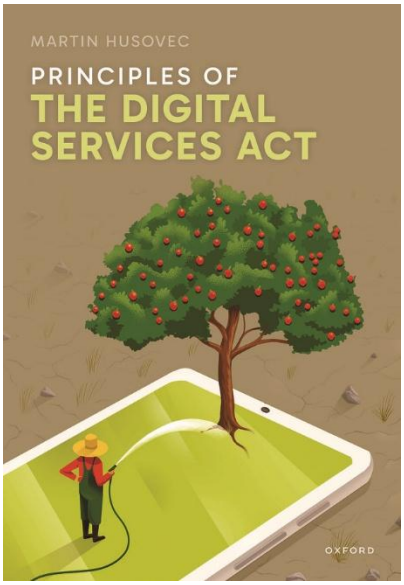
Martin Husovec, *Principles of the Digital Services Act* (OUP, 2024)



11
Fair Moderation Process

Contents

11.1 DSA's approach to content moderation

The DSA foresees numerous procedural requirements that are meant to put pressure on the quality of content moderation decisions and to allow due redress to *all* parties involved – not just users acting as content creators. The DSA prepares elaborate incentives for notifiers, providers, and affected third parties. The DSA also introduces a series of incentives based on trust that can be earned and lost: for instance, notifiers who have a track record of accurate notifications shall enjoy priority in the assessment; providers who fail to make the right decisions can face external complaints that they must reimburse. In my view, this incentive approach is well-suited to what has become an industry of its own, with large external suppliers of know-how and human expertise.[1]

As a result, the DSA can be said to fuse some principles from constitutional and administrative law,[2] but its approach to content moderation is still a form of market regulation. The DSA is probably better described as due process regulation for companies that includes a set of macro incentives for the ecosystem of content moderation. Such process fairness rules are complemented by transparency rules and more limited substantive fairness rules that I discuss in the next chapter (Chapter 12). Only providers with a systemic role in the EU's digital ecosystem face more extensive fair design obligations which ask them to test, re-design, and improve their content moderation systems along with other parts of the design of their services.

11.1.1 Scope of the procedural rules

Not all providers of the digital ecosystem are regulated by the DSA equally. Due diligence obligations under the DSA follow an asymmetric logic, that intends to tackle the more pertinent problems of content moderation where those arise, without unnecessarily imposing red tape on services which are more removed from the relevant content, are smaller, or have less of an impact on society. The graphic below depicts the basic layering logic for content moderation rules (see Figure 11.1).

[1] Adam Satariano and Mike Isaac, 'The Silent Partner Cleaning Up Facebook for $500 Million a Year Accenture' *The New York Times* (New York, 31 August 2021) <https://www.nytimes.com/2021/08/31/technology/facebook-accenture-content-moderation.html> accessed 29 August 2023.

[2] Nicolas P Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (CUP 2019) 8.

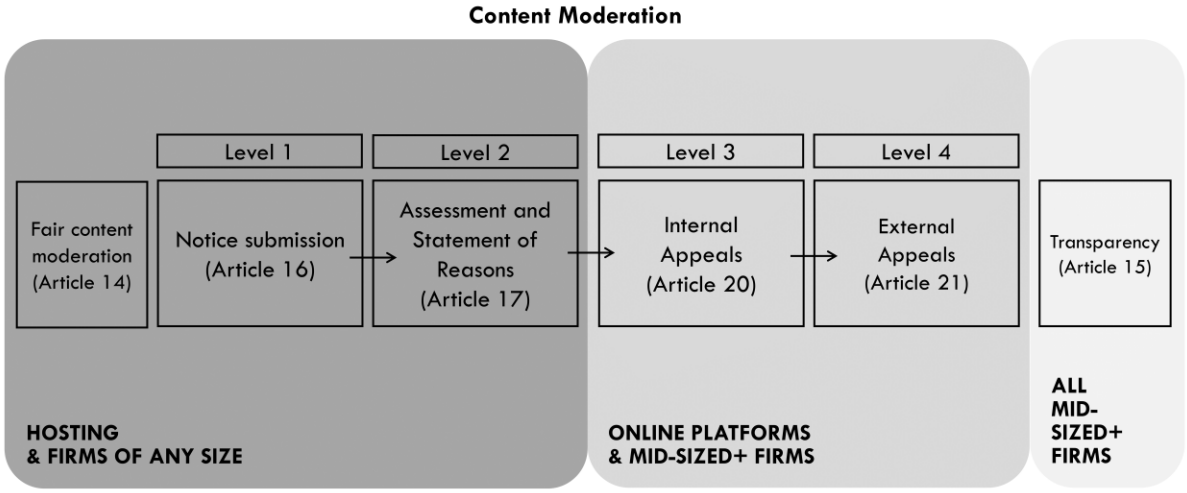


Figure 11.1 Content moderation rules (Husovec)

