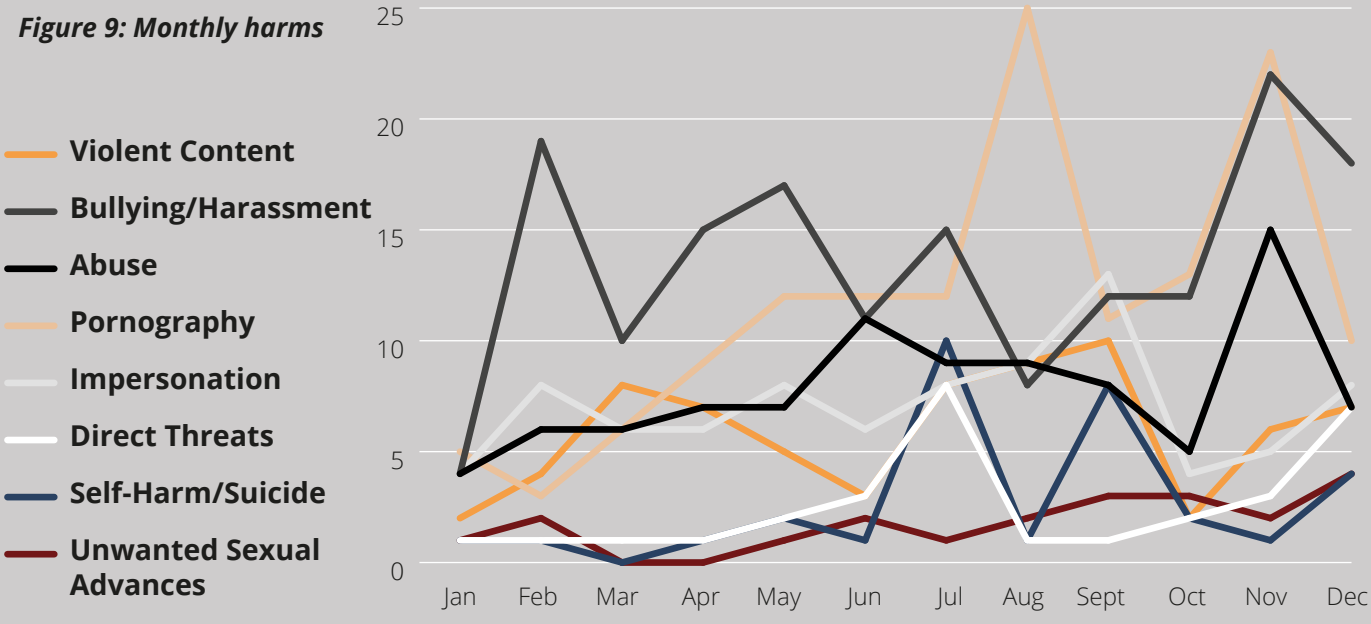


Figure 9: Monthly harms



Oftentimes, reports made to the service not only concerned online harms, but wider offline issues and contexts. Figure 10 shows the wide range, and the proportion of wider issues. The most common associated offline issue/context was harassment (80 cases), followed by hate speech (64 cases) and intimate image abuse (60 cases).

5. Referral routes:

Fifty-two per cent of clients were signposted to relevant advice and support services in 2020. Clients were referred to correct reporting links for industry platforms or different services 811 times. This is due to the fact that a single ticket may have multiple referral routes. The most referred to service in 2020 was the correct reporting link for partner industry platforms (22% of referrals) followed by the police (21%). Figure 11 shows the frequency breakdown for the way in which clients were helped.

Figure 10: Wider issues by type

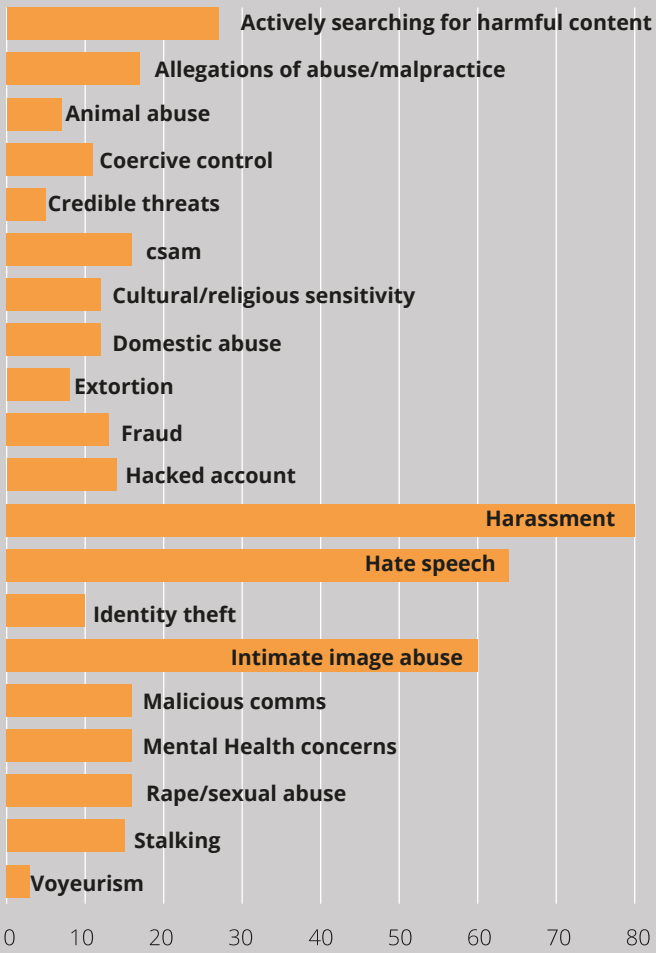
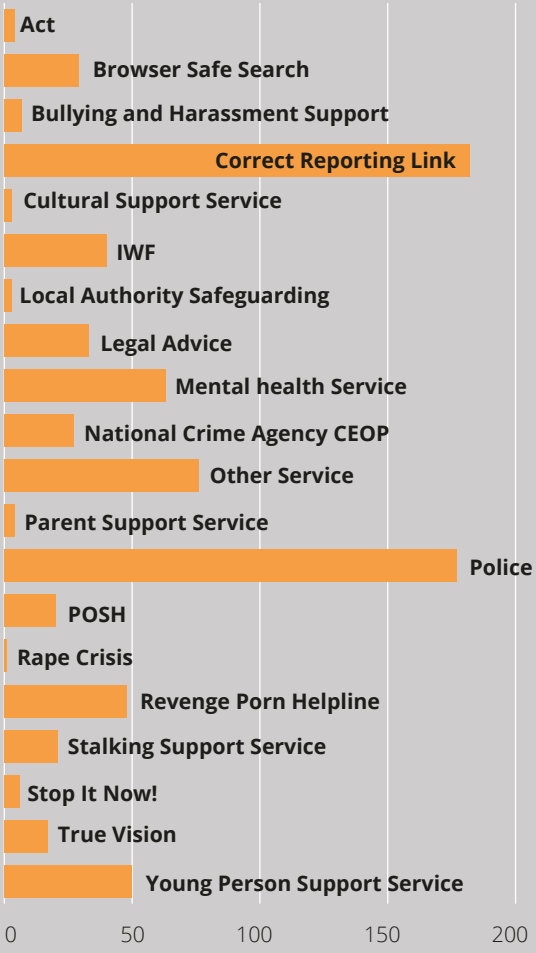


Figure 11: Referred to services



6. How were clients helped?

In addition to escalating content to industry and redirecting to correct industry reporting links (as discussed above under 'RHC response'), practitioners were able to assist clients by providing them with information and/or clarification on the nature of their issue or redirecting them to reporting links on other (non-industry) sites. They also offered emotional support, alongside signposting clients to other agencies and services, for either emotional or practical support (as discussed above under 'referral routes').

