



Trusted notifiers and the privatization of online enforcement

Schwemer, Sebastian Felix

Published in:
Computer Law & Security Review

DOI:
[10.1016/j.clsr.2019.105339](https://doi.org/10.1016/j.clsr.2019.105339)

Publication date:
2019

Citation for published version (APA):
Schwemer, S. F. (2019). Trusted notifiers and the privatization of online enforcement. *Computer Law & Security Review*, 35(6), [105339]. <https://doi.org/10.1016/j.clsr.2019.105339>

Trusted notifiers and the privatization of online enforcement

Sebastian Felix Schwemer*

Centre for Information and Innovation Law (CIIR), Faculty of Law, University of
Copenhagen and Danish Internet Forum (DIFO), Copenhagen, Denmark

Original place of publication: Sebastian Felix Schwemer, “Trusted notifiers and the
privatization of online enforcement“ *Computer Law & Security Review*, Volume 35, Issue 6,
November 2019, 105339, <https://doi.org/10.1016/j.clsr.2019.105339>

Abstract

Online content is increasingly enforced by private parties based on private regulation. One recent trend in the takedown of unlawful online content is the emergence of models, where trusted third parties –private or public– are given privileged notification channels for flagging infringing content.

Despite increasing practical importance, these arrangements have received little scholarly attention. This article explores the functioning of trusted notifier-models and how they are addressed by the European lawmaker in the context of two intermediaries, online platforms and domain name registries. Depending on intermediary, trusted notifier-models can both be seen as extension of the existing notice-and-takedown regimes and an additional voluntary expedited-enforcement layer. The author argues that these trusted notifier-models are problematic given the broad room of autonomy that the legislator is leaving to private parties. Whereas models involving public authorities are subject to general administrative law principles as well as constitutional and human rights safeguards, the framework for private regulation (i.e. without intervention of public actors) is less clear. In the field of domain names, these legitimacy issues give raise to special concern given the detached relation between domain names and website content. The paper criticizes the lack of insights into existing arrangements and calls for increased transparency. The author concludes that a legislative minimum framework is desirable.

Keywords: Trusted notifiers, private ordering, online platforms, domain name system, online enforcement, privatization of enforcement, Recommendation (EU) 2018/334

* Centre for Information and Innovation Law (CIIR), Faculty of Law, University of Copenhagen, Karen Blixens
Plads 16, 2300 Copenhagen, Denmark.
E-mail address: sebastian.felix.schwemer@jur.ku.dk.

1. Introduction

Eighteen years ago, the European legislator stated that “[i]n many cases (...) intermediaries are best placed to bring (...) infringing activities to an end.”¹ This premise has kept its validity throughout the years but the answer to the question what this observation means for the role of intermediaries has changed. For almost twenty years, the liability exemption regime of the E-Commerce Directive has led to a more or less well-functioning² notice-and-takedown regime in the European landscape, which has been refined and adapted by the courts³ and, over time, attracted significant legal scholarship. Given the substantial development of the Internet since the adoption of this framework and the large amounts of content being made available today, intermediaries’ role is constantly revisited. Currently, notably in connection to hosting providers or online platforms as gatekeepers, the European framework for intermediaries and unlawful content is undergoing substantial changes: on the one hand, the European legislator, in certain areas, proposes moving from the traditional notice-and-takedown towards a proactive regime, where unlawful content is automatically identified and blocked before being available on online platforms.⁴ On the other hand, the legislator works towards increasing the efficiency of the current regime. The focus of this article is on the latter.

In both trajectories, the cat-and-mouse game of enforcing content is increasingly moving from the parliamentary and judiciary arena to the negotiation table of private actors.⁵ This development towards non-law-based measures⁶ is encouraged by the European lawmaker. The emergence of more or less voluntary arrangements, e.g. in forms of memoranda of understanding or codes of practices, has also been endorsed by others: in 2011, the Organisation for Economic Co-operation and Development (OECD) noted in its *Recommendation on Principles for Internet Policymaking* that these “should encourage and facilitate voluntary cooperative efforts by the private sector to (...) address illegal activity (...) taking place over the Internet.”⁷ Frosio (2017) aptly describes these policy tendencies as a shift from “intermediary liability to intermediary responsibility”.⁸

¹ Recital 59 InfoSoc Directive. See also Jonathan Zittrain, ‘A History of Online Gatekeeping’ (2006) *Harvard Journal of Law and Technology* 254.

² In 2016, the European Commission deemed it generally fit for purpose, whereas the European Parliament in 2017 called for clarifying action.

³ Such as the Court of Justice’s case law and national courts, e.g. in Germany in the context of notice-and-staydown.

⁴ E.g. in relation to copyright content with Article 17 of Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market and in relation to terrorist content with Article 6 of the proposed Regulation on terrorist content. See Thomas Riis and Sebastian F Schwemer, ‘Leaving the European Safe Harbor, Sailing Towards Algorithmic Content Regulation’ (2019) 22 *Journal of Internet Law* 1–21.

⁵ Already the E-Commerce Directive relied on codes of conduct for its implementation, see Art. 16 and recital 49 Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L 178/1 (E-Commerce Directive).

⁶ See the differentiation between law-based and non-law-based blocking of content by the Council of Europe (CoE), see Council of Europe, Commissioner for Human Rights, ‘The rule of law on the Internet and in the wider digital world’ (Issue paper, December 2014) 13.

⁷ Organisation for European Economic Co-operation (OECD), ‘OECD Council Recommendation on Principles For Internet Policymaking’ (2011) 7.

⁸ Giancarlo Frosio, ‘Why keep a dog and bark yourself? From Intermediary Liability to Responsibility’ (2017) Centre for International Intellectual Property Studies Research Paper No. 2017-11, 4 <https://papers.ssrn.com/abstract_id=2976023>

In the field of online content platforms, the European Commission recently endorsed the establishment of non-judicial takedown mechanisms, which rely on self-regulation.⁹ One proposed solution concerns privileged channels for takedown notices by so-called “trusted flaggers” or “trusted notifiers”, i.e. private or public entities with specific expertise in identifying illegal content. Such arrangements are or have been operational at different intermediary-levels; for example, for the content-related takedown of generic top-level (gTLD) domain names, the DNS-blocking of websites containing child sexual abuse material, or the takedown of terrorist propaganda on Youtube.

Whereas many questions related to the liability exemption and accountability framework are well-studied¹⁰, trusted notifier arrangements have up until now –despite increasing practical importance– received little scholarly attention. The goal of this article is to fill this gap and to shed light on the workings and governance of trusted notifier-models put in place by two different kinds of intermediaries, firstly online platforms (hosting services) relating to the application layer of the Internet and secondly domain registries relating to the infrastructure layer.¹¹

This article is structured in the following way: first, it explores the general characteristics of trusted notifier-models. Then, it turns towards the recent regulatory intervention regarding online platforms and how underlying legitimacy concerns are suggested to be addressed in the absence of specific secondary legislation. Following these insights, it looks into the content-related takedown of domain names, where trusted notifier-models are relevant but have remained off lawmakers’ radar for now.¹² By looking at this phenomenon in a horizontal fashion for different intermediaries, it finally provides a critical discussion and suggestions for improving the regulatory landscape.

2. Characteristics of the trusted notifier-model

When we look at the role of intermediaries in the takedown of unlawful content, there are roughly three scenarios that can be differentiated: firstly, an intermediary could take down content on its own independent initiative. Secondly, an intermediary could be obliged by a court order to take down content. Thirdly, an intermediary could be obliged to take down

⁹ One conceptual definition understands regulation “as the means by which the state ‘seeks to encourage direct behaviour which it is assumed would not occur without such intervention’ and as such should be seen as distinct from the operation of the markets, even though the latter is underpinned by legal rules” see Julia Black, ‘Critical reflections on regulation’ (2002) Centre for Analysis of Risk and Regulation, London School of Economics and Political Science, 11 <<http://eprints.lse.ac.uk/35985>> accessed 13 December 2018. See Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online [2018] OJ L 63/50.

¹⁰ See e.g. Christina Angelopoulos et al., *Study of fundamental rights limitations for online enforcement through self-regulation*, Institute for Information Law (IViR), 2016 <<https://www.ivir.nl/publicaties/download/1796>>; Jaani Riordan, *The Liability of Internet Intermediaries* (Oxford University Press, 2016); Tatiana-Eleni Synodinou, ‘Intermediaries’ liability for online copyright infringement in the EU: Evolutions and confusions’ (2015) *Computer Law & Security Review* 57; and in the context of injunctions against intermediaries Martin Husovec, ‘Accountable, Not Liable: Injunctions Against Intermediaries’ (2016) TILEC Discussion Paper No. 2016-012, <https://papers.ssrn.com/abstract_id=2773768>

¹¹ Husovec, above n 10, 24 and 66, in a similar distinction calls them “proximate services” and “remote services”.