Mere conduit- and *caching-types* of services, such as internet access providers, some messaging apps,[3] or content delivery networks, are *not* obliged to moderate content. That being said, the DSA recognises that while many of these services traditionally did not moderate content, the trend is changing.[4] If caching services do decide to engage in content moderation, their decisions must be transparent and conform to some elementary requirements of fairness. Thus, for instance, if a content delivery network uses a child-abuse scanning technology on its clients,[5] such use will fall under Article 14 as it constitutes a content moderation practice. However, such practice, even though it can result in restriction of accounts, will not be subject to explicit due process obligations, such as to justify decisions and review their appeals. The same applies to mere conduit services, but with the caveat that content moderation practices by some providers might be prohibited under net neutrality laws.[6] For instance, an internet access provider is not allowed to block infringing websites without proper legal basis.[7]

Hosting-type services are subject to much more elaborate rules. They are expected to moderate illegal content because they could otherwise lose their liability exemption. However, due diligence obligations are independent of the question of liability of providers of intermediary services: providers could comply with

[3] See Chapter 6 of this book.

[4] For example, see Cloudflare, a global cloud platform, blocking Kiwifarms: Matthew Prince, 'Blocking Kiwifarms' (*The Cloudflare Blog*, 3 September 2022) <http://blog.cloudflare.com/kiwifarms-blocked/> accessed 3 September 2023..

[5] See the CSAM Scanning Tool: Justin Paine and John Graham-Cumming, 'Announcing the CSAM Scanning Tool, Free for All Cloudflare Customers' (*The Cloudflare Blog*, 18 December 2019) <http://blog.cloudflare.com/the-csam-scanning-tool/> accessed 3 September 2023.

[6] See the Open Internet Regulation, although it does not apply to WiFi operators as mere conduits: European Parliament and Council Regulation (EU) 2015/2120 laying down measures concerning open internet access [2015] OJ L310/1 (Open Internet Regulation).

[7] ibid, art 3(3)(a).

all the DSA procedures and still expose themselves to liability for third party content.[8] Similarly, providers who already lost a liability exemption, are still required to comply with the DSA content moderation procedures. Thus, the availability of liability exemptions and compliance with DSA rules are entirely separate from each other (Chapter 9).

Among hosting-type providers, online platforms are subject to most of the obligations. This very much reflects the magnitude of the problems posed by online platforms today, where such platforms, in their position as 'public-facing' providers, constitute the main source of friction. Finally, very large online platforms (VLOPs) have special obligations in content moderation, given their importance and reach 'in facilitating public debate, economic transactions and the dissemination to the public of the information, opinions and ideas and in influencing how recipients obtain and communicate information online'.[9]

DSA defines content moderation as follows:

> the activities, whether automated or not, undertaken by providers of intermediary services aimed, in particular, at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions, provided by recipients of the service, including measures taken that affect the availability, visibility, and accessibility of that illegal content or that information, such as demotion, demonetisation, disabling of access to, or removal thereof, or the recipients' ability to provide that information, such as the termination or suspension of a recipient's account[10]

Thus, the DSA's definition of content moderation is very broad, covering any 'activity' by providers that 'addresses' the compliance of users' content with rules introduced by parliaments or platforms themselves.[11] The term 'content', whether illegal or against terms and conditions, is also only a shorthand for users' content or their behaviour.[12] This effectively means that any activity can be viewed as content moderation if it addresses user-generated content. However, this legal definition is *not* used by the DSA to trigger due process rights. It mostly serves to delimit transparency obligations. Only a subset of content moderation activities trigger further process fairness obligations. The procedural rights are only owed for decisions that are exhaustively listed in Article 17(1). Still, some could argue, and I would agree to a limited extent, that the DSA should graduate the strength of procedural rights based on the severity of the intervention's impact; for example, highly structured dispute resolution processes for account terminations or monetisation disputes, and lesser procedural protections for the mere labelling of sensitive content.[13]

---

[8] DSA, recital 41.

