

Response to the public consultation on the
*Regulation of Harmful Content on Online
Platforms and the Implementation of the
Revised Audiovisual Media Services Directive*

Éibhear Ó hAnluain*

April 15, 2019

Contents

1 Introduction	1
1.1 The nature of the internet from the perspective of the technology	2
1.2 Self-hosting	3
1.3 Abuse	5
1.4 Harmful content definition	6
2 Answers to consultation questions	7
2.1 Strand 1 – National Legislative Proposal	7
2.2 Strand 2 – Video Sharing Platform Services	9
2.3 Strands 3 & 4 – Audiovisual Media Services	10
2.4 Strands 1 & 2 – European & International Context	10
2.5 Strands 1-4 – Regulatory Structures	11
2.6 Strands 1 & 2 – Sanctions/Powers	11

1 Introduction

My name is Éibhear Ó hAnluain.

I have been working in the IT industry since 1994, initially as a software engineer, more recently as an IT systems architect, and I am currently a consultant IT systems architect employed by Dublin-based consultancy organisation¹.

I am responding to this public consultation in my personal capacity, and my views here are not necessarily those of my employer, nor those of any of my employer’s clients.

*eibhear.geo@gmail.com, 086 8565 666, <http://www.gibiris.org/eo-blog/>

¹www.linkedin.com/in/eibhearohanluain

In this submission I am seeking to highlight 3 concerns with respect to the "Regulation of Harmful Content on Online Platforms".

- How such a regulation could affect small or hobbyist services
- How such a regulation could be abused by bad-faith actors
- How such a regulation could define "Harmful Content"

I would like to outline some initial thoughts on these matters first before addressing the specific questions of the consultation.

1.1 The nature of the internet from the perspective of the technology

1.1.1 Technical protocols

Formally, "the Internet" is a mechanism for identifying computers on a network, and for ensuring that messages from one computer on the network get to another computer. For this purpose, each computer is assigned an address (e.g. 78.153.214.9). This system is called The Internet Protocol².

These dotted-notation addresses are associated with more easy-to-remember name-based addresses by means of a system called the "Domain Name System"³.

There are a number of protocols⁴ for transmitting messages over the Internet, with two of the more common being "TCP"⁵ and "UDP"⁶.

The software required to implement these communications protocols is installed onto all forms of internet-connected devices, ranging from objects as small as (or smaller than) heart pacemakers, to as large as the largest super-computers.

This software is not aware of the size or capacity of the device it's installed on. Similarly, the protocols mentioned above have no regard to the purpose its host computer has, nor to who owns it, nor to how large it is.

The "World Wide Web" (the Web), from a technological perspective, is *not* the Internet. The Web is a set of defined protocols that make use of the Internet. Unlike the Internet and transmission protocols – which are designed to require each computer to regard all others as peers – the Web operates a little more on a client-server basis: the software package, often referred to as a web browser, on one computer is used to request specific information from the software package, often referred to as the web server, on the other computer.

However, despite the "client-server" nature of the Web, due to the simplicity of the software needed for a computer to be a web server, you can find web serving software operating on extremely small "IoT" devices.

²As defined in RFC 791: <https://tools.ietf.org/html/rfc791>

³As defined by RFC 1034 (<https://tools.ietf.org/html/rfc1034>) and RFC 1035 (<https://tools.ietf.org/html/rfc1035>).

⁴For the purposes of this document, a protocol is a set of instructions detailing how two or more computers should express queries and responses to each other

⁵<https://tools.ietf.org/html/rfc1035>

⁶https://en.wikipedia.org/wiki/User_Datagram_Protocol

1.1.2 Low barrier of entry for useful technology

The above demonstrates that someone with a computer, a connection to the internet and sufficient time and determination can set up a web service that will function just like the services we're all familiar with.