Qualitative data also offers insight into the way in which clients were helped. Client testimonials (either communicated directly to practitioners or through follow-up surveys) revealed the positive impact of RHC's assistance.

Clients remarked on the quality of support given by practitioners making comments such as:

> *I found the ladies at the online safety helpline extremely helpful and respectful. They didn't judge me, they knew my family needed help and support and they gave their time and commitment to ensure the videos were removed promptly.*

> *Incredibly grateful for the support, felt listened to when previously was ignored by [industry platform]*

> *Continue doing the good work because it's really been very useful to know you are there to help.*

> *RHC is greatly appreciated for their dedication they provided to customers. This company is making a difference. I hope this company continues to help keep the internet safer for children. We are their voices.*

Clients also remarked on the likelihood of using the service again if another issue occurs:

> *RHC is a fantastic resource, it's great to have somewhere to go if just reporting doesn't work.*

> *I would definitely come to you guys again if I have an issue in the future*

7. Client demographics

RHC collects basic demographic information from clients (age and gender), alongside recording the location of the harmful content. The gender of RHC clients was predominantly female (53%), followed by male (38%) with 142 reports from female clients as compared to 244 reports from male clients. The service also had eight reports from non-binary individuals. Figure 12 shows the gender of clients, represented as a proportion of total reports.

The age group most likely to report to RHC was 19–30 (268 reports) as was the case in 2019, closely followed by 31–50 (209 reports). Figure 13 shows the age group of clients, represented as a proportion of total cases.

Figure 12: Clients by gender

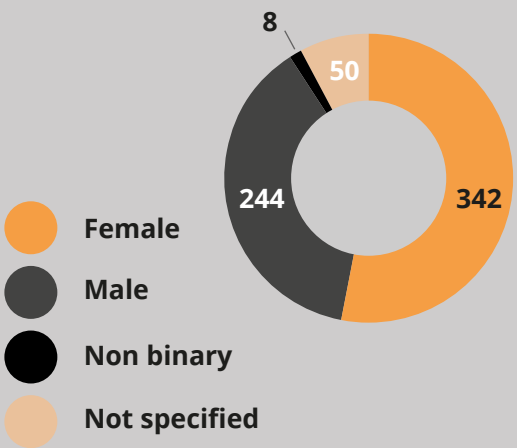
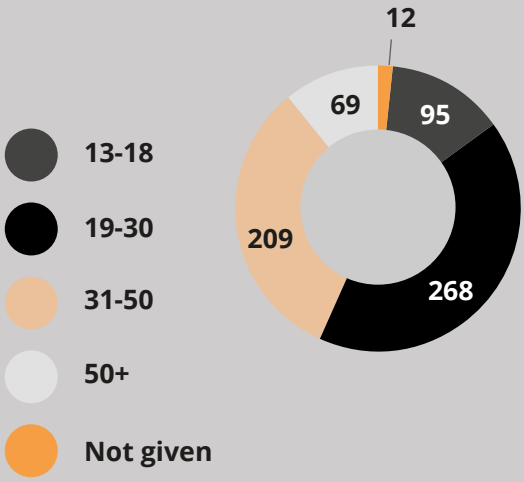
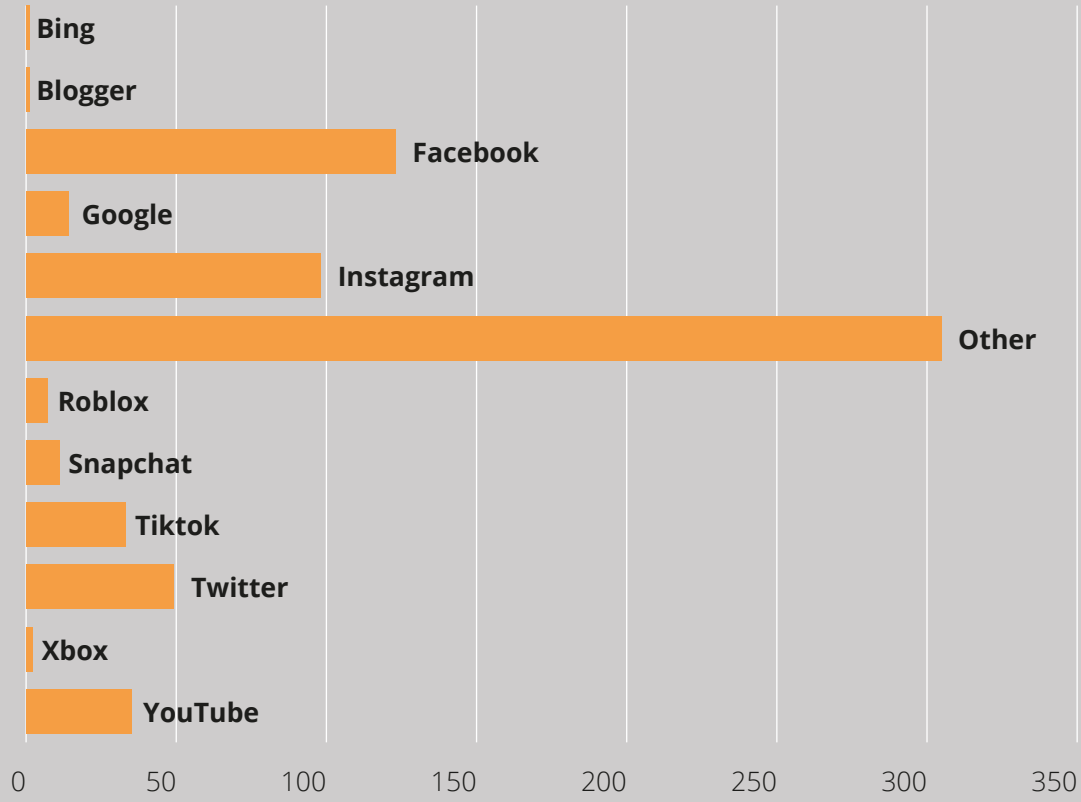


Figure 13: Clients by age



Of all industry platforms partnered with, harmful content was most likely to be located on Facebook (123 pieces of content), Instagram (98), Twitter (49), YouTube (35), TikTok (33), Google (14) and Snapchat (11) as illustrated in Figure 14. Three hundred and five pieces of content were located on sites other than industry platforms with which RHC work in partnership. Clients often made reports about multiple pieces of harmful content, located on a range of platforms, which is why the total pieces of harmful content is greater than the total number of reports.