[9] DSA, recital 75.

[10] DSA, art 3(t).

[11] See DSA, art 3(t).

[12] See Article 3(h) DSA, where it defines 'illegal content' as '*information that, in itself or in relation to an activity*… is not in compliance with [law]' (emphasis mine). Compare this to Article 3(t), where in the context of describing content that violates a provider's terms and conditions and is thus subject to content moderation, the DSA refers to such content merely as '*information* incompatible with terms and conditions' (emphasis mine). While the extension of the definition to include even 'information in relation to an activity' is not expressly replicated in the Article 3(t)'s description of content, the same extended scope should be adopted as well, given that the intention was clearly to cover the enforcement of both content and behaviour restrictions on the relevant services.

[13] See Evelyn Douek, 'Content Moderation as Systems Thinking' (2022) 136(2) Harvard Law Review 526, 582.

11.1.2 Design of interfaces

DSA not only regulates *procedural steps* for notification of illegal content but also the *principles* for the design of communication interfaces given effect to those rights. Consider Article 16(1):

> Those mechanisms shall be easy to access and user-friendly, and shall allow for the submission of notices exclusively by electronic means.

The DSA thus places an explicit requirement on all hosting providers to ensure widespread accessibility of their notification interfaces. The providers cannot hide them or make them cumbersome to use. If they impose restrictions, they must be able to justify why they have created additional friction and imposed a burden on notifiers, and why other measures of achieving the same purpose cannot be implemented. Article 16(1) thus explicitly challenges designs of notice submission systems that hide behind security or privacy to avoid obtaining large number of notices about illegal content by notifiers. In the past, some providers would undertake various measures to avoid receiving notices (e.g. by putting caps on them from specific IP addresses or accounts), or make their submission costly in order to limit their content moderation workload. These practices are explicitly prohibited. Any security measures must be justified because the notification mechanisms must be as user-friendly and accessible as possible.[14]

While the binding principle of *user-friendly design* is not explicitly mentioned in the context of the statement of reasons (Article 17), in my view, it applies to *the entire process of notice and action*. The principle is reiterated for how the information about redress is communicated (Article 17(3)(f)), and even the statement of reasons must be 'be clear and easily comprehensible' (Article 17(4)). Finally, this also follows from Article 12(1), Article 14(1), and finally, Article 44(1)(b) which in the context of standardisation speaks of 'templates, design and process standards for communicating with the recipients of the service in a user-friendly manner on restrictions resulting from terms and conditions and changes thereto'. Thus, the principle of user-friendly design is a binding obligation underlying the entire notice and action system.[15]

The DSA explicitly provides that the means of notification must be exclusively 'electronic'.[16] Thus, a practice where providers ask for letters to be sent by post is clearly not allowed. The European Commission, after consulting the Board, is tasked to support and promote the development of voluntary standards for the electronic submission of notices for all notifiers (Article 44(1)(a)). Trusted flaggers, unlike other notifiers, can gain access to more advanced programming interfaces to submit their notices which can again be standardised (Article 44(1)(c)).

The standardisation coupled with ease of submission is hoped to create a system where notifications easily flow between various entities through interoperable systems (e.g. NGOs easily generating notifications for a range of services without high administrative costs). In combination with incentives for quality of notices (e.g. trusted flaggers) and content moderation redress options that DSA creates (e.g. out-of-court dispute settlement and complaint processes), the hope is that while providers cannot reduce their workload, they can insist on quality. If the quality of the input level improves, that should translate to quality of the output

---

[14] Obviously, this does not mean that targeted measures cannot be taken against abusers of the notice submission systems. On the contrary, as explained in Chapter 12 when discussing Article 23, the DSA sometimes even mandates actions against abusers.

[15] While Article 16 only applies to the processing of notifications regarding illegal content, Articles 12 (contact point) and 14 (fair design of content moderation) apply universally, regardless of the basis for the provider's actions. They, therefore, suggest that the principle is binding to both.