¹² This article focusses on domain name takedowns in relation to *content* hosted under the website accessible via the domain name, not in relation to the domain *name* itself. In other words, this article does not look at situations where the domain name *as such* infringes rights and leads to a takedown. On the notion ‘takedown’, see Sebastian F Schwemer, ‘On domain registries and unlawful content’ (2018) 26 *International Journal of Law and Information Technology* 273, 277.

content, when it gains the knowledge or awareness of illegal activity.¹³ The latter is the starting point for the liability exemption regime for hosting providers of the E-Commerce Directive and the notice-and-takedown regime.¹⁴ It is also *one* conceptual starting point for a fourth expedited model that relies on specific third parties, so-called “trusted notifiers” or “trusted flaggers”.

A trusted notifier-system refers according to the European Commission to a mechanism, where a privileged notification channel is provided by an intermediary to a third party, which is particularly knowledgeable or has particular expertise to identify unlawful content.¹⁵ Under the respective conditions of these mechanisms, the notifier can request the takedown of, for example, content on a website or the domain name without judiciary involvement in the initial takedown, which in turn regularly is open to judicial review.

Examples for such notifiers range from individual or organized networks of private organizations, civil society organizations and semi-public bodies, to public authorities. Today, there exist several models for non-judicial content takedowns based on trusted notifier-akin arrangements, oftentimes corresponding to industry sectors or types of content.¹⁶ On the European level, for example, the Internal Referral Unit of Europol since July 2015 identifies content related to terrorist and violent extremist propaganda and refers relevant content to Internet service providers.¹⁷ The International Association of Internet Hotlines (INHOPE) provides hotlines for reporting child abuse material and promises quick and effective content removal in collaboration with industry and law enforcement.¹⁸ Related systems exist with specialized civil society organizations or semi-public bodies regarding illegal hate speech, such as racist or xenophobic content.¹⁹ Similar arrangements are also in place at the national level: in Denmark, since 2005, a voluntary content filtering regime is in place for child sexual abuse material, based on a collaboration between the Danish National Police, non-governmental organization Save the Children (Red Barnet) and the majority of Internet access service providers.²⁰ Also online platforms have engaged in voluntary measures, for example, a Code of Conduct countering illegal hate speech by Facebook, Microsoft, Twitter and Youtube

¹³ Cf. Article 14 (1) E-Commerce Directive. According to Article 14(1) lit. b of the E-Commerce Directive, an online hosting platform is shielded from liability for the information stored by a user under the condition that “the provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information.”

¹⁴ On notice and takedown procedures in the context of counterfeit goods see e.g. Knud Wallberg, ‘Notice and takedown of counterfeit goods in the Digital Single Market: a balancing of fundamental rights’ (2017) 12 *Journal of Intellectual Property Law and Practice* 922, 924ff and in the context of ISPs see Clement Salung Petersen and Thomas Riis, ‘Private enforcement of IP law by internet service providers: notice and action procedures’ in Thomas Riis (ed), *User Generated Law, Re-Constructing Intellectual Property Law in a Knowledge Society* (Edward Elgar 2016), 228–251.

¹⁵ See European Commission, ‘Tackling Illegal Content Online’ (Communication) COM(2017) 555 final, 8.

¹⁶ Child sexual abuse content, for example, is prominently addressed. Perplexingly, Moore et al. (2009) found that child sexual abuse websites “are removed much more slowly than almost any other type of unlawful content.” see Tyler Moore, Richard Clayton, Ross Anderson, ‘The Economics of Online Crime’ (2009) 23 *Journal of Economic Perspectives* 3, 16.

¹⁷ See Europol, ‘Europol’s internet referral unit to combat terrorist and violent extremist content’ (Press release, 1 July 2015) <www.europol.europa.eu/newsroom/news/europol-s-internet-referral-unit-to-combat-terrorist-and-violent-extremist-propaganda> accessed 24 October 2018. On the UK schemes related to terrorist material and extreme pornography, see TJ McIntyre, “Internet Censorship in the United Kingdom: National Schemes and European Norms” in Lilian Edwards (ed.), *Law, Policy and the Internet* (Hart Publishing 2019) 291–330.

¹⁸ See European Commission, above n 15, 8.

¹⁹ *Ibid.*

²⁰ So-called ‘netfilterordning’, see <www.ft.dk/samling/20111/almdel/reu/spm/125/svar/851602/1069893.pdf> accessed 24 October 2018 (in Danish). Notably, the evaluation of content is at the discretion of the police and does not have to be based on a court order.

supported by the European Commission.²¹ In the domain name industry, too, the emergence of trusted notifier-models for content-related domain name remedies can be observed.²² Generally, however, there is only very sparse public information available on these arrangements. This makes trusted notifiers a difficult phenomenon to study and to assess their use, accuracy, effectiveness or due process.

On a general level, these models are argued to come with several advantages: compared to notifications by ‘regular’ users, the European Commission notes as benefits of trusted notifiers “higher quality notices and faster take-downs”, refraining, however, from further specifying the quality improvements of notices.²³ In other words, the mechanism is argued to lead to a higher efficiency in the notice-and-takedown regime based on an expedited process. Thus, a principle purpose of implementing such measures is to increase the efficiency of enforcement over ‘regular’ notice-and-takedown channels. Regularly, intermediaries are lacking the competence and resources to evaluate the legality of content²⁴; trusted notifiers, on the other hand, so it is presumed, have exactly that competence.²⁵ In the example of intellectual property (IP) rights infringements, rights holders, for example, have information about protected works and whether they are licensed or not, an information that the intermediary regularly lacks. Husovec argues that “under some circumstances even a right holder can be the sole cheapest-cost avoider”, namely in instances where rights holders have access to a removal interface.²⁶

There is little information about the costs of such trusted notifier arrangements. On the intermediaries’ side, they are likely to depend on existing process and technologies for the handling of notices. On the notifier side, right holders and other parties will regularly only push for such mechanisms, where a faster takedown is perceived beneficial compared to the pre-existing regime. Already today, different parties such as rights holders or interest organizations screen content. In any case, implementation costs are likely smaller compared to e.g. the introduction of automated filtering mechanisms. The incentives for intermediaries to engage in trusted-notifier models can be manifold. As a starting point stands an extension of the notice-and-takedown regime based on the respective legislative frameworks, secondary liability questions and preliminary injunctions.²⁷ Trusted notifier systems might also be able to reduce the risk of court proceedings. Legal uncertainty and other considerations might be a driver too, as examples from the domain name industry will show.

²¹ European Commission, ‘EU Internet Forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online’ (Press release, 3 December 2015) IP/15/6243. See also Lilian Edwards, “‘With Great Power Comes Great Responsibility?’: The Rise of Platform Liability” in Lilian Edwards (ed.), *Law, Policy and the Internet* (Hart Publishing 2019), 271–272.

²² See in detail the examples below.

²³ Empirical research has displayed the varying validity of copyright-related takedown notices, see e.g. JM Urban, J Karaganis and BL Schonfeld, *Notice and Takedown in Everyday Practice* (Berkeley Law 2016).

²⁴ See e.g. Aleksandra Kuczerawy, ‘Intermediary liability & freedom of expression: Recent developments in the EU notice & action initiative’ (2015) 48 *Computer Law & Security Review* 46, 48.

²⁵ Husovec, above n 10, 36, for example, makes the point in context IP rights and “contextual infringements”, where “identification and determination of such infringements requires knowledge of (1) societal circumstances (...) (2) external public information (...) (3) confidential information”, that right holders are better positioned to identify infringing content.

²⁶ Husovec, above n 10, 37.

²⁷ See e.g. Matthias Leistner, Structural aspects of secondary (provider) liability in Europe, *Journal of Intellectual Property Law & Practice*, 2014, Vol. 9, No. 1, pp. 75–90 and the big study.

At the same time, there exist concerns when enforcement arrangements are not based on a legal basis but voluntary codes or contractual arrangements as regards, for example, legitimacy, rule of law, transparency and control mechanisms that prevent excessive or unlawful takedown mechanisms. These concerns are aggravated by a principal–agent relation between intermediary and notifier. Thus, the efficiency gain comes at a cost. These will be discussed in the sections below.

3. Voluntary introduction of trusted notifier models at platform-level by means of soft law

In recent years, the European lawmaker has primarily looked towards online platforms²⁸ –i.e. the application layer– and their role in the regulation of content.²⁹ For these intermediaries, it is regularly technically feasible to make unavailable the actual infringing content *as such*, i.e. without disabling the website as a whole. From a proportionality perspective, the appeal of measures restricted to the concrete infringing content is self-evident.