Figure 14: Location of content



Overleaf Figure 15 shows the proportion of content on platforms by age. Figure 16 indicates the harms by type that were linked to sites not within the remit of RHC. These include additional harms centred around child sexual abuse material (CSAM) which were reported to the IWF and CSAM-narrative which include fan-fiction stories of underage characters. Privacy violation was also included as an additional harm identified on other sites. Of these, pornography was the most reported harm on other sites (36 reports) followed by violent content (23 reports) and bullying and harassment (20 reports).

Figure 15: Proportion of content on platforms by age

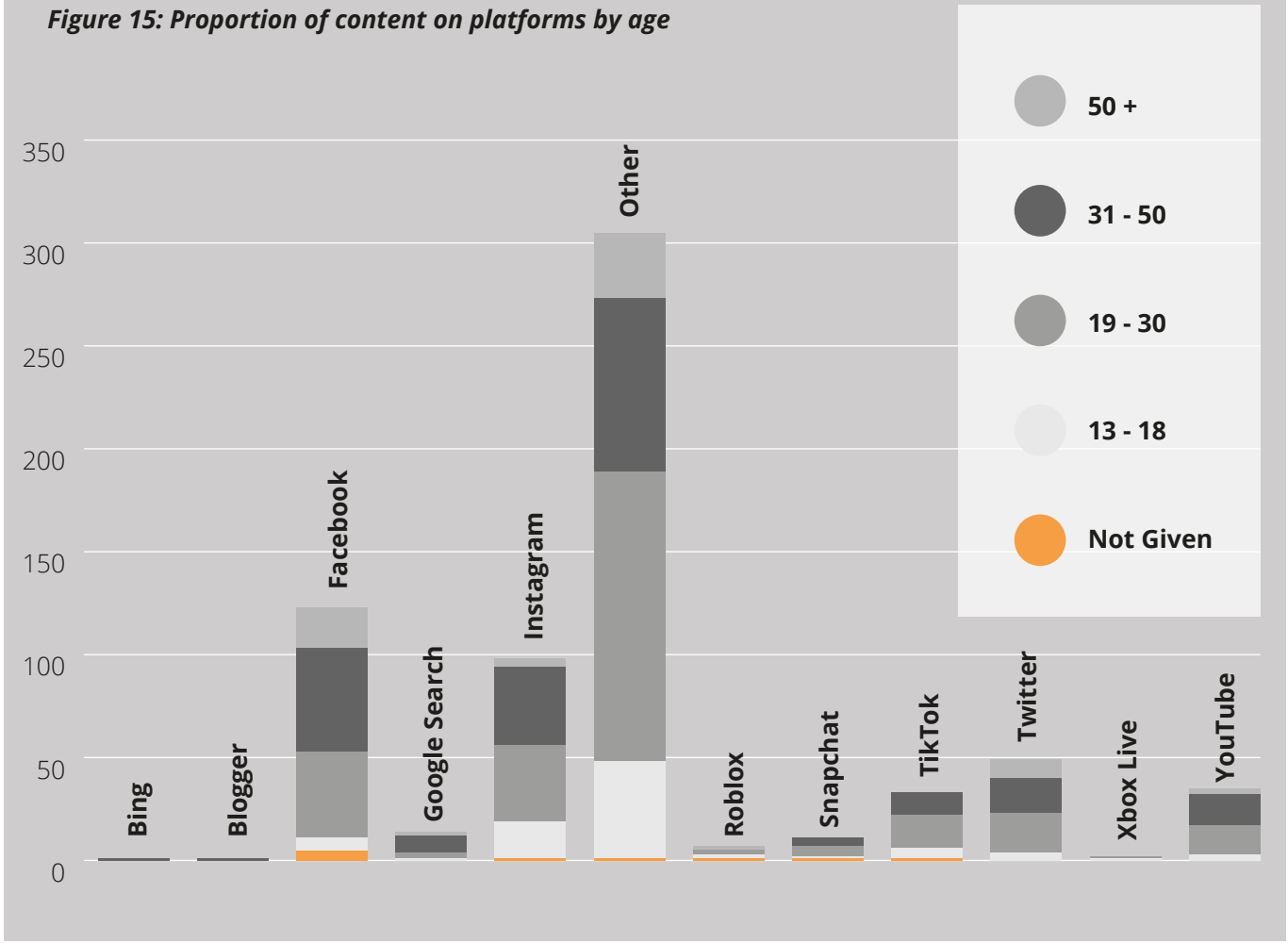
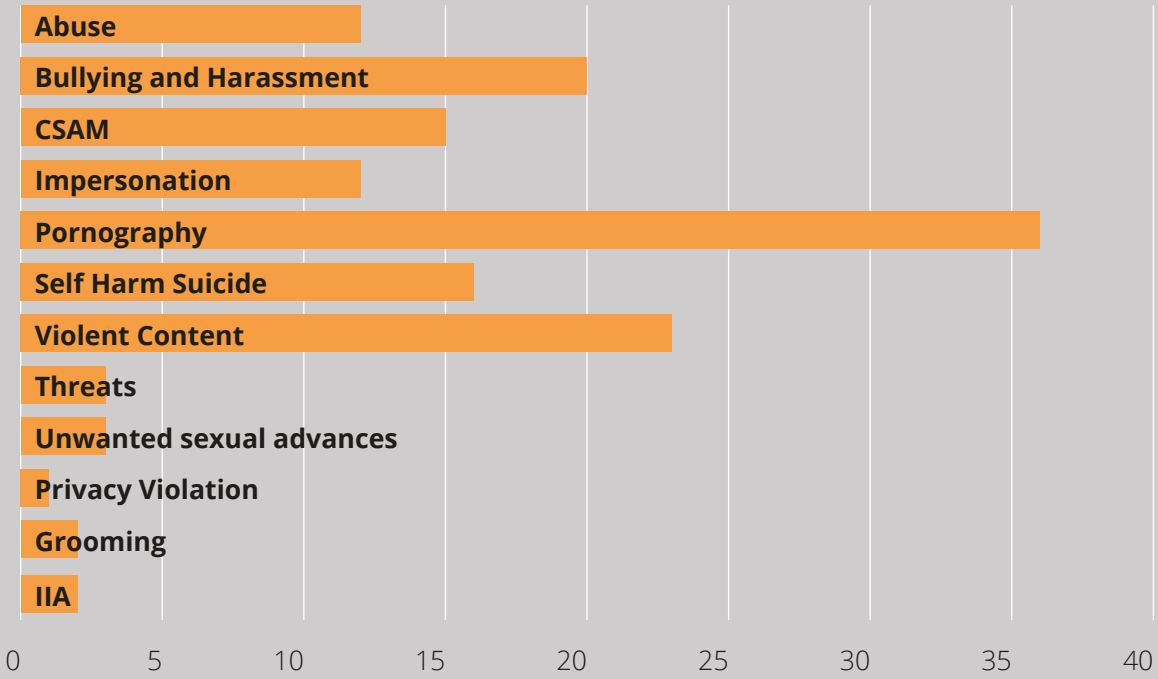