This is exemplified by the development of certain internet-related technology in recent decades:

- The *Linux* operating system kernel is named after its inventor, Linus Torvalds, who started work on it in 1991 as a college project – he wanted to write a computer operating system that was accessible to all, and which functioned in a specific way. The Linux operating system now forms the basis of a significant proportion of internet connected computing devices globally⁷ (including 73% of smartphones and tablet computers, through Google's Android, and somewhere between 36% and 66% of internet-facing server computers), and 100% of supercomputers.
- The *Apache* web server started development when a group of 8 software developers decided they wanted to add functionality to one of the original web server software packages, *NCSA httpd*. The Apache web server now powers 43.6% of all web sites⁸.
- The *Firefox* web browser was initiated by three software developers who wanted to make a light-weight browser based on the Mozilla code-base. At the height of its popularity, *Firefox* was used in 34% of web-page requests, despite not coming installed by default on any computer or mobile device. However, its real impact is that it was instrumental in breaking the monopoly that Microsoft's Internet Explorer held since the late '90s, resulting in far richer and more secure web.

1.2 Self-hosting

1.2.1 The nature of self-hosting

Both the *Linux* operating system kernel and the *Firefox* web browser can be considered truly disruptive technologies. In both of their domains, their arrival resulted in a dramatic improvements in internet and other technologies.

This affect isn't unique to those examples. There are many alternatives to the systems that we are familiar with, all developed by individuals, or small, enthusiastic teams:

- *Twitter* isn't the only micro-blogging service: there's also *GNU Social*, *Mastodon*.
- One alternative to *Facebook* is *diaspora**

⁷https://en.wikipedia.org/wiki/Usage_share_of_operating_systems

⁸https://w3techs.com/technologies/overview/web_server/all. Incidentally, the no. 2 on that web page, with nearly 42% share of websites is *nginx*. It also started out as a project by an individual who wanted to solve a particular project.

- *Nextcloud* and *Owncloud* are examples of alternatives to *Dropbox*.

In the cases of all these alternatives, users can sign up for accounts on "instances" operated by third-party providers, or users can set up their own instances and operate the services themselves.

Many of these services can federate with others. Federation in this context means that there can be multiple instances of a service, communicating with each other over a defined protocol, sharing updates and posts. For users, federation means that they can interact with other users who aren't necessarily on the same node or instance. For administrators of instances, federation means that they can configure their instances according to their own preferences, rather than having to abide by the rules or technical implementation of someone else.

1.2.2 Real examples of self-hosting

I host a number of such services:

- *Éibhear/Gibiris* is my blog site.
- *Social Gibiris* is a micro-blogging service that is federated with others using the *AtomPub* technology. Thus, *Social Gibiris* is federated with many other instances of *GNU Social*, *Mastodon* and *Pleroma*.
- *git.gibiris.org* is a source-code sharing site that I use to make publicly available some of the software that I develop for myself.
- *news.gibiris.org* is a news-aggregation that allows me to gather all the news sources of interest to me into one location, which I can then access from wherever I am.
- *cloud.gibiris.org* is a file-sharing platform that I use with my family when we are collaborating on projects (e.g. school projects, home improvement projects, etc.)
- *matrix.gibiris.org* is an instant-messaging system which I set up for the purposes of communicating with my family and close friends.

Most of these services are hosted on a computer within my home. 3 of these services provide information to the general public, and the other three are accessible only to those who set up accounts.

2 of those services, *git.gibiris.org* and *Social Gibiris* can process or post user-uploaded information.

1.2.3 Regulation of self-hosted services

While it is attractive to create regulations to manage the large, profit-making organisations, it is imperative that such regulations don't harm the desire of those who want to create their own services.

Any regulation that applies liability on the service for someone else's words or behaviour, is a regulation that can be adhered to only by organisations with large amounts of money to hand. For example, if the regulation was to apply liability on me for posting made by someone else (and **somewhere** else – these are federated services) on the 2 implicated services that I run, I would have to shut them down, as I would not be able to put in place the necessary infrastructure that would mitigate my liability⁹. Given that my services are intended to provide a positive benefit to me, my family members and my friends, and that I have no desire to facilitate harmful behaviour on those services, a law forcing me to shut these services down benefits no one.

Similarly, a regulation that demands responses from services on the assumption that the service will be manned at all times, requires individuals who are self-hosting their services to be available at all times (i.e. to be able to respond regardless of whether they are asleep, or overseas on a family holiday, etc.)

This submission comes from this perspective: that small operators should not be unduly harmed by regulations; the likelihood of this harm coming to pass is greater when such small operators are not even considered during the development of the regulations. If the regulations have the (hopefully unintended) effect of harming such small operators, the result will not just be the loss of these services, but also the loss of opportunity to make the Web richer by means of the imposition of artificial barriers to entry. Such regulations will inhibit the development of ideas that pop into the heads of individuals, who will realise them with nothing more than a computer connected to the internet.