Self-regulation plays an important role in the playbook for tackling illegal online content on online platforms: following an announcement in its Digital Single Market Strategy mid-term review, in September 2017, the European Commission presented non-binding guidelines and principles for online platforms to increase the proactive prevention, detection and removal of illegal content inciting hatred, violence and terrorism online.³⁰ Its Communication on “Tackling Illegal Content Online, Towards an enhanced responsibility of online platforms” stipulates that regarding the detection and notification of illegal content

“Online platforms should cooperate more closely with **competent national authorities**, by appointing points of contact to ensure they can be contacted rapidly to remove illegal content. To speed up detection, online platforms are encouraged to work closely with **trusted flaggers**, i.e. specialised entities with expert knowledge on what constitutes illegal content. Additionally, they should establish easily accessible mechanisms to allow **users** to flag illegal content and to invest in automatic detection technologies.”³¹

²⁸ The uniform notion “online platform” is not unproblematic: the European Parliament, for example, points towards difficulties of a single definition of online platform, given the variety of platforms. Instead, the Parliament proposes to distinguish and defines platforms (...) “in relevant sector-specific legislation at EU level according to their characteristics, classifications and principles and following a problem-driven approach.” (See European Parliament, Resolution of 15 June 2017 on online platforms and the digital single market (2016/2276(INI)). See also Andrej Savin, ‘Regulating internet platforms in the EU - The emergence of the “Level playing Field”’ (2018) 34 Computer Law & Security Review 1215.

²⁹ In relation to copyright-protected works, the Commission suggested, for example, the introduction of a monitoring obligation for certain information society service providers and effectively introducing a notice and stay-down regime, see above n 4. Similarly, in May 2016, the Commission proposed a rule on (technical) measures to restrict access to harmful content by minors in, see Article 12 of European Commission, ‘Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services in view of changing market realities’ COM(2016) 287 final.

³⁰ European Commission, ‘Security Union: Commission steps up efforts to tackle illegal content online’ (Press release, 28 September 2017) IP/17/3493. Also to be seen in context of European Parliament, above n 28, e.g. point 34. Notably, the press release does not mention copyright-related infringements, whereas the underlying communication refers to copyright-related infringements at several points.

³¹ *Ibid.*, emphasis added. The Communication “strongly recommends” online platforms to prevent the re-appearance of illegal content by using automation technologies, effectively reiterating the Commission’s policy choice for proactive filtering technologies beyond the copyright vertical.

The Commission anticipated online platforms to “proactively implement the guidelines” and aimed at tracking the progress until May 2018 before considering to introduce accompanying legislative measures complementing the existing framework. Already on 1 March 2018, the Commission followed up with a soft law tool: Recommendation (EU) 2018/334 on measures to effectively tackle illegal content online.³² This Recommendation follows a horizontal approach addressing both all types of illegal content (Chapter 2) and terrorist content (Chapter 3) and is directed towards both Member States and hosting service providers.³³ According to Chapter 1, para. 4 lit. b, ‘illegal content’ “means any information which is not in compliance with Union law or the law of a Member State concerned”. The broad scope is to be seen towards the horizontal scope of the E-Commerce Directive³⁴ but is noteworthy towards the background of the Commission’s earlier statements that different types of illegal content would “require different policy approaches in respect of notice-and-action procedures.”³⁵

In the Recommendation, the introduction of systems akin to trusted notifiers is foreseen in two dimensions: for non-state private actors and for public authorities. In line with the Commission’s preceding Communication, it suggests:

“Cooperation between hosting service providers and trusted flaggers should be encouraged. In particular, fast-track procedures should be provided to process notices submitted by **trusted flaggers**.”³⁶

Such trusted flagger is defined in point 4 lit. g as “an individual or entity which is considered by a hosting service provider to have particular expertise and responsibilities for the purposes of tackling illegal content online”. Secondly, the Recommendation also foresees privileged notice channels for public authorities, namely in point 23 where it stipulates that “[f]ast-track procedures should be provided to process notices submitted by competent authorities.”

Let us recall that the Recommendation is directed towards Member States and hosting service providers. Interestingly, the Commission envisions a –potentially considerable– expansion of scope of the proposed solutions though: in Recital 15, it recalls that online hosting service providers play a “particularly important role in tackling illegal content online” and that the Recommendation “therefore primarily relates to the activities and responsibilities of those providers”. The Recital continues that “[h]owever, where appropriate, the recommendations made can also be applied, mutatis mutandis, in relation to other affected online services providers.” The Recommendation refrains from defining these ‘online service providers’ and providing guidance or examples, potentially leaving a wide discretion to Member States and affected services. It seems, however, that the envisioned model is the Commission’s preferred

³² Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online C/2018/1177 [2018] OJ L 63/50.

³³ Such ‘hosting service provider’ is defined in Chapter 1 para. 4 lit. a as “a provider of information society services consisting of the storage of information provided by the recipient of the service at his or her request, within the meaning of Article 14 of Directive 2000/31/EC, irrespective of its place of establishment, which directs its activities to consumers residing in the Union”.

³⁴ See Recital 14 of Recommendation (EU) 2018/334.

³⁵ European Commission, ‘Online Platforms and the Digital Single Market, Opportunities and Challenges for Europe’ (Communication) COM(2016) 288 final.

³⁶ Chapter 2, point 25 of Recommendation (EU) 2018/334, emphasis added.

solution for potentially a wide variety of Internet intermediaries. I will come back to this observation when looking at the specific intermediary case of domain registries.

A Recommendation is by its very nature a weak instrument vis-à-vis a Directive or Regulation. In relation to terrorist content, Jean-Claude Juncker justified a proposal for Regulation on terrorist content in its State of the Union speech from September 2018 with the observation that the recommendations from March 2018 “have brought positive results, overall progress has not been sufficient.”³⁷ In relation to other forms of illegal content, no follow-up legislation has been proposed so far, but another assessment (and potentially opening) of the E-Commerce Directive’s regime during the 2019–2024 Commission cycle in form of a Digital Services Act appears to be on the legislative agenda.³⁸

4. Mitigating issues of privatized enforcement I: should we trust a notifier?

Privatized enforcement has generally been associated with a variety of issues related to, for example, the rule of law, legal certainty, accountability, democracy deficit, presumption of innocence, right to due process, and potentially right to privacy and freedom of speech and communication.³⁹ A central issue of trusted notifier relates to their legitimacy.⁴⁰ One approach defines legitimacy as a “(...) generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed systems of norms, values, beliefs and definitions.”⁴¹ In my view, legitimacy in the context of these trusted notifier-models, is partly about being representative, something which can be seen in some of the referred examples: public authorities, for example (such as police), enjoy general trust and gained legitimacy in that they form an essential part of the state and are subject to (democratic) oversight. Another example are organizations that can be specifically representative on a knowledge level (e.g. children protection organisation) or given a special legitimation (e.g. rights holders’ societies). Likely it is often representatives of involved parties that are especially knowledgeable, think for example of an industry organization. In those instances, however, a principal–agent problem is inherent: the agent, i.e. the trusted notifier and in the example, the industry organization, has a primary interest in making infringing content unavailable. Whereas we can assume that the principal, i.e. the intermediary, shares a similar goal, the interests of the intermediary are likely to embrace more diverse and differentiated objectives that go beyond the mere inaccessibility of the respective, allegedly illegal content. This could lead to situations, where the equilibrium between interests balances out at the principal-level, whereas it is more biased towards a certain result at the agent-level. Child protection organisations or rights holders’ societies, for example, could be more willing to accept a higher

³⁷ European Commission, ‘State of the Union 2018: Commission proposes new rules to get terrorist content off the web’, Press release, IP/18/5561, Strasbourg, 12 September 2018 < http://europa.eu/rapid/press-release_IP-18-5561_en.htm>

³⁸ See leaked note: European Commission, *Digital Service Act note DG Connect June 2019*, < <https://cdn.netzpolitik.org/wp-upload/2019/07/Digital-Services-Act-note-DG-Connect-June-2019.pdf>>

³⁹ See e.g. European Digital Rights (EDRI), ‘Human Rights and privatised law enforcement’ (25 February 2014), 3 <https://edri.org/wp-content/uploads/2014/02/EDRI_HumanRights_and_PrivLaw_web.pdf>; Angelopoulos et al., above n 10. For a discussion of legitimacy of institutions in the context of ICANN, see Pierre Mounier, ‘Internet governance and the question of legitimacy’, in Eric Brousseau, Meryem Marzouki and Cécile Méadel, *Governance, Regulations and Powers on the Internet* (Cambridge University Press, 2012) 170–185.