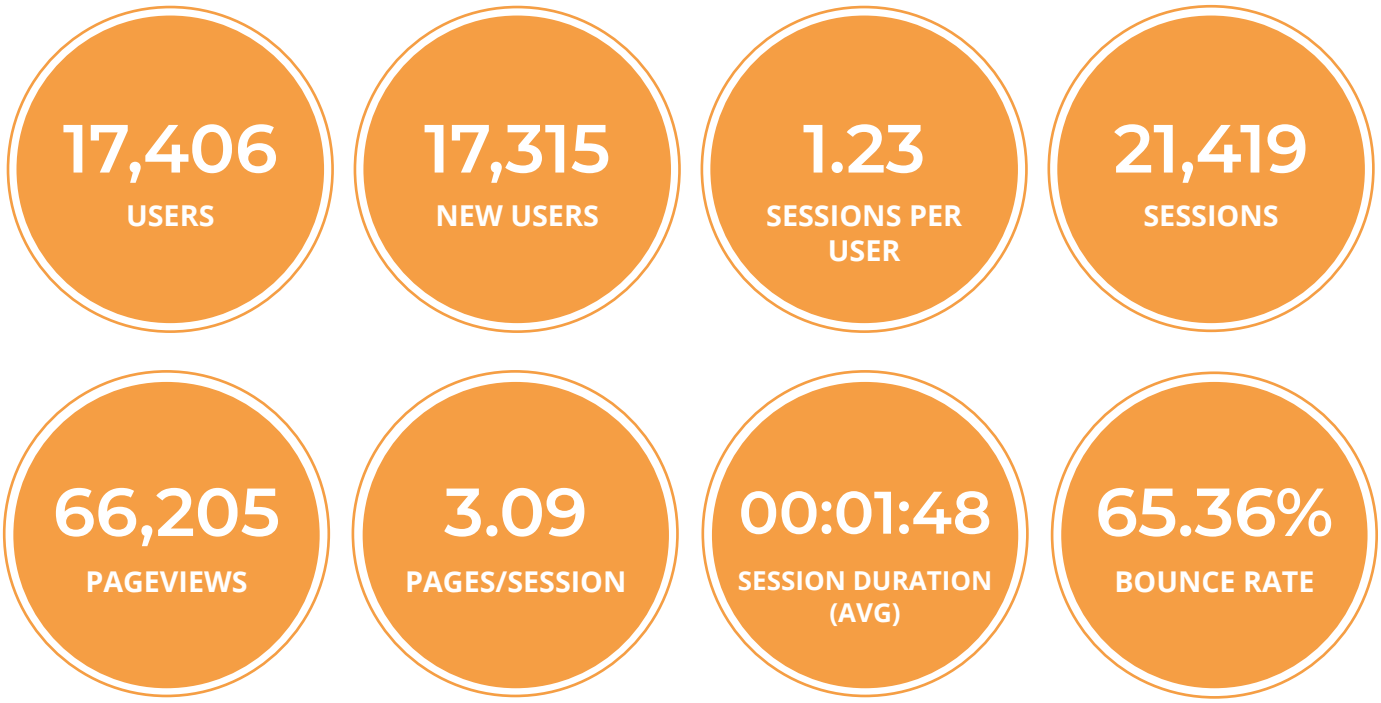
Figure 16: Sites not in remit containing harmful content



8. RHC Website Stats:

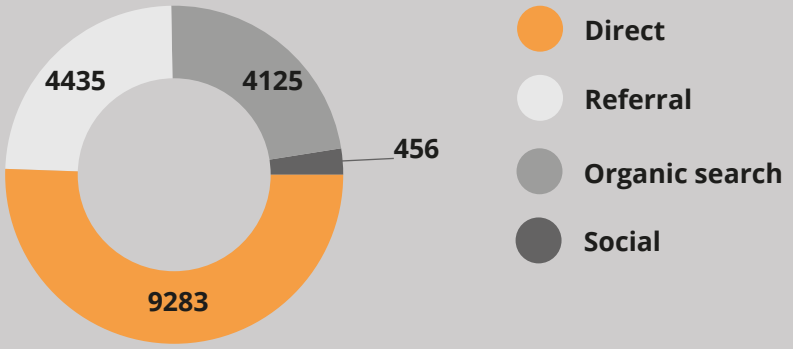
During 2020 the RHC website received 17,315 visitors, 80.9% of which were new visitors. The majority of website visitors (50.7%) went directly to the website, 24.2% of visitors were referred from other websites such as the Internet Watch Foundation, True Vision and the UK Safer Internet Centre website, followed by 22.5% that found the site from search engines.