1.3 Abuse

All systems that seek to protect people from harmful or other objectionable material (e.g. copyright infringement, terrorism propaganda, etc.) have, to date, been amenable to abuse. For example, in a recent court filing, Google claimed that 99.97% of infringement notices it received in from a single party in January 2017 were bogus¹⁰:

A significant portion of the recent increases in DMCA submission volumes for Google Search stem from notices that appear to be duplicative, unnecessary, or mistaken. As we explained at the San Francisco Roundtable, a substantial number of takedown requests submitted to Google are for URLs that have never been in our search index, and therefore could never have appeared in our search results. For example, in January 2017, the most prolific submitter submitted notices that Google honored for 16,457,433 URLs. But on further inspection, 16,450,129 (99.97%) of those URLs were not in our search

⁹This assumes that my services aren't forced to shut down by the new EU Copyright Directive anyway

¹⁰<https://www.techdirt.com/articles/20170223/06160336772/google-report-9995-percent-dmca-takedown-notices-are-bot-generated-bullshit-buckshot.shtml>

index in the first place. Nor is this problem limited to one submitter: in total, 99.95% of all URLs processed from our Trusted Copyright Removal Program in January 2017 were not in our index.

Aside from the percentage of URLs noted that don't exist in Google's index, that a single entity would submit more than 16 million URLs for delisting in a single month is staggering, and demonstrates a compelling point: there is no downside for a bad-faith actor seeking to take advantage of a system for suppressing information¹¹.

More recently, there is the story of abuse of the GDPR's *Right to be Forgotten*. An individual from Europe made a claim in 2014, under the original *Right to be Forgotten*, to have stories related to him excluded from Google searches for him. This seemed to have been an acceptable usage under those rules. However, that this claim was made and processed seems also to be a matter of public interest, and some stories were written in the online press regarding it. Subsequently, the same individual used the *Right to be Forgotten* to have **these** stories excluded from Google searches.

This cat-and-mouse game continues to the extent that the individual is (successfully) requiring Google to remove stories **about his use** of the GDPR's *Right to be Forgotten*. Even stories that cover **only** his *Right to be Forgotten* claims, making no reference at all to the original (objected-to) story¹². This is clearly an abuse of the law: Google risks serious sanction from data protection authorities if it decides to invoke the "... exercising the right of freedom of expression and information" exception¹³ and it is determined that the exception didn't apply. However, the claimant suffers no sanction if it is determined that the exception *does* apply.

In systems that facilitate censorship¹⁴, it is important to do more than merely assert that service providers should protect fundamental rights for expression and information. In a regime where sending an e-mail costs nearly nothing, where a service risks serious penalties (up to and including having to shutdown) and where a claimant suffers nothing for abusive claims, the regime is guaranteed to be abused.

1.4 Harmful content definition

This submission will not offer any suggestions as to what should be considered "harmful content". However, I am of the belief that if "harmful content" is

¹¹The law being used in this specific case is the US Digital Millennium Copyright Act. It contains a provision that claims of copyright ownership on the part of the claimant are to be made under penalty of perjury. However, that provision is very weak, and seems not to be a deterrent for a determined agent: <https://torrentfreak.com/warner-bros-our-false-dmca-takedowns-are-not-a-crime-131115>

¹²<https://www.techdirt.com/articles/20190320/09481541833>

¹³GDPR, Article 17, paragraph 3(a)

¹⁴While seeking to achieve a valuable and socially important goal, this legislation, and all others of its nature, facilitates censorship: as a society, we should not be so squeamish about admitting this.

not narrowly defined, the system will allow bad actors to abuse it, and in the context where there is no risk to making claims, and great risk in not taking down the reported postings, loose definitions will only make it easier for non-harmful content to be removed.