[16] DSA, art 16.

level, i.e. content moderation decisions. The design principles behind notice and action systems are thus important building blocks for setting up good incentives for all the players involved in content moderation.[17]

11.1.3 Trusted flaggers

While the DSA's focus is firmly on providers, their actions can sometimes be only as good as the environment to which they are responding to. Thus, the DSA also comes up with some rules for other actors who are relevant in content moderation, such as trusted flaggers and other notifiers.

The role of trusted private actors has been long foreseen. Already the Communication on illegal content (2017)[18] and the Recommendation on illegal content (2018),[19] made clear that independent 'third-party' experts could play a key role in addressing the challenge of tackling illegal online content. These documents encouraged online platforms to work closely together with 'trusted flaggers'.[20] However, they were also thin on the procedural safeguards and allocation of roles and responsibilities.

By introducing the notion of trusted flaggers into regulation, the DSA tries to remedy in a roundabout way one of the original sins of the pre-DSA legal framework — that is, its inability to distinguish between actors who notified with due care and those who were reckless while doing so. Prior to the DSA, everyone's notifications had to be treated equally, which resulted in an incentive for notifiers to increase volume rather than quality. The DSA aims to change this. Partly building upon industry convention,[21] but also on the literature,[22] this new regime shows how badly the original system structured its incentives.

To improve the situation, the DSA creates a new tier of notifiers — the so-called 'trusted flaggers'. The underlying logic is that repeated actors whose work is known for its quality can be trusted more and given privileges in content moderation procedures, such as priority of review. Such privileges then act as an incentive for other frequent actors to improve their quality of notification to attain such status. The more

---

[17] For more on this, Lenka Fiala and Martin Husovec, 'Using Experimental Evidence to Improve Delegated Enforcement' (2022) 71 International Review of Law and Economics 106079; also Alexandre de Streel and Martin Husovec, 'The E-Commerce Directive as the Cornerstone of the Internal Market – Assessment and Options for Reform' (Study Requested by IMCO committee, PE 648797, European Parliament 2020) 35 <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/648797/IPOL_STU(2020)648797_EN.pdf> accessed 1 August 2023.

[18] Commission, 'Tackling Illegal Content Online: Towards an Enhanced Responsibility of Online Platforms' (Communication) COM(2017) 555 final.

[19] Commission Recommendation (EU) 2018/334 on measures to effectively tackle illegal content online [2018] OJ L63/50.

[20] It is a common practice by the best-known online platforms to create a privileged reporting channel for those users or organisations which are particularly active in reporting violations with an extraordinarily high rate of accuracy. These programmes vary from commercial users (rights holders, usually both for copyright and trademarks) to public policy objectives, such as hotlines or market surveillance authorities. Encouragements of such practices were already reflected in several codes of conduct, European Commission, 'The EU Code of Conduct on Countering Illegal Hate Speech Online' <https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en> accessed 6 September 2023; European Commission, 'Memorandum of Understanding on the Sale of Counterfeit Goods on the Internet' <https://single-market-economy.ec.europa.eu/industry/strategy/intellectual-property/enforcement-intellectual-property-rights/memorandum-understanding-sale-counterfeit-goods-internet_en> accessed 6 September 2023.

[21] Naomi Appelman and Paddy Leerssen, 'On "Trusted" Flaggers' [2022] Yale-Wikimedia Initiative on Intermediaries & Information <https://law.yale.edu/sites/default/files/area/center/isp/documents/trustedflaggers_ispessayseries_2022.pdf> accessed 29 August 2023.

[22] Fiala and Husovec (n 17).

trusted flaggers the ecosystem has (provided that there is sufficient quality control in getting the status),[23] the fewer mistakes the providers themselves make, and the fewer complaints necessary to correct their mistakes. In the end, not only victims but also content creators and their audience benefit.

Member States should therefore encourage civil societies throughout the EU to create enough skilled and well-resourced trusted flaggers locally. Unlike with other parts of the DSA, here the task lies heavily with the local civil societies, as German organisations are unlikely to police hate speech in Slovak, or Danish. This societal infrastructure, which the DSA only incentivises and envisions, needs to be built in each Member State for each area of content moderation. The situation is thus similar to the case of out-of-court dispute settlement bodies, which often have to originate from the same Member State in which the issue they help to resolve arises.