⁴⁰ For a discussion on legitimacy in the Internet Governance context see also ‘Utility and Legitimacy Issues’ in Lee A Bygrave, *Internet Governance by Contract* (Oxford University Press, 2015) 133–136.

⁴¹ Mark C Suchman, ‘Managing legitimacy: Strategic and institutional approaches’ (2005) 20 *Academy of Management Rev.* 571, 574 cited in Bygrave, above n 40, 134.

false positive rate because their main interest is the unavailability of certain content. In other words, situations where the agent wants to enforce “as much as possible”, whereas the principal wants to enforce “as much as necessary”.⁴² Additionally, trusted notifier-models inherently come with a stark information asymmetry: the whole point of these arrangement is that the agent, i.e. the notifier, is more knowledgeable than the principal. Thus, for the presumption effect of notices by trusted notifiers to be legitimate, certain safeguards are necessary in order to mitigate the principal-agent problem.

But not only legitimacy depends on the assessment of whether the models involve non-state private actors or public authorities: this distinction impacts on the set of rules with which the models need to comply. Many constitutional duties or human rights only directly obligate States, not private companies.⁴³ Thus, private entities “can impose (and be “encouraged” to impose) restrictions on access to information without being subject to the constitutional or international law constraints that apply to state limitations of the right to freedom of expression”.⁴⁴ This makes for a difficult assessment in instances where arrangements between private actors are facilitated or endorsed by public authorities. In the context of blocking decisions by private actors this can, in the Council of Europe’s (CoE) Commissioner for Human Rights’ view, lead

“(…) to a situation in which access to selected websites for the vast majority of the population is determined, not on the basis of public law, but on the basis of decisions by private-law entities that are not directly subject to human rights law. In particular, ISPs can stipulate in their general terms and conditions that they are free to decide, by themselves, whether to block access to specific sites if they deem those sites (at their own discretion) to be contrary to company policies.”⁴⁵

This issue is of course not restricted to expedited trusted notifier-models. The CoE’s Committee of Ministers, for example, addressed this issue in its *Declaration on ICANN, human rights and the rule of law* from 2015 arguing that human rights and fundamental freedoms of internet users “should prevail over the general terms and conditions of service of private-sector Internet companies and the technical mandates of specialised entities, such as the Internet Corporation for Assigned Names and Numbers (ICANN).”⁴⁶ In a similar vein, albeit more reserved, also the European Commission in its Recommendation (EU) 2018/334 foresees that

“(…) decisions taken by hosting service providers to remove or disable access to content which they store **should take due account of the fundamental rights** and the **legitimate interests** of their users as well as of the central role which those providers tend to play in

⁴² Borrowed from Husovec, above n 10, 50.

⁴³ Cf. European Convention of Human Rights, EU Charter of Fundamental Rights, International Covenant on Civil and Political Rights. See Christina Angelopoulos et al. above n 10; see also United Nations, Office of the High Commissioner for Human Rights, ‘An Open Letter to States Concerning an International Legally Binding Instrument on Business and Human Rights’ (2018) <www.ohchr.org/Documents/Issues/Development/IEDebt/OpenLettertoStates_1Oct2018.pdf> accessed 13 December 2018; Emily B. Laidlaw, *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility* (Cambridge University Press, 2015) 84–85.

⁴⁴ Council of Europe, Commissioner for Human Rights, above n 6, 16.

⁴⁵ *Ibid.*, 70.

⁴⁶ Council of Europe, Declaration of the Committee of Ministers on ICANN, human rights and the rule of law, (Adopted by the Committee of Ministers on 3 June 2015 at the 1229th meeting of the Ministers’ Deputies), point 4.

facilitating public debate and the distribution and reception of facts, opinions and ideas in accordance with the law.”⁴⁷

Thus, the also Recommendation foresees an application of fundamental rights intermediaries to some extent. Given the absence of specific secondary legislation, this section focuses on sketching the outlines of the framework by relying on the European Commission’s Recommendation and various documents by the Council of Europe (CoE) and the United Nations (UN), which have addressed associated drawbacks and put forward suggestions for their mitigation.

The necessity of a legal basis for non-judicial enforcement regimes

Especially in the context of non-judicial enforcement and the involvement of state actors, the rule of law becomes relevant.⁴⁸ In March 2018, the CoE Committee of Ministers adopted a Recommendation on the roles and responsibilities of internet intermediaries.⁴⁹ In its appendix on the Guidelines for States on actions taken vis-à-vis internet intermediaries, it recommends that:

“Any request, demand or other action **by public authorities** addressed to internet intermediaries that interferes with human rights and fundamental freedoms shall be **prescribed by law**, exercised within the limits conferred by law and constitute a necessary and proportionate measure in a democratic society. States should **not exert pressure** on internet intermediaries through non-legal means.”⁵⁰

Especially, the second sentence in the cited paragraph is interesting: my reading is that the Committee of Ministers is quite aware of the non-legislative influence that public actors enjoy. To me, the use of a Recommendation, which is non-binding but encouraging to comply under the threat of potential legislative measures in case of non-compliance, represents almost a prototype of the situation, which the CoE discourages. Thus, the European Commission’s push regarding privileged channels based on a soft law tool⁵¹ seems at odds with the CoE’s guidelines. At the same time, it is questionable whether actions based on the Commission’s (more or less) clear input regarding non-law based enforcement mechanisms would even qualify as voluntary actions by private actors.⁵² In light of new modes of governance and in the context of the Commission’s Better Regulation Agenda, non-law-based interventions operate side by side with, or instead of, legislative interventions.⁵³ Against this background, I find it unconvincing to qualify models that are based on soft law as voluntary private actions.

⁴⁷ Recital 13 of Recommendation (EU) 2018/334, last sentence, emphasis added.

⁴⁸ See also Council of Europe, Commissioner for Human Rights, above n 6, 43.

⁴⁹ Council of Europe, Committee of Ministers, CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries (Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers’ Deputies). The Committee is the primary decision body of the CoE; its recommendations are not binding to the Member States.

⁵⁰ Appendix, point 1.1.1, emphasis added.

⁵¹ See Article 288 TFEU.

⁵² Council of Europe, Commissioner for Human Rights, above n 6, 14.

⁵³ See David M Trubek, Patrick Cottrell and Mark Nance, “Soft Law”, “Hard Law” and EU Integration’ in Gráinne de Búrca and Joanne Scott (eds), *Law and Governance in the EU and the US* (Hart Publishing 2006) 65; Rob van Gestel and Hans-W. Micklitz, ‘Revitalizing Doctrinal Legal Research in Europe: What About Methodology?’ in Ulla Neergaard, Ruth Nielsen and Lynn Roseberry (eds), *European Legal Method - Padoxes and Revitalisation* (DJØF Publishing 2011) 46; Stephen Weatherill, ‘The Challenge of Better Regulation’ in Stephen Weatherill (ed), *Better*

The situation looks somewhat different for private actors. In 2011, the United Nations' Human Rights Council's Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, recommended (private) intermediaries in a catalogue to

“only implement restrictions to these rights after judicial intervention; be transparent to the user involved about measures taken, and where applicable to the wider public; provide, if possible, forewarning to users before the implementation of restrictive measures; and minimize the impact of restrictions strictly to the content involved. Finally, there must be effective remedies for affected users, including the possibility of appeal through the procedures provided by the intermediary and by a competent judicial authority.”⁵⁴

In my reading, the Special Rapporteur's position from 2011 leaves little room for restrictions based on non-judicial takedowns arrangements. In a similar vein, in 2014, also the CoE's Commissioner for Human Rights recommended in an issue paper regarding blocking and filtering that:

“Member states should ensure that any restrictions on access to Internet content affecting users under their jurisdiction are based on a strict and predictable **legal framework** regulating the scope of any such restrictions and affording the guarantee of judicial oversight to prevent possible abuses. In addition, **domestic courts must** examine whether any blocking measure is necessary, effective and proportionate, and in particular whether it is targeted enough so as to impact only on the specific content that requires blocking.”⁵⁵

Furthermore, the Commissioner underlined that states “should not rely on or encourage private actors (...) to carry out blocking outside a framework meeting the criteria described above.”⁵⁶ Thus, also in a private actor-setting, the predictability of the legal framework and judicial involvement regarding the assessment of specific blocking measures constitute key concerns. In other words, also regarding private action, the European Commission's soft law approach appears to be at odds with the CoE's Commissioner's view. More recently, however, in 2018, the CoE's Committee of Ministers, has once more visited their stance on responsibilities of internet intermediaries. It stipulates that

“Any interference by intermediaries with the free and open flow of information and ideas (...) should be based on **clear and transparent policies** and be limited to specific legitimate purposes, such as restricting access to illegal content, as determined either by law or by a judicial authority or other independent administrative authority whose decisions are subject to judicial review, **or** in accordance with their **own content-restriction policies or codes of ethics**, which **may include flagging mechanisms**.”⁵⁷

Thus, it appears that the recent position of the CoE Committee of Ministers is much more favorable of intermediaries' own content takedown policies than in earlier positions.