Figure 17: website statistics 2020

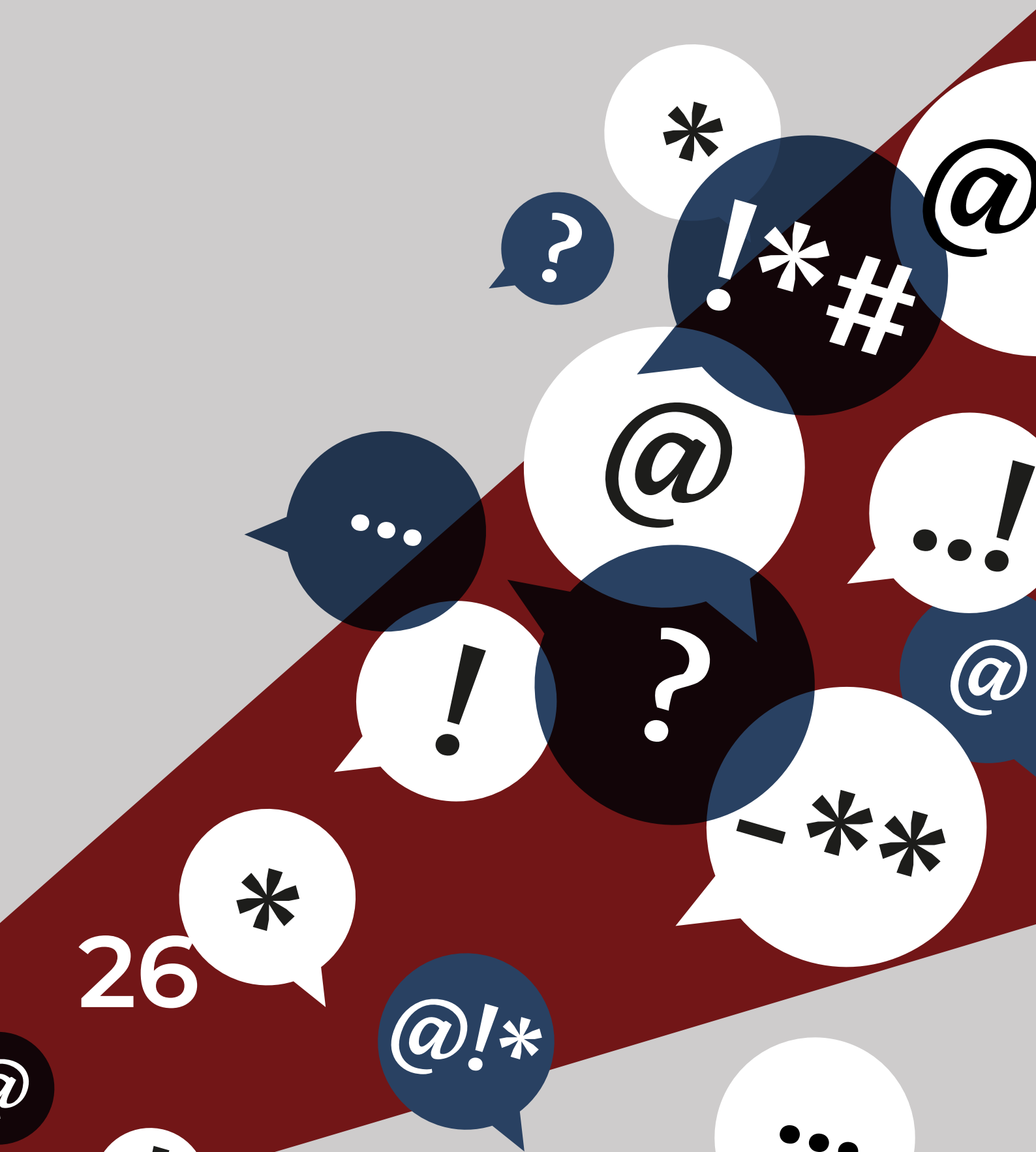


Website visitors predominantly used desktops to access the website (65.7%), followed by 31.6% of users using mobile devices.

Figure 18: RHC Acquisition channels 2020



Exploring reports in more depth



1. Domestic Abuse

Trend one: Throughout 2020, many wider issues (figure 9) overlapped and intersected with one another. The first trend identified on RHC was a cluster of domestic abuse, coercive control and harassment issues. This usually took the form of an ex-partner or an ex-partner's current partner hacking into an old social media account and posting private images. In 91% of reports where domestic abuse was reported as a wider issue, there were further harms associated. In 58% of these reports, domestic abuse was linked to either harassment, coercive control or both. In 50% of reports, coercive control was an active issue and in 41%, harassment was an issue. The majority of reports of this cluster were raised by either a loved one or a friend of the victim (59%) while 41% of reports were reported by the victims themselves.

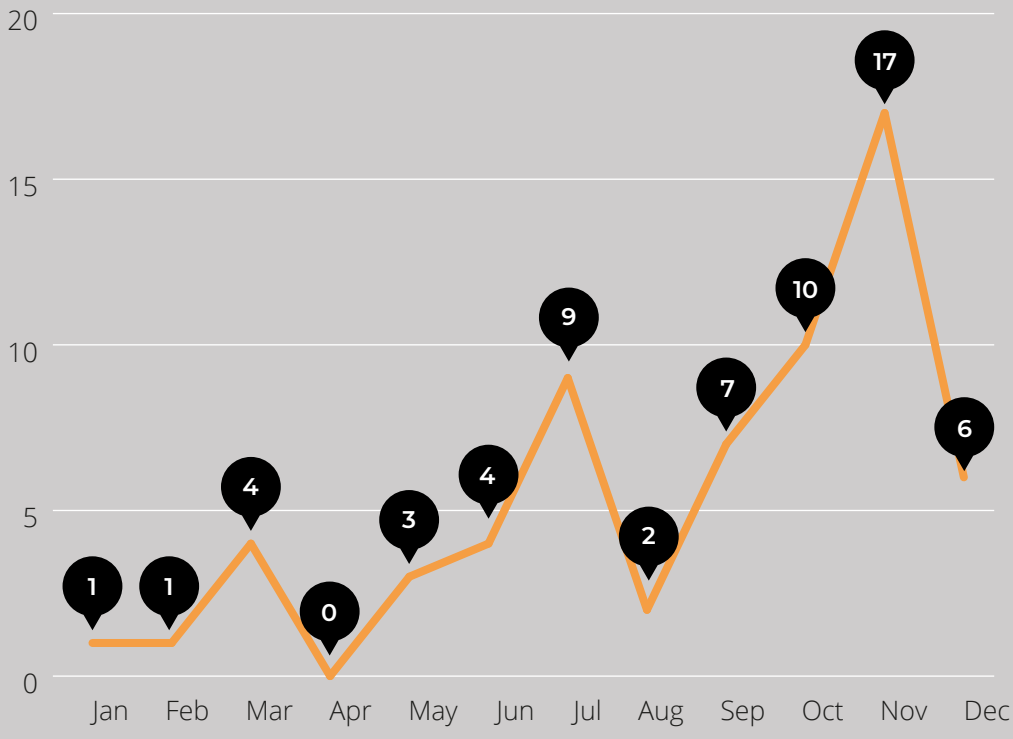
In these reports, a third of the reported content was located on Facebook followed by 25% on Instagram and 25% on 'other' sites not within the remit of RHC. Those reporting domestic abuse-related content were predominantly aged between 19– 30 (41%) and 30–50 (33%). Additionally, 75% of reports came from women (25% reported by men). In 75% of reports, the perpetrator was known by the victim and in 15% of cases, the reported content led to mental health issues. In 33% of reports there was an element of cultural or religious sensitivity, with ex-partners posting pictures deemed to be against specific cultural or religious values. In 25% of reports, intimate image abuse was also an additional issue. In 35% of reports, RHC recommended that the client should contact the police to report ongoing domestic abuse and harassment issues.

Technology is playing a growing role in cases of domestic abuse in the UK. Between April 2020 and May 2021, domestic abuse charity Refuge found a 97% increase in the number of domestic abuse cases requiring specialist tech support. Avast found an increase of 93% in the use of stalkerware and spyware apps from March 2020. In 2019 the Office for National Statistics found that out of the 376 prosecutions for Intimate Image Abuse recorded in the year ending March 2019, 83% (313) were flagged as being domestic abuse related.

2. Rise in hate speech

Trend two: the second trend identified noted a rise in reports where a wider issue of hate speech was identified. In 2020, 64 reports were linked to hate speech, an increase of 255% from the 18 cases identified in 2019. The increase in reports identifying hate speech as an additional issue throughout the year could be explained by a greater awareness of what indicates hate speech after the Black Lives Matter protests over the summer of 2020. RHC also launched their 'Negate the Hate' resource as part of the wider SWGfL helplines universities campaign over the autumn of 2020. This could help to explain the sharp increase of reports in September and October. Thirty-four per cent of reports linked to hate speech flagged content that was racist in nature.