Trusted flaggers are entities awarded special status by national Digital Service Coordinators (DSCs). Their notifications must be 'given priority and are processed and decided upon without undue delay' (Article 22(1)). However, Recital 61 emphasizes that this only applies when they are 'acting within their designated area of expertise'.

It is expected that Article 22's reference to 'necessary technical and organisational measures' will lead to standardisation in the area of interfaces through which trusted flaggers can submit their notification in bulk and automatically from their internal systems.[24] While Article 22 only applies to illegal content, nothing stops providers from relying on the existing certifications in the context of terms of service violations or awarding other flaggers the same status as other entities that do not meet the criteria under Article 22.

The status is awarded only to entities who have expertise and competence in a specific area of content moderation, are independent of providers, and have a track record of carrying out their activities 'diligently, accurately and objectively' (Article 22(2)). Trusted flaggers must publish annual reports concerning notifications they have submitted with numbers according to the identity of providers, type of illegal content, and actions taken by the respective providers. The database of trusted flaggers is maintained by the European Commission (Article 22(5)).

While under the Commission proposal individual entities (such as rightsholders) could not be awarded this status, the final text allows any entity, even if it does not represent collective interests, to become a trusted flagger. Since trusted flaggers have to be entities which are independent from any platform, the reports must explain the procedures in place to ensure their independence. It is unclear why Article 22(3) limits independence to providers of platforms and does not include other stakeholders (Article 22(2)). The DSA's goal is generally to make sure that the trusted flagger represents the collective interests of some stakeholders, which is why Recital 61 encourages 'industry associations representing their members' interests' to apply for the status of trusted flaggers, 'without prejudice to the right of private entities or individuals to enter into bilateral agreements with the providers of online platforms'. The same recital also clarifies that even authorities, such as enforcement authorities, Europol, or semi-public bodies, can equally gain the status.

[23] However, Recital 61 DSA establishes that '[t]o avoid diminishing the added value of such mechanism, the overall number of trusted flaggers awarded in accordance with this Regulation should be limited'.

[24] Article 44 DSA also includes a call for developing voluntary standards for the 'electronic submission of notices by trusted flaggers under Article 22, including through application programming interfaces'.

Entities that were awarded the status of trusted flagger can be stripped of that status by the same DSCs. The responsible DSC can act upon complaints by the affected platforms, or on its own initiative. Following an investigation, during which the entity is confronted with the allegations, DSCs can decide to revoke the awarded status. If the procedure was initiated due to a complaint from a provider, the status can be also suspended during the revocation procedure (Article 22(6)). The typical reason for revocation is a violation of the elements of trust: objectivity, diligence, and accuracy of notifications or internal appeals. Trusted flaggers are in a special position here. Their trusted status can be suspended with much lower thresholds of mistakes, given that they should be known for diligent, objective, and accurate notifications or complaints. Finally, the European Commission is empowered to issue guidance further detailing the conditions for the award, revocation, and suspension of trusted flagger status, in case over time there is divergence in the interpretation by the different DSCs (Article 22(8)).

11.2 Step 1: Notice handling (Article 16)

Regulating the process of content moderation starts with anticipating how external notices are received by providers. The procedure dealing with notifications is now closely regulated under the DSA.

It was evident from its Impact Assessment that, for the European Commission, a basic obligation to put in place 'notice and action' mechanisms was essential to improve the status quo, for several reasons: to harmonise a very fragmented legal framework, to transform into the binding form the rules included in the 2018 Recommendation,[25] and to ensure that users could easily flag manifestly illegal content. The Commission was clearly conscious of the compliance costs that such an obligation involves. It expected that 'the introduction of the standard, minimum requirements for notices, procedures and conditions, as well as reporting templates, should further decrease the expected costs for small companies, supporting them in tackling illegal content and increasing, in turn, the legal certainty'.[26]

To be clear, the DSA does *not* regulate notice-receiving mechanisms of hosting service providers that seek to find violations of *terms and conditions*. Article 16 applies only to notification of illegal content. The reason for this is simple. The DSA makes a conscious choice not to legitimise and incentivise public enforcement of private rules, that is, the removal of content contractually restricted by the providers despite its lawful nature (e.g. nudity). This policy choice is aligned with the rest of the DSA, which intends to keep separate channels for public enforcement of 'public rules', such as orders by authorities, or notifications by others, and for private enforcement of 'private rules'. In the latter case, the DSA nevertheless aims at least to correct the negative impact that such private enforcement can have on users by subjecting them to the same process.