Regulation (Hart Publishing 2007) 47; Benoît Frydman, Ludovic Hennebel and Gregory Lewkowicz, “Co-regulation and the rule of law”, in Eric Brousseau, Meryem Marzouki and Cécile Méadel (eds.), *Governance, Regulations and Powers on the Internet* (Cambridge University Press, 2012) 133–150.

⁵⁴ United Nations, General Assembly, Human Rights Council, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue, 16 May 2011, A/HRC/17/27, point 47, 14.

⁵⁵ Council of Europe, Commissioner for Human Rights, above n 6, point 16, emphasis added.

⁵⁶ *Ibid.*, point 17, emphasis added.

⁵⁷ Council of Europe, Committee of Ministers, above n 49, Section 2 on Responsibilities of internet intermediaries with respect to human rights and fundamental freedoms, point 2.1.3, emphasis added.

Towards the background of large amounts of unlawful content and intermediaries' gate-keeper function, this seems a practical and result-oriented view. Yet, there exist concerns as to the privatization of enforcement. The question is then, how these concerns can be mitigated.

5. Mitigating issues of privatized enforcement II: when to trust a notifier?

In the following, I will look at a few aspects, where the mitigation of negative effects is especially relevant: the selection of trusted notifiers, transparency and control mechanisms, and the risk of over-removal. The section takes its starting point in the only instrument directly addressing trusted notifiers at the EU-level at this time, Recommendation (EU) 2018/334.

Selection of trusted notifiers

When we consider the legitimacy aspects and the inherent principal-agent problem discussed above, one central aspect relates to the selection of trusted notifiers. Who should be able to appoint these actors and under what standards? Conceptually, the selection of a trusted notifier is a positive discrimination of certain actors. In point 26, the Recommendation encourages online platforms "to publish clear and objective conditions for determining which individuals or entities they consider as trusted flaggers." Thus, the Commission effectively proposes that online platforms themselves determine who is to be seen a trusted flagger. The selection conditions are further specified in point 27, where it is held that they "should aim to ensure that the individuals or entities concerned have the necessary expertise and carry out their activities as trusted flaggers in a diligent and objective manner, based on respect for the values on which the Union is founded."⁵⁸ The Recommendation further stipulates in Recital 29 that such "[c]ooperation should only be open to individuals and entities which respect the values on which the Union is founded as set out in Article 2 of the Treaty on European Union and meet certain appropriate conditions, which should moreover be clear and objective and be made publicly available." There is special emphasis on the respect of the values on which the EU is founded, namely "human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail."

Thus, there are indeed some broad suggested selection criteria. Their very existence underlines the importance of the selection process. Effectively, however, it is left to the individual platform to refine the abstract set of values and define the concrete criteria that a trusted notifier needs to meet. Another, related question is how to select trusted notifiers in cross-border situations, which are inherent to these platforms. Finally, there is also little guidance on whether the selection of trusted notifiers is a one-time accreditation process or rather an iterative process whether the privilege is monitored and can be withdrawn, for example when quality of notices is low or abuse is taking place.

Transparency and control mechanisms

Public and administrative law and principles do as a starting point not infuse private mechanisms. There exists no requirement for private measures to offer procedural due process safeguards comparable to those usually related to the granting of a preliminary injunction.

⁵⁸ Chapter 2, point 27 of Recommendation (EU) 2018/334.

Considering the empirical evidence on inaccurate takedown notices,⁵⁹ this of concern. One mitigating means is the right to counter notices, e.g. prior to takedown and within a certain timeframe.⁶⁰

Recommendation (EU) 2018/334 contains several general safeguards, e.g. on transparency (points 16–17) and on the protection against abusive behavior, such as the submission of notices and counter-notices in bad faith (point 21), which relate both to private and public notifying parties. Additionally, the Recommendation foresees safeguards on the side of the hosting service provider in point 19 to “(...) act in a diligent and proportionate manner in respect of content that they store, in particular when processing notices and counter-notices and when deciding on the possible removal of or disabling of access to content considered to be illegal”.

Risk of rubber-stamping and over-removal

Trusted notifiers models are appealing because the system, at its surface, solves a problem related to the competence to decide often difficult legal questions on the legality of content. This is recognized by Recital 29 of Commission Recommendation (EU) 2018/334 which stipulates that notices submitted by such parties should be treated “(...) as a matter of priority and with an appropriate degree of confidence as regards their accuracy”. It is, however, unclear what an appropriate degree would be and to what extent and how an intermediary is to evaluate the accuracy.

In this principal–agent situation there exists a risk of rubberstamp-accepting takedown notices by trusted notices leading to over-compliance and excessive content takedown, in situations where intermediaries, for example, seek to avoid legal repercussion for failure to act on a received notice.⁶¹ Furthermore, for example in the case IP rights, if there exist no sanctions for over-removal, rights holders “will rely on automation” for any kind of infringement⁶², even in instances, where automation is expected to lead to high false-positives. Empirical evidence underlines the tendency of over-removal by intermediaries.⁶³ This concern is aggravated towards the background that government agencies, interest organizations and rightholders are more likely to file a lawsuit than the infringer and when the intermediary possesses limited resources or competences to evaluate the situation.⁶⁴ Given the privileged takedown channel, safeguards are necessary to ensure that a notifier continues to deliver high quality notices and false positives are held low in the long run. It has been suggested before, for example, that “automated submission should be available only under the condition of extremely low margin of errors and allow intermediaries to temporarily terminate the access to it when the condition

⁵⁹ See Daphne Keller, ‘Empirical Evidence of “Over-Removal” by Internet Companies under Intermediary Liability Laws’ (Stanford University, The Center for Internet & Society, 12 October 2015) <<http://cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws>> accessed 23 October 2018.

⁶⁰ See e.g. Husovec, above n 10, 73, with reference to Jennifer Urban and Laura Quilter, ‘Efficient Process or “Chilling Effects”? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act’ (2006) 22 Santa Clara Computer and High Technology Law Journal 621, 689.

⁶¹ See also Council of Europe, Commissioner for Human Rights, above n 6, 70.

⁶² Husovec, above n 10, 45.

⁶³ See Keller, above n 59.

⁶⁴ See also Internet & Jurisdiction, ‘Domains & Jurisdiction Program: Cross-Border Domain Suspension, Problem Framing’ (May 2017) 5 <<https://www.internetjurisdiction.net/uploads/pdfs/Papers/Domains-Jurisdiction-Program-Paper.pdf>> accessed 13 December 2018.

is violated”.⁶⁵ Yet, there appear to be few safeguards in the European playbook that mitigate this risk at this point.

In conclusion, it seems that the European Commission’s approach in the Recommendation leaves many questions unaddressed and provides only a vague, and non-binding template.

6. Turning towards an odd example: content-related domain name takedowns

Let us now turn towards the case of content or use-related domain name takedowns.⁶⁶ Domain names facilitate the access to content by translating IP addresses into mnemonic codes. These situations are special in that not the allegedly infringing content is taken down, but rather the accessibility to the website via a domain name *as such* disabled. The major difference of domain name vis-à-vis platform takedowns relates thus inter alia to proportionality in that the access to the website via the respective domain name as a whole (including e-mail) disappears. Looking at domain names might seem odd at first, when considering the regulatory focus on online platforms and other intermediaries, which are closer to the content. Yet, this section will reveal that trusted notifier-akin regimes indeed are used in the industry but appear to be overlooked by the regulator.

6.1 Traditional sources for domain name takedowns

Before looking at voluntary forms of content-related domain name takedowns, it is helpful to briefly sketch traditional sources for domain name takedowns: in the majority cases instances the takedown of a domain name is related to the domain name *as such*, such as in the case of trademark infringements.⁶⁷ Many country-code top-level (ccTLD) domain name registries and generic top-level (gTLD) domain name registries have different processes and rules in place, oftentimes based on the ICANN framework. Already in 2001, for example, the World Intellectual Property Organisation (WIPO) issued its ‘ccTLD Best Practices for the Prevention and Resolution of Intellectual Property Disputes’.⁶⁸ The takedown of domain names related to *its use*, on the other hand, regularly finds its basis either in court orders or in specific legislation.⁶⁹

First and foremost, a domain registry can be obliged to suspend a domain name on the basis of a court order subject to the respective national procedural rules. In Denmark, for example, the national ccTLD registry has received 1.025 court orders to suspend domains in

⁶⁵ Husovec, above n 10, 73.