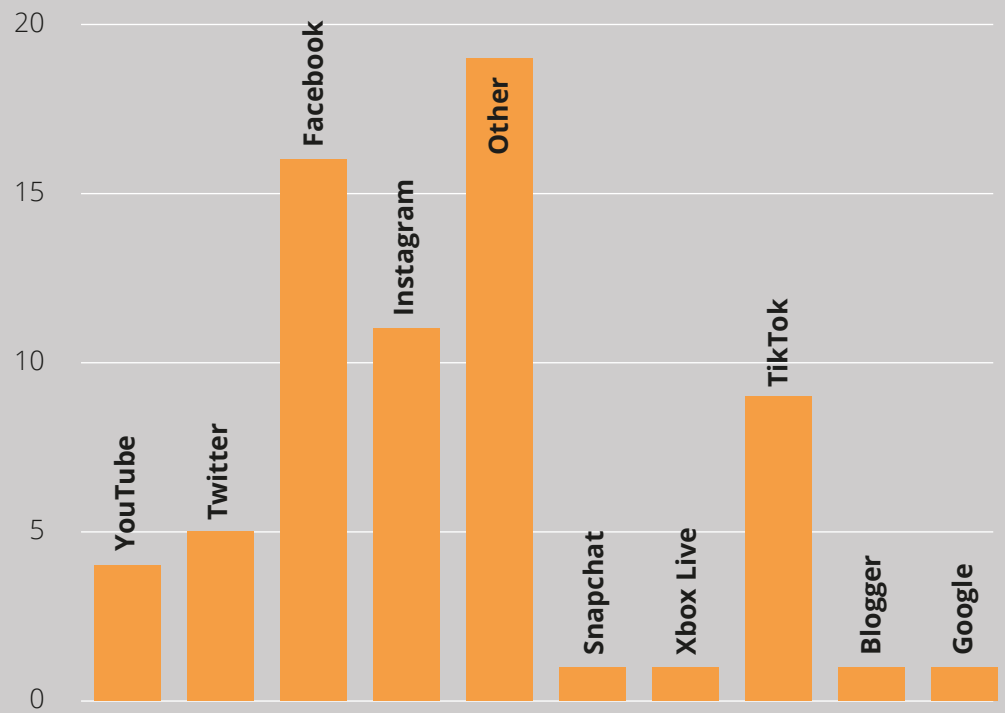
Figure 19: Hate speech in 2020



This trend was split fairly evenly between genders with 46% or reporters identifying as female and 30% identifying as male. Additionally, hate speech was also the issue most reported by those identifying as non-binary (4%). Those that reported issues linked to hate speech were predominantly aged between 19–30 (46%), with those aged 31–50 accounting for 31% of reports.

This issue was found predominantly in reports linked to bullying and repeated harassment (23 reports) followed by abuse cases (19 reports). Those reporting hate speech were predominantly observers (75%) with 21% of individuals reporting hate speech online that they were victims of. When analysing the location of hate speech content, 19 reports were found on 'other' websites while Facebook was found to be the partner organisation with the most instances of hate speech (16 reports) followed by Instagram (11 reports).

Figure 20: Platforms where hate speech was located



The overall increase in hate speech was reflected in numerous reports in 2020 and all indicated that hate speech was on the rise. Digital Awareness UK found that "in 2020, we've seen young people in their millions passionately take to social media in support of causes they believe in, following campaigns like the Black Lives Matter movement. But these efforts have sadly been met by a marked increase in posts encouraging harmful ideologies, such as anti-immigration, homophobia, xenophobia, racism and anti-Semitism." The Commission for Countering Extremism stated that "Since the outbreak of the coronavirus (COVID-19) pandemic, the Commission for Countering Extremism has heard increasing reports of extremists exploiting the crisis to sow division and undermine the social fabric of our country." The Home Office also found an 8% increase in hate crimes in the year ending March 2020 from the previous year.

3. Young males actively searching for harmful content and reporting it.

Trend three: An emerging trend and concern identified in RHC cases from 2020 was focused around young men actively searching for and reporting pornographic content. Of all the 'other sites' not within the remit of RHC, the majority were reported for pornographic content (25% of cases). Additionally, pornography was the only harm that was predominantly reported by males. The most common age group that reported pornographic content was 19–30 (54%) followed by 31–50 (26%), 13–18 (11%) and 50+ (8%). In the most common age category 67% of reporters were male. There were 17 reports from people aged 19–-30 who were actively searching for harmful content making this the most likely age group to be searching for harmful content (62%). The gender differences between those actively searching for harmful content were 48% men, 40% women and 11% unspecified. In cases involving those actively searching for harmful content 44% of reporters were reporting pornographic content. Men were twice as likely to be searching for harmful pornographic content than women (66% of men vs 34% of women).

It is evident then that younger males are the most at risk of actively searching for pornographic content online. Rissel et al (2016) found that younger individuals had higher rates of reporting a bad effect from pornography, which may include guilt as a result of anti-pornography discourse in educational materials which could, in turn, run the risk of inculcating self-hatred in young people who consume pornography. Chelsen (2011) also found that the more time a student spent accessing pornographic content online, the more likely they were to report feeling guilty about it. One potential explanation for this could be that these young men are reporting this content as a way to alleviate and manage the guilt they feel about accessing it. More research should be done in this field in order to understand whether guilt is driving young men to report pornography in order to manage their feelings.

Recommendations

1. The value of the service:

Report Harmful Content is evidently meeting its objective of helping everyone to report harmful content online. RHC deals with reports from a range of demographics, across multiple platforms. During 2020, RHC practitioners dealt with a wide variety of online harms, the majority of which overlapped with wider issues, both on and offline. The value of the service lies in the way in which it addresses online harms, not in isolation, but holistically. Practitioners draw upon a range of escalation options, support services and referral routes in order to offer support that is uniquely tailored to individual reports.

Not only is RHC effective at tackling the complexity of online harm, it is also efficient. The high percentage of content which was successfully actioned by industry, 90%, clearly demonstrates this. The high level of referrals to RHC from the police, alongside the openness for police to work on reports in conjunction with practitioners, demonstrates the way in which RHC continues to be a trusted service used in conjunction with official criminal procedures.

Finally, the steady growth in reports throughout 2020 and the increase in cases from 2019 evidences the clear and increasing demand for this service. Over the past 10 years, the UK Safer Internet Centre has received funding from the European Union – this has been instrumental in its ability to operate. Without funding for the next financial year, the UK Safer Internet Centre will potentially be unable to deliver critical programmes and services such as RHC. Practitioners are also keen for the service to expand and develop, however, they are currently working at full capacity. To this end, an increase in funding is also desperately needed to meet existing demand and to equip practitioners to deal with the widening range of issues reported.