2 Answers to consultation questions

2.1 Strand 1 – National Legislative Proposal

2.1.1 Question 1 – Systems

- The legislation should state in an unequivocal manner that it is not the role of web services to adjudicate on whether specific user-uploaded pieces (text, videos, sound recordings, etc.) can be considered harmful under the legislation. The law should make it clear that where there is a controversy on this matter, the courts will make such rulings.
- As regard a system, this submission would support a notice-counternotice-and-appeal approach. Such an approach affords the service operator and the accused party an opportunity to address the complaint before the complained-of material is taken offline. The following should be incorporated:
 1. A notice to a service operator that a user-uploaded piece is harmful should contain the following information:
 - That the notice is being raised under this legislation (citing section, if relevant).
 - That the person raising the notice is the harmed party, or that the person raising the notice is doing so on behalf, and at the request, of the harmed party. Where the harmed party doesn't want to be identified, the notice could be raised on their behalf by someone else. However, totally anonymous notifications under this legislation should not be permitted, as it would not be possible to determine the good-faith nature of the notice.
 - The specific (narrowly tailored) definition of "harmful content" in the legislation that is being reported.
 2. A notice to the user who uploaded the complained-of material regarding the complaint. This will allow the user to remove the material, or to challenge the complaint. An opportunity to challenge a complaint is necessary to forestall invalid complaints that seek to have information removed that would not be considered harmful under the legislation.
 3. Adequate time periods for both the complainant and the posting user to respond.
 4. Where responses aren't forthcoming...

- ... if the posting user doesn't respond to the initial complaint, the posting is to be taken down
 - ... if the complaining user doesn't respond to the posting user's response, the posting is left up.
5. Within a reasonable and defined period of time, the service provider will assess the initial complaint, the counter-notice, and the complainant's response to the counter-notice, and will decide whether to take the material down or to leaving it up, *citing clear reasons for the decision*.
 6. Where either party is not happy with the decision, they can appeal to the regulator, and if the regulator contradicts the service operator's decision, the service operator must abide by the regulator's ruling. In its consideration of the ruling, the regulator must be required to consider the rights of both parties.
- Responsibilities and obligations of the service provider **must** relate to the size of the service. For example, it's not reasonable to ask the service provider to respond within an amount of time for those services that would not have someone available within that time. Self-hosters or small, single-location, operations would not be able to respond within an hour if the complaint is made at 4am!
 - This system should not apply to complaints that a posting violates the service's terms and conditions. If the complaint isn't explicitly made under this legislation, it should not fall within the regulator's remit. **Under no circumstances should merely violating a service's terms and conditions (or "community standards") be considered an offence under this legislation.**

2.1.2 Question 2 – Statutory tests

The service operator should be protected from liability under the rules if the service can show the following:

- That the initial complaint was responded to appropriately and within a reasonable amount of time.
- That an appeal was responded to within a reasonable amount of time.
- That the poster and complainant were each offered an opportunity to respond
- That the responses, and any appeals, were given due consideration.
- That the final decision (whether to keep the post up or pull it down) was well-reasoned, and considered the context in which the post was made.
- That, where appeals have been made to the regulator, the service responds to any order from the regulator in a reasonable manner and within a reasonable amount of time.

2.1.3 Question 3 – Which platforms to be considered in scope

This submission is concerned to ensure that assumptions not be made that all affected platforms will be large, for-profit organisations with scores, or hundreds, or thousands of staff acting as moderators of user-uploads.

The legislation should also not assume that platforms that want to deal with user uploads **should** be of a particular nature, or size.

To make either assumption would be to chill lawful interactions between internet-connected parties, and would further entrench the larger players on the internet.

2.1.4 Question 4 – Definitions

- Please see my introductory comments on this matter.
- Definitions of "harmful content" must aim to be as narrow as possible, in order to avoid the potential of the legislation being used to target political speech.
- In respect of serious cyberbullying, it should be considered harmful content under the legislation not just when it targets a child. It should be considered cyberbullying and harmful even if it is an adult, if the complaint states that s/he is being harmed or fears harm should the complained-of behaviour continue.
 - In the event that the target of the cyberbullying is a public figure, there should be an additional burden on the complainant to state that the behaviour represents real intent to cause harm, and is more than people with opposing political or social views "shooting their mouths off".

2.2 Strand 2 – Video Sharing Platform Services

2.2.1 Question 5 – What are video-sharing services

This submission is not providing an answer to this question.

2.2.2 Question 6 – Relationship between Regulator and VSPS

This submission is not providing an answer to this question.