⁶⁶ On domain registries and their role regarding unlawful content see Maarten Truyens and Patrick Van Eecke, ‘Liability of Domain Name Registries: Don’t Shoot the Messenger’ (2016) 32 Computer Law & Security Review 327 and Schwemer, above n 12.

⁶⁷ An example would be the domain name <www.rolex-watches.dk> by a registrant that is not entitled to use the trademark.

⁶⁸ World Intellectual Property Organisation (WIPO), ‘ccTLD Best Practices for the Prevention and Resolution of Intellectual Property Disputes’ (Version 1: June 20, 2001), <www.wipo.int/amc/en/domains/bestpractices/> accessed 13 December 2018.

⁶⁹ On the legislative frameworks and practices regarding blocking, filtering and takedown of illegal content in the forty-seven Council of Europe Member States, see the comprehensive comparative study commissioned by the Council of Europe: Swiss Institute of Comparative Law, ‘Comparative Study on Blocking, Filtering and Take-Down of Illegal Internet Content’ (Lausanne, 2015).

2017 and 1.021 in the first six months of 2018 alone.⁷⁰ Another noteworthy case is the seizure of more than 4.500 domain names of websites which sold counterfeit products, in a concerted effort by Europol in the operation ‘In our Sites’.⁷¹ Court orders, directed both at domain registries and/or other intermediaries, have become a frequently used measure by both rightholders and public prosecutors.

There exist also other secondary, partly sectoral, legislation or proposals, which could provide a basis for the use-related takedown of a domain name. The European Commission’s proposal for a Consumer Protection Cooperation Regulation from May 2016⁷², for example, stipulated in Article 8 (2) lit. 1 that the competent national authority shall have at least the power to inter alia “close down a website, domain or similar digital site, service or account or a part of it, including by requesting a third party or other public authority to implement such measures” under certain conditions.⁷³ Directive 2011/92/EU on combating sexual abuse and sexual exploitation of children and child pornography in Article 25(1) introduced an obligation on Member States to “take the necessary measures to ensure the prompt removal of web pages containing or disseminating child pornography hosted in their territory and to endeavor to obtain the removal of such pages hosted outside of their territory.”⁷⁴ In relation to terrorist content, Directive (EU) 2017/54⁷⁵ regulates in Article 21 (1) and (2) on “Measures against public provocation content online” that Member States shall take the necessary measures to ensure the prompt removal or blocking of such content. Also, national legislation is in some instances addressing domain name takedowns (e.g. Switzerland⁷⁶ or proposal in Denmark⁷⁷).

⁷⁰ DK Hostmaster for example provides the Danish police and other public authorities with guidelines on how to request data from the registry, see <www.dk-hostmaster.dk/en/release-data-police> accessed 13 December 2018. The legal basis for preliminary injunctions in Denmark is found in the general rules, section 314 of the Danish Administration of Justice Act.

⁷¹ See Europol, ‘Operation In Our Sites (IOS)’ <www.europol.europa.eu/activities-services/europol-in-action/operations/operation-in-our-sites-ios> accessed 24 October 2018.

⁷² European Commission, ‘Proposal for a Regulation of the European Parliament and of the Council on cooperation between national authorities responsible for the enforcement of consumer protection laws’ COM(2016) 283 final.

⁷³ Notably in the Council’s general approach, references in Article 3 to website and domain names were deleted. The CPC Regulation (Regulation (EU) 2017/2394 on cooperation between national authorities responsible for the enforcement of consumer protection laws and repealing Regulation (EC) No 2006/2004 [2017] OJ L 345/1) can be seen as a tendency for how the takedown at the domain-name level is coming into the focus. Also in the proposed Regulation laying down rules and procedures for compliance with and enforcement of Union harmonisation legislation on products, COM(2017) 795 final, from 19 December 2017, a basis for takedown by market surveillance authorities.

⁷⁴ See e.g. Article 25 of Directive 2011/92/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA [2011] OJ L 26/1.

⁷⁵ Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA [2017] OJ L 88/6.

⁷⁶ The case of the regulation of <.ch> and <.swiss> top level-domains is interesting, as it constitutes one of the most comprehensive regulatory frameworks that also addresses the use of domain names for phishing or distribution of malware. See Bundesamt für Kommunikation (BAKOM), ‘Wirksamere Massnahmen zur Bekämpfung von Cyberkriminalität auf .ch- und .swiss-Websites’ (press release, 15 September 2017) <www.bakom.admin.ch/bakom/de/home/das-bakom/medieninformationen/medienmitteilungen.msg-id-68117.html> accessed 24 October 2018.

⁷⁷ In Denmark, the government in 2016 proposed an amendment to the Administration of Justice Act that would have allowed police to take down domains for any violation of its penal code. After severe public backlash the proposal was amended to only cover certain violations related to terror.

Let it suffice here to note that several more or less specific rules are in place for the takedown or blocking of content.

6.2 Emergence of mechanisms for content-related takedown of domain names

In the, up until recently, absence of regulation by traditional governmental actors in the domain name space, it is interesting to consult the stance of the multi-stakeholder governance organization Internet Corporation for Assigned Names and Numbers (ICANN). ICANN has established a system for the resolution of disputes stemming from abusive registration and use of domain names *as such* in its Uniform Domain-Name Dispute Resolution Policy (UDRP)⁷⁸, which is applied by registrars. It does not contain redress related to the unlawful activity on the underlying website. In 2015, ICANN's Chief Contract Compliance Officer stated:

“Allow me to say this clearly and succinctly – ICANN is not a global regulator of Internet content (...) ICANN was never granted, nor was it ever intended that ICANN be granted, the authority to act as a regulator of Internet content.”⁷⁹

This position seems to not be that clear-cut anymore. At recent ICANN meetings, for example, questions on DNS and content regulation have been featured prominently on the agenda.⁸⁰ Bridy (2017) comments that ICANN supports “voluntary enforcement agreements between DNS intermediaries and right holders (...).”⁸¹

Regularly, given the contractual relations between registry, registrar and registrant, it is these contracts that provide the basis for the takedown of domain names. The respective domain registry can regulate, to a varying degree, mechanisms in their terms of conditions, which every registrant accepts when entering into a contract with the registry. Thus, an indication for the industry's stance can also be derived from to what extent there exists a contractual basis for the takedown of a domain name for content- or use-related reasons. Brenden et al. (2017) examined the registrars' terms of service of 74 ICANN contracted parties (accounting for 90% of all gTLD domain registrations worldwide) for morality clauses that give domain registrars the right to cancel domains for content-related reasons.⁸² In their empirical study they found that around 59% of registrars employ such clause. Also, some ccTLDs appear to reserve the right to terminate the domain name agreement with a registrant for use- or content-related reasons.⁸³ There exists however, little insight as to whether these provisions – and if so, in what situations – have been used.

⁷⁸ Available at <www.icann.org/resources/pages/udrp-rules-2015-03-11-en> accessed 13 December 2018.

⁷⁹ Allen R Grogan, ‘ICANN Is Not the Internet Content Police’, (ICANN Blog, 12 Jun 2015) <www.icann.org/news/blog/icann-is-not-the-internet-content-police> accessed 24 October 2018.

⁸⁰ See e.g. ICANN57 High Interest Topics Review Page, <<https://meetings.icann.org/en/hitreviews>>

⁸¹ Annemarie Bridy, ‘Notice and Takedown in the Domain Name System: ICANN's Ambivalent Drift into Online Content Regulation’ (2017) 20 <https://papers.ssrn.com/abstract_id=2920805>

⁸² Brenden Kuerbis, Ishan Mehta, Milton Mueller, ‘In Search of Amoral Registrars: Content Regulation and Domain Name Policy’ (2017) Internet Governance Project White Paper, Georgia Tech <www.internetgovernance.org/wp-content/uploads/AmoralReg-PAPER-final.pdf> accessed 12 March 2018; see section 3.18 of ICANN Registrar Accreditation Agreement. See also Caroline Briceux, ‘Regulating Online Content through the Internet Architecture’ (2016) Journal of Intellectual Property, Information Technology and Electronic Commerce Law 229.

⁸³ The terms and conditions of the Austrian and German ccTLD registry, nic.at and DENIC, for example, do not provide directly clauses related to a termination of contract based on the content accessible under the domain name. Interestingly, however, both terms refer to a termination for “important reasons” (“wichtiger Grund”) and come with a list of examples, indicated by “in particular” (“insbesondere”).