2. Responding to emerging trends:

- Significantly, the first trend identified shows an increasing overlapping of wider issues. Out of the 279 cases that included wider issues, a third of these cases found multiple issues at play. Anecdotal evidence from those working on RHC found that the overlapping of issues has only increased with the Covid-19 pandemic and is likely to be higher still in 2021. This additional complexity in cases, potentially resulting from the past year, will require practitioners to gain a greater understanding of the varying forms of online harm alongside the wider issues associated. Furthermore, in order to tackle this increase in overlapping issues, RHC practitioners should continue to work holistically with individuals in addressing their needs. This could mean directing clients to additional support services and law-enforcement channels, as well as monitoring cases over a prolonged period of time and ensuring channels of communication are kept open should the issue resurface. The large overlap of issues dealt with by RHC and the Revenge Porn Helpline demonstrates a need for more tailored support for clients and could be an area to explore in future service development.

- The second trend, indicating a rise in hate speech reports, reflects many reports released through 2020. During 2020 there were also numerous events that led to a rise in hate crimes. The Covid-19 pandemic led to an upset in many individuals' personal, social and economic lives which, in turn, has increased anxiety and fear levels (Ahorsu at 2020). This increase in emotions has led to incidents across the world such as online harassment and abuse. These incidents include the global Black Lives Matter protests (Ziems et al 2020). The increase in hate speech that RHC is dealing with is concerning as it reflects a systemic problem which seems to be deteriorating rather than improving. Many industry platforms have proactive content moderation processes in place for removing this type of content, however, these will need to continue to be adapted to match the pace of the ever-changing societal landscape. It is evident that social media companies need to develop more consistent policies in swiftly identifying and removing hate speech online

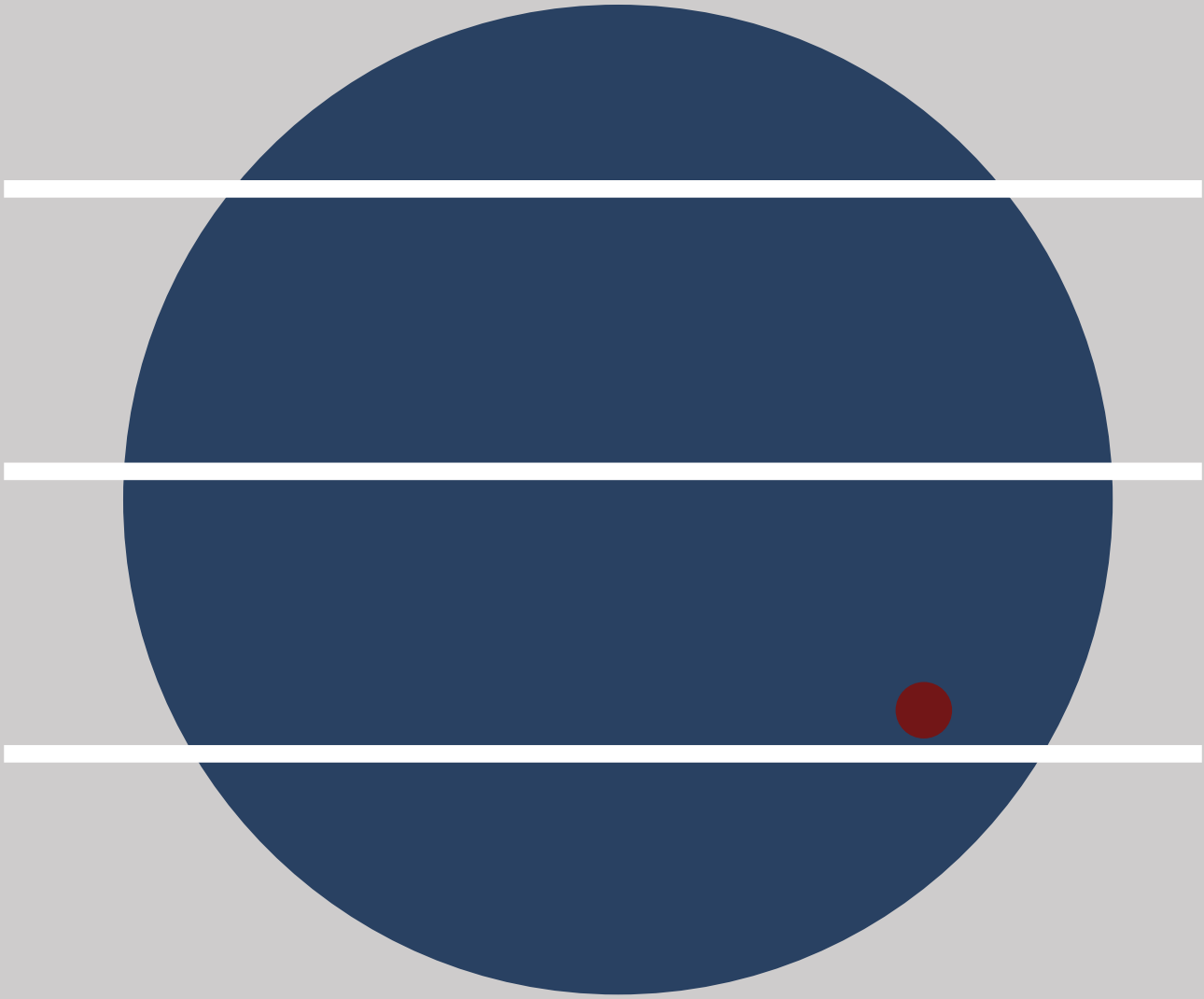
(Williams & Mishcon de Reya, 2019). The Online Harms Bill should mitigate this somewhat in its use of Ofcom as an independent regulator awarding the body with the power to block and fine services that do not protect their users. However, some critics argue that this will not go far enough with the NSPCC remaining critical, arguing that the law should threaten criminal sanctions against senior managers (Essex Barrett 2020).

- Increasingly, the need for greater research in the area of online harms and accompanying issues is also needed. Some interesting observations occurred over the past year, including the third trend. RHC practitioners found that younger men were actively searching for harmful pornography that they could report, potentially as a way to alleviate their feelings of guilt and shame about accessing and viewing this content. Stop It Now have resources for individuals who are concerned about their own behaviour in using the internet to access pornography featuring minors, assisting them in building a healthier, happier life. However, very little research has been undertaken in this area to prove or disprove these observations. Indeed, more research could be undertaken on individuals actively searching for harmful content as a whole.