2.2.3 Question 7 – Review by Regulator

The regulator should require the following reports to be published by online services regarding complaints made under this legislation:

- Number of complaints, broken down by nature of complaint
- Number of complaints that were appealed to the service, broken down by nature of complaint and basis of appeal

- Number of appeals upheld, broken down by reason for appeal
- Number of appeals rejected, broken down by reason for rejection.
- Number of complaints/appeals that were appealed further to the regulator.

2.3 Strands 3 & 4 – Audiovisual Media Services

2.3.1 Question 8 – "Content" rules for television broadcasting and on-demand services

This submission is not providing an answer to this question.

2.3.2 Question 9 – Funding

RTÉ and its subsidiary services should continue to be funded by the government, either through the licence fee, general taxation or a mixture of both. RTÉ's editorial independence should be re-iterated in this law (and strengthened, if required, specifically to assure independence from the editorial demands of advertisers). It should be anticipated that RTÉ will eventually broadcast only over the internet, and that it will be both a live-streaming service (e.g. providing programming in a manner similar to its current broadcast schedule), **and** an on-demand service.

Funding of services other than RTÉ should only be considered for services operated by non-profit organisations such as trusts or charities, and such funding should also come with an assurance of editorial independence for the recipients.

2.4 Strands 1 & 2 – European & International Context

2.4.1 Question 10 – Freedoms

- Core to the consideration of the legislation is that everyone posting to services are presumed to be innocent of an offence, and their postings should also be presumed **not** to offend the law.
- Accusations of harm **must** be tested to determine if they are being made to suppress legal speech. This is particularly true where the person making the allegation is a public figure, or is representing a public figure.
- Where a service applies – or is required to apply – sanctions on users who repeatedly post harmful information, similar sanctions should also be applied to users who repeatedly make **false** accusations under the law.

2.4.2 Question 11 – Limited liability

Any regulatory system that makes service providers liable for what their **users** say on those services will result in one or a combination of the following effects:

1. Service will stop permitting users to make postings.

2. Where the value of a service is wholly, or in part, that it allows its users to post to it, the service may have to shut down.
3. Services will be sued or prosecuted for the actions of its users **regardless** of the effort and good faith they put in to "moderating" what is posted on their service – a concept that is borderline ludicrous in the off-line world. This would be analogous to a car manufacturer being liable for the consequences of car occupants not wearing their seat-belts.

There must be clarity in the regulations that a service is protected as long as it acts in a good-faith manner to deal with postings made by users that are determined to have been illegal. This reflects Ireland's obligations under various trade agreements to grant safe-harbour protections to internet services.

The regulation must also protect platforms and their users against bad-faith accusations of harm, particularly from public figures. If it is easier to use an accusation of "harmful content" than to claim libel, public figures will use that facility to suppress information they would like not to be known.

2.5 Strands 1-4 – Regulatory Structures

2.5.1 Question 12 – Regulatory structure

This submission is not providing an answer to this question.

2.5.2 Question 13 – Funding of regulatory structure

This submission is not providing an answer to this question.

2.6 Strands 1 & 2 – Sanctions/Powers

2.6.1 Question 14 – Functions and powers

This submission is not providing an answer to this question.

2.6.2 Question 15 – Sanctions

The following should be taken into account when considering sanctions on platforms

- The nature of the operation
 - Large, global, profit-based private organisations providing services to the general population. (examples include YouTube, Facebook, Twitter).
 - Smaller, local, profit-based private organisations providing services to the general population, focused on the region (examples might include boards.ie, everymum.ie, etc.)

- Small, non-profit forums set up by locally-based and -focused organisations such as soccer clubs, or school parents' associations¹⁵
- Individuals, hosting their own platforms.
- The good-faith efforts of the operation to respond to accusations of harm.
- The capacity of the service to respond – smaller operations can't afford 24-hour monitoring to respond to such accusations, and the law should not require it. Such services should be able to avail of bad-faith actors seeking to interfere with their operations by overwhelming them with false accusations of harm that need to be dealt with.
- Who the accuser is – public figures should be prevented from using accusations of "harmful content" to remove information that is merely critical of them or their behaviour.

2.6.3 Question 16 – Thresholds

This submission is not providing an answer to this question.

¹⁵There is often the temptation to advise these organisations to use larger platforms like Facebook or Google. Some organisations may not want to avail of those services, and the reasons for this are not relevant. What's important is that deciding not to use these platforms is valid, and these decisions should be protected and encouraged, not inhibited.