Today, there exist different models for the non-judicial takedown, some of which are addressing websites as such whereas others target specific forms of content (e.g. related to *ordre public* such as child abuse material, terrorism, vis-a-vis counterfeit or IP violations).

There are examples of trusted notifier-akin arrangements between national ccTLD registries and public authorities. In the United Kingdom, for example, since 2014 the police and several other law enforcement agencies function akin to a trusted notifier to the local ccTLD registry, NOMINET.⁸⁴ Between November 2016 and October 2017, NOMINET received more than 16.000 takedown requests, whereof as many as 13.000 stemming from the UK's Police Intellectual Property Crime Unit (PIPCU).⁸⁵

There is no information available whether or to what extent this model involving a public authority is based on a legal basis, as recommended by the CoE Committee of Ministers.⁸⁶

Takedown arrangements at the registry-level are not restricted to public agencies, however. There are also instances, where gTLD registries collaborate with private actors around trusted notifier models. In the United States, a trusted notifier scheme regarding copyright is practiced and envisioned by the Motion Picture Association of America (MPAA) and the Recording Industry Association of America (RIIA) respectively.⁸⁷ In 2016, the domain registries Donuts and Radix which represent approximately 200 generic top-level domains such as <.movie> and <.tech>, introduced a trusted notifier regime with MPAA.⁸⁸ Under this model "the MPAA can request the registries to take action against a domain name by presenting evidence that it has evidence that the domain name is being used for "clear and pervasive copyright infringement," and that it has first attempted to contact the registrar and hosting provider for resolution."⁸⁹ The participating registries are required to respond within 10 days. It is being pointed out by the Electronic Frontier Foundation that there is no requirement that the registrant is notified of the takedown.⁹⁰ In a similar vein, Bridy (2017) points towards various risks such as participating registries defaulting "to a rubber stamp approach."⁹¹

Another example for a broad private industry initiative is the Healthy Domain Initiative. In 2017, the Domain Name Association, an industry association consisting of several ccTLD, notably the Canadian and the Dutch ccTLD registry, and major gTLD registries as

⁸⁴ NOMINET, 'Nominet formalises approach to tackling criminal activity on .UK domains' (3 April 2014) <www.nominet.uk/nominet-formalises-approach-to-tackling-criminal-activity-on-uk-domains/> accessed 24 October 2018.

⁸⁵ World Trademark Review, 'Domain name suspensions in ".uk" TLD double as PIPCU takedown requests surge' (15 November 2017) <www.worldtrademarkreview.com/Blog/Detail.aspx?g=bd1a428f-053b-4a15-b978-994b6d7aaa2a> accessed 24 October 2018. See detailed analysis by Bridy, above n 82.

⁸⁶ See section 4 above.

⁸⁷ Bridy, above n 82, 20.

⁸⁸ Motion Picture Association of America, 'MPAA/Radix Partnership Highlights Momentum Behind Voluntary Initiatives' (Press release, 13 May 2017) <www.mpa.org/new-mpaaradix-partnership-highlights-the-continuing-momentum-of-voluntary-initiatives/> accessed 23 October 2018. Motion Picture Association of America, 'Donuts and the MPAA establish new partnership to reduce online piracy' (Press release, 9 February 2016) <www.mpa.org/wp-content/uploads/2016/02/Donuts-and-MPAA-Establish-New-Partnership-2.9.16.pdf> accessed 23 October 2018.

⁸⁹ Electronic Frontier Foundation, 'Which Internet registries offer the best protection for domain owners?' (27 July 2017) 5 <www.eff.org/files/2017/08/02/domain_registry_whitepaper.pdf> accessed 23 October 2018.

⁹⁰ Ibid 6.

⁹¹ Bridy, above n 82, 20.

well as registrars, proposed its ‘Registry / Registrar Healthy Practices’.⁹² Critically received⁹³, the practices consists of recommendations relating to online security abuse such as malware, phishing and pharming; child abuse mitigation systems (e.g. trusted notifier systems); “rogue” online pharmacies; and voluntary third party handling of copyright infringement cases. In the context of child abuse, one of the primary recommendations relates to the establishment of trusted notifier systems, where “a party is pre-vetted (e.g. NCMEC, IWF, INHOPE) and recognized by the contracted party as capable of providing the relevant and complete evidence needed to take action against the registrant.”⁹⁴ In the context of copyright infringement cases, the practices suggest the construction of a “voluntary framework for copyright infringement disputes, so copyright holders could use a more efficient and cost-effective system for clear cases of copyright abuse other than going to court and registries and registrars are not forced to act as “judges” and “jurors” on copyright complaints”.⁹⁵

Also the European ccTLD registry EURid, administering <.eu>, has as recently as June 2018 announced a collaboration with a private actor: a Memorandum of Understanding with the International AntiCounterfeiting Coalition (IACC).⁹⁶ The content of the MoU is, however, secret, and the parties did not respond for this research whether the agreement contains the establishment of a privileged access channel akin to trusted notifiers.

Industry-led solutions in form of DNS-governance self-regulation in this sphere are partly welcomed.⁹⁷ Others see in these models “a failure of the private governance processes associated with administration of the DNS.”⁹⁸ There also exists evidence for the unjustified seizure of websites, based on information provided by the notifier.⁹⁹ The multi-stakeholder policy network Internet & Jurisdiction notes that in relation to trusted notifier-models, “evaluations by notifiers are often established without sufficiently clear procedures or mechanisms for redress or may be based on the laws of only one particular country or the interest of trade associations.”¹⁰⁰ As noted above, the Commission’s Recommendation from 1 March 2018, does not directly address trusted notifier models outside hosting platforms.

7. Discussion

In this article, I have looked at the trend to rely on non-judicial takedown mechanisms in form of trusted notifier-models for unlawful websites and content. Whereas there is generally little empirical data available, we know that the European regulator –at least in the sphere of online platforms– encourages these ‘voluntary’ privatized enforcement measures by intermediaries. In

⁹² Domain Name Association, ‘Healthy Domain Initiative’ (August 2017) <<https://thedna.org/healthy-domain-practices/>> accessed 24 October 2018.

⁹³ See e.g. Jeremy Malcolm and Mitch Stolz, ‘Healthy Domains Initiative Isn’t Healthy for the Internet’ (Electronic Frontier Foundation, 9 February 2017) <<https://www.eff.org/deeplinks/2017/02/healthy-domains-initiative-censorship-through-shadow-regulation>> accessed 24 October 2018.

⁹⁴ Domain Name Association, above n 93, 5.

⁹⁵ Ibid 5.

⁹⁶ EURid, ‘EURid and IACC Team Up to Fight Cybercrime’ (Press release, 27 June 2018) <<https://eurid.eu/en/news/eurid-and-iacc-team-up-to-fight-cybercrime/>> accessed 24 October 2018.

⁹⁷ Paul Vixie, ‘Notice, Takedown, Borders, and Scale’ (CircleID, 1 March 2017) <www.circleid.com/posts/20170301_notice_takedown_borders_and_scale/> accessed 24 October 2018.

⁹⁸ Annemarie Bridy, ‘A response to Paul Vixie’s “Notice, takedown, borders, and Scale”’ (Stanford University, The Center for Internet and Society, 3 March 2017) <<http://cyberlaw.stanford.edu/blog/2017/03/response-paul-vixie-s-notice-takedown-borders-and-scale>> accessed 24 October 2018.

⁹⁹ See Karen Kopel, ‘Operation Seizing Our Sites: How the Federal Government is Taking Domain Names Without Prior Notice’ (2013) 23 Berkeley Technology Law Journal 859.

¹⁰⁰ Internet & Jurisdiction, above n 65, 7.

March 2018, the European Commission has put forward the introduction of trusted notifier systems broadly for online platforms. Furthermore, this tendency towards privatized expedited enforcement mechanisms is not only apparent in relation to the application layer (i.e. online platforms) but also the infrastructure layer (i.e. domain name registries).

Trusted notifier systems, when drawbacks are mitigated, can contribute to more efficient enforcement. Theoretically, notifiers are much better equipped to identify unlawful content than intermediaries. Additionally, intermediaries can –at least to some degree– refrain from twiddling with content-related issues. But the situation also depends on the intermediary one is looking at.¹⁰¹ Online platforms, on the one hand, have notice-and-takedown procedures in place, often accompanied by content moderation practices that result in takedown of content without a third-party notice. Compared to current national regimes based on the E-Commerce Directive, trusted notifier regimes are just an adjustment of the notice-and-takedown system with an expedited process, by introducing a presumption for privileged third parties. Domain registries, on the other hand, as general-purpose technical infrastructure providers have so far engaged little in content-debates and -actions. Also in the context of responsibility, however, and especially compared to the direct enforcement by intermediaries themselves, these private mechanisms based on expert-governance models can come with efficiency gains.