3. Greater industry partnership

- The data identifies that there is a need for greater industry partnership. In 2020, 52% of cases weren't within the remit of RHC, while RHC signposted relevant services and agencies they were unable to report the content, predominantly as they were on sites that have not yet partnered with RHC. As a result, if more platforms were to partner with RHC this number could be reduced. They were on sites that have not yet partnered with RHC. As a result if more platforms were to partner with RHC this number could be reduced.
- Moreover, 23% of escalations by RHC practitioners to industry partners took over 72 hours to be actioned by industry. During the global pandemic, industry partners have had to adapt moderation processes in light of remote working. This resulted in some delays in responses to escalations. For the majority of platforms where delays were identified, these were short-lived with normal prompt response times resumed midway through the first national lockdown. However, for some larger industry partners, there continues to be significant delays responding to escalations from RHC attributed to Covid-19. As such, we would recommend that larger industry partners further streamline processes and increasing capacity to account for the inevitable increase in escalations from RHC, ensuring more content is actioned rapidly.
- 22% of clients reporting during 2020 were redirected to the correct reporting route for an industry platform (Figure 10). A large portion of these clients were not users of the industry platforms where they had experienced/witnessed harmful content. This, coupled with evidence from interviews conducted with RHC practitioners, highlighted a need for clearer navigation of reporting routes alongside easier access to reporting routes for non-users of industry platforms.
- Ofcom has proposed, as part of their upcoming regulation of Video Service Providers (VSP's), that all industry platforms in scope should be required to have an impartial dispute resolution procedure in place. RHC is an established impartial dispute resolution provider. Ofcom have indicated that the Online Harms Bill is likely to supersede VSP regulation, it is recommended that industry in scope make use of RHC, adhering to the requirements ahead of the regulation coming into force.

Conclusion



The 2020 RHC Annual Report has presented results from mixed-methods research carried out on all reports dealt with from January 2020 to December 2020. It used data collected from RHC tickets, the RHC website and practitioners in order to identify the top-level statistics that analysed client demographics, service response, website statistics and the nature of reports. Some emerging trends and recommendations were also identified with reference to the RHC data collected, practitioner experiences and external sources.

The trends identified in the 2020 report were determined with regards to both the service evaluation as well as pertinent policy and legislation. With the recent publication of the Online Harms Bill, it appears that policymakers are aware of some of the key issues and trends alongside industry attempts at tackling online abuse. However, the report demonstrates that there is more work that needs to be done to ensure harmful but legal content as well as illegal content and associated offline harms are dealt with. The report demonstrates that the most vulnerable need to be protected with minority groups, women and young people with mental health issues being the focus of the trends.

There are some limitations to this report. While certain client demographics are collected by RHC practitioners, others such as race or sexuality are not. As a result, the report was unable to analyse the experiences of these minorities and the online harms landscape. To this end, readers are asked to bear in mind that some minority experiences have not been represented. Moreover, there has been little mention on the effects of the Covid-19 pandemic in relation to the data, while the data analysed didn't necessarily identify any trends specifically linked to the pandemic, a more in-depth comparative analysis could take place once restrictions are eased. Finally, the effects of the Online Harms Bill on harmful content online and specifically RHC was unable to be analysed in this report, this could be an additional focus in future reports.

Vitally, this report has identified the increased popularity and need of the RHC service. However, with European funding due to end at the end of December 2021, future funding is needed to ensure this vital service can continue to meet service demands. In 2020, 4% of those using RHC expressed suicidal ideation meaning that, in its first full year of operation, RHC potentially helped to save 25 lives. Over the past 10 years, the UK Safer Internet Centre, which runs RHC, has received funding from the European Union, this has been instrumental in its ability to operate. While RHC will continue to seek additional partnerships with industry, without fresh government support for the next financial year, the UK Safer Internet Centre will be unable to deliver critical programmes and services.

Bibliography

Ahorsu, D., Lin, C.Y., Imani, V., Saffari, M., Griffiths, M. and Pakpour, A. (2020). The fear of covid-19 scale: development and initial validation. International journal of mental health and addiction. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7100496/

Avast (2021). Use of Stalkerware and Spyware Apps Increase by 93% since Lockdown Began in the UK. Retrieved from https://press.avast.com/use-of-stalkerware-and-spyware-apps-increase-by-93-since-lockdown-began-in-the-uk

Chelsen, P. (2011). An examination of Internet pornography usage among male students at Evangelical Christian colleges. Retrieved from https://www.proquest.com/docview/919995526?pq-origsite=gscholar&fromopenview=true

Commission for Countering Extremism (2020). How hateful extremists are exploiting the pandemic. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/906724/CCE_Briefing_Note_001.pdf

Department for Digital, Culture, Media and Sport (DCMS). (2019). Online harms white paper. Department for Digital, Culture, Media and Sport. Retrieved from https://www.gov.uk/government/consultations/onl ine-harms-white-paper

Digital Awareness UK (2020). Tackling the rise in hate speech during COVID. Retrieved from: https://newscentre.vodafone.co.uk/digital-parenting/tackling-the-rise-in-hate-speech-during-the-pandemic/

Essex Barrett (2020). More Harm Than Good?: Ofcom given the power to penalise online services under the Online Harms Bills. Retrieved from https://www.littlelaw.co.uk/2020/12/27/more-harm-than-good-ofcom-given-the-power-to-penalise-online-services-under-the-online-harms-bills/

Galop (2020). Online Hate Crime Report. Retrieved from https://www.report-it.org.uk/files/online-crime-2020_0.pdf

Home Office (2020). Hate crime, England and Wales, 2019 to 2020. Retrieved from https://www.gov.uk/government/statistics/hate-crime-england-and-wales-2019-to-2020/hate-crime-england-and-wales-2019-to-2020

Kumar, A., Pranesh. R. and Pandey, S (2020). TweetBLM: A Hate Speech Dataset and Analysis of Black Lives Matter-related Microblogs on Twitter. Retrieved from https://openreview.net/pdf?id=XG2YfqOQFwO

ONS (2019). Domestic abuse prevalence and trends, England and Wales: year ending March 2019. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/domesticabuseprevalenceandtrendsenglandandwales/yearendingmarch2019

Refuge (2021). Refuge launches Tech Safety Website. Available at https://www.refuge.org.uk/refuge-launches-domestic-abuse-tech-safety-website/

RHC (2020). Report Harmful Content Pilot Year Evaluation. Retrieved from https://d1afx9quaogywf.cloudfront.net/sites/default/files/RHC%20Report%20Final%20with%20Logos_0.pdf

Rissel, C., Richters, J., de Visser, R., McKee., A, Yeung, A. & Caruana, T (2016). A Profile of Pornography Users in Australia: Findings From the Second Australian Study of Health and Relationships. The Journal of Sex Research. 54:227240.

Stop It Now! (2021). Concerned about your own thoughts or behaviours? Retrieved from https://www.stopitnow.org.uk/concerned-about-your-own-thoughts-or-behaviour/concerned-about-use-of-the-internet/why-change/

Williams, M. (2019). The connection between online hate speech and real world hate crime. Oxford University Press blog. Retrieved from https://blog.oup.com/2019/10/connection-between-online-hate-speech-real-world-hate-crime/

Ziems, C., He, B., Soni, S. and Kumar, S. (2020). Racism is a virus: Anti-asian hate and counterhate in social media during the covid-19 crisis. Retrieved from https://arxiv.org/pdf/2005.12423.pdf