This brings me to a more general point: I mentioned before the existing notice-and-takedown models based on the liability regime of the E-Commerce Directive as *one* starting point for trusted notifier-models. Here, as seen, trusted notifiers simply make existing notice-and-takedown models more efficient. Yet a *different* conceptual starting point is broader and unrelated to the legal framework of the E-Commerce Directive: the political or corporate desire to tackle unlawful content online related to ‘responsibility’ rather than ‘liability’. In this narrative, trusted notifier-systems can be seen as attempt to solve the role of intermediaries regarding unlawful content towards the background of the political expectation to take more responsibility.¹⁰² I am not arguing whether or to what extent trusted notifier-models are beneficial or desirable at the application or infrastructure layer. This is likely to depend on the political landscape rather than the contribution by legal scholars, economists or related disciplines. The political climate towards enhanced responsibility (vis-à-vis liability) and the European lawmaker’s encouragement to rely on non-judicial takedown mechanisms of online platforms, might have spill-over effects on the debate of the role of other intermediaries, including domain registries, though. Industry and interest organizations are likely to have an interest in domain registries’ potential role to block access to content and some registries are less reluctant to introduce trusted notifier regimes, too.¹⁰³ Some agreements between domain registries and registrants appear to contain a basis to enforce content- or use-related infringements.

Liability and responsibility are two separate aspects though¹⁰⁴ and in the absence of secondary legislation, actors need to carefully balance the benefits and drawbacks of privatized

¹⁰¹ See also Husovec, above n 10, 24, differentiating between proximate and remote services.

¹⁰² In this context, see also interrelation to corporate social responsibility e.g. in Emily B. Laidlaw, *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility* (Cambridge University Press, 2015), 58–115.

¹⁰³ See above. Some ccTLDs have a clear stance that only judicial orders would result in the take down of domain names. See e.g. Torrentfreak, ‘Pirate Bay Finds Safe Haven in Iceland, Switches to .IS Domain’ (25 April 2013) <<https://torrentfreak.com/pirate-bay-finds-safe-haven-in-iceland-switches-to-is-domain-130425/>> accessed 24 October 2018.

¹⁰⁴ See also Frosio, above n 8, 4.

enforcement mechanisms before sacrificing at the altar of efficiency. When looking at intermediaries and their role regarding unlawful content, one must have a differentiated view that acknowledges the different and technical functional roles of intermediaries. One size does not always fit all.

Many concerns raised in relation to notice-and-takedown regimes are also present with regard to trusted notifier regimes. The emergence of and reliance on these mechanisms, where norm-setting, monitoring and enforcement are centralized in a few private Internet actors, presents multiple challenges: these forms of quasi-content-regulation and private norm-setting raise legitimacy concerns given their detachment from traditional democratically legitimized regulation.

As shown above, there exists a discrepancy as to the applicable framework for these models, depending on the degree of ‘privatization’ of the arrangement. This is illustrated by Figure 1 below. Many of the functions performed by these private flagging and takedown mechanisms have traditionally been located in the sphere of public actors. Whereas public actors are subject to a variety of legal safeguards, private actors are less controlled and enjoy contractual freedom. In instances where trusted notifiers are public entities, such as certain authorities or governmental bodies, public law rules and principles infuse trusted notifier-models. This effect is less clear, when models are formed by private actors: both in instances where models are encouraged, endorsed or facilitated by state actors and even more so in instances where private intermediaries establish trusted notifier-models with private entities without state involvement.

Figure 1: Trusted notifier models and their regulatory framework

	Trusted notifier-model	Legal framework
Privatization ↑	Private-Private	Contractual freedom?
	Private-Private (Public endorsement)	Hybrid?
	Private-Public	Fundamental rights, constitutional safeguards, administrative law?

Today, these arrangements are largely unaddressed by secondary law. The current EU framework simply consists of a non-binding Recommendation which is blurry in its scope and content, in addition to leaving many aspects unaddressed.¹⁰⁵ Whereas well-intended, Recommendation (EU) 2018/334 is unlikely sufficient to create a level field or standard for these trusted notifier-models. The main issues addressed in this paper relate to the selection of trusted notifiers, the mitigation of the risk of over-removal and the principal-agent problem but there exist likely other aspects that need to be addressed. In the case of domain names, my concerns towards trusted notifier models are aggravated: the administration of domain names constitutes a monopoly and whereas this is often performed by private companies, they constitute a critical resource for society and certain public policy objectives are deeply enshrined, especially in the ccTLD sphere. More fundamentally and related to the technical layer, a domain name merely renders the access to content easier but is unrelated to the hosting of the content.¹⁰⁶ The proportionality aspects of domain name-related measure are especially important, as only the access to the domain name as a whole can be disabled, making it a potentially far-reaching measure. It is clear that the current regulatory initiatives do not have

¹⁰⁵ The E-Commerce Directive too, left it to the Member States to specify the notice-and-takedown regime.

¹⁰⁶ See Schwemer, above n 12, 6–7.

the domain name system in their crosshair. It is uncertain, however, to what extent domain registries could fall under the Recommendation's scope, leaving these intermediaries with even less guidance on trusted notifier-models.¹⁰⁷

Towards this background, the lack of a clear framework is problematic: when industry or legislator encourage such models, a clear framework is desirable both to ensure coherence and to transfer some of the traditionally 'public' safeguards to the private sector –especially as the emergence of these models is state-induced or state-supported in many instances. Given the anecdotal evidence from trusted notifier-models in this paper, the Council of Europe's works and the Commission's Recommendation seem to have had little influence on models so far. When private parties enforce content, many of the traditional administrative law principles do not apply. Husovec argues in the case of rights holders that statutory schemes are quickly outdated and slow to deploy, which is why both intermediaries and rights holders should prefer voluntary schemes.¹⁰⁸ Yet, given the identified issues, intervention that addresses certain minimum requirements but also remains flexible to allow for future developments appears necessary. The question is then, under what kind of framework these notifier regimes should operate. To simply broadly apply administrative law to these models seems unrealistic in most instances. At the same time, it is in the public interest that at least some safeguards are applied in order to mitigate the principal-agent trade-off: these relate to certain procedural minimum requirements and thresholds as well as levels of care on the notifier side. At the minimum level stands transparency. It is impossible to assess the functioning and desirability of trusted notifier-models without insight in their workings. This relates to a central issue that became visible during the research for this paper: the general unavailability of information on these privatized takedown models.¹⁰⁹ This could partly be mitigated by legislative intervention. Not only for academic research, but also for the legitimacy of these models as well as the development of best practices it is paramount that enforcement collaborations and their mechanisms are available to public scrutiny. To mitigate undesirable effects of trusted notifier-models, it is also in place to establish an obligation to provide an explanation for the takedown and to put in place redress mechanisms. Finally, given the principal-agent relation, there need to be taken precautions that prevent the abuse of notices and safeguard that notifiers provide high quality notices, e.g. by restricting automated notices to instances where false positives compared to human-generated notices are extremely rare.

Many of these aspects are, in fact, touched upon by the Commission's Recommendation. And whereas a soft law tool can be favorable because of its inherent flexibility, at least the most essential criteria should be binding for private intermediaries. This could, for example, be addressed when the E-Commerce Directive is undergoing a fitness check. Until then, intermediaries –both at the content and at the infrastructure layer– are well-advised to implement the Recommendation's principles.

¹⁰⁷ Although not explicitly mentioned in the CoE Recommendation of 2018, it is evident from the CoE's general information website on Internet intermediaries that also "entities that distribute domain names" are considered intermediaries. See <www.coe.int/en/web/freedom-expression/internet-intermediaries> accessed 23 October 2018.

¹⁰⁸ Husovec, above n 10, 33.

¹⁰⁹ A notable exception in the sphere of copyright takedown notices is the Lumen project, whose data has been analysed e.g. in Daniel Kiat Noon Seng, 'The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices' (2014) *Virginia Journal of Law and Technology*, 369, <<https://ssrn.com/abstract=2411915>>

Acknowledgements

This article is part of an industrial PostDoc research project that has been funded by the Danish Innovation Fund and the Danish Internet Forum (DIFO). This research represents solely the view of the author, who enjoyed full academic freedom but acknowledges that results may be in the interest of the co-funding organization. I thank the anonymous reviewer, the members of the Copenhagen Internet & Society Research group and the Norwegian Research Center for Computers at Law for their valuable comments.