As explained above, *mere conduit* and *caching* services are not the traditional players in content moderation. Hence, Article 16 (and Article 17)[27] does not apply to them. This is also due to the fact that under the liability exemptions protecting such categories, there is no 'knowledge standard' following which providers could be held liable if they had actual knowledge of the illegality of the content they transmit or temporarily store. Article 16 is closely linked to Article 6 (the hosting liability exemption), as explained in

[25] Commission Recommendation (EU) 2018/334 on measures to effectively tackle illegal content online [2018] OJ L63/50.

[26] Commission, '"Impact Assessment" (Commission Staff Working Document) Accompanying the Document "Proposal For A Regulation Of The European Parliament And Of The Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC"' SWD(2020) 348 final, para 200.

[27] Discussed fully below in Chapter 11.3.

Chapter 7. Thus, internet access providers, open Wi-Fi providers, domain name authorities, messaging apps, videotelephony tools, web browsers, VPN services, and content delivery networks remain outside those rules. These services only have the universal obligation to conduct content moderation fairly under Article 14 *if* they decide to engage in content moderation. Article 15 prescribes some level of transparency, but again, only if they decide to moderate.[28] The fairness required by Article 14 for other providers cannot be equated with the elaborate set of procedural rules prescribed for hosting services by Article 16, although they undeniably give some indication, and can be followed by such providers voluntarily.

Hosting services are thus the core target of Article 16. The relevant services include social networks, various content-sharing services (pictures, text, or video), auction sites, marketplaces, user-edited encyclopaedias, and discussion forums. However, even among hosting services, there exist some less traditional, infrastructure-like services that only moderate content on occasion, such as webhosting,[29] cloud computing services, map services, or app stores.[30] Nonetheless, all these hosting services are subject to the basic due diligence obligations concerning notice submission and handling under Article 16 and 17. For hosting services that are *not* online platforms, such as webhosting and cloud services, or which do not qualify as medium-sized or bigger firms (Chapter 9), Articles 16 and 17 (and also 18) remain the only obligations they are subject to in the area of content moderation.

Despite these limitations, Article 16 sets out an important framework under which all providers of *hosting* services are subject to an elaborate notice and action choreography, which was subject to fierce discussions during negotiations.[31] As explained by Roche Laguna, 'while the Council's stance seemed quite similar to the Commission's proposal and they did not propose strong variations, the European Parliament represented both traditionally-confronted positions: on the one hand, a fierce defence of freedom of expression and information online, hence limiting notices' effect to the bare minimum; on the other hand, an impulse to expand the effect of such notices, including so-called "stay-down" effects, and facilitating the submission of notices, following demands of some industries'.[32] In the end, the result is very close to the Commission's original proposal.

11.2.1 The mechanism

Providers of hosting services are uniquely placed to provide the mechanism required for this notice and action choreography. As recital 50 explains, they play a particularly important role in tackling illegal content online as they store the information provided by and at the request of the recipients of the service and typically give other recipients access thereto, sometimes on a large scale.[33]

Under Article 16(1) DSA, they are obliged to 'put mechanisms in place to allow any individual or entity to notify them of the presence on their service of specific items of information' that they consider being illegal

[28] See the separate chapter on transparency, Chapter 16 of this book.

[29] For example, AWS suspending webhosting for Parler, John Paczkowski and Ryan Mac, 'Amazon Is Booting Parler Off Of Its Web Hosting Service' *BuzzFeed News* (10 January 2021) <https://www.buzzfeednews.com/article/johnpaczkowski/amazon-parler-aws> accessed 4 September 2023.

[30] Sarah Perez and Brian Heater, 'Apple Suspends Parler from App Store' *TechCrunch* (10 January 2021) <https://techcrunch.com/2021/01/09/apple-suspend-parler-from-app-store/> accessed 4 September 2023.

[31] *Interview with Irene Roche Laguna*, Deputy Head of Unit for Coordination and Regulatory Compliance, European Commission (2021 - 2023).

[32] ibid.

[33] DSA, recital 50.

content. The illegality depends on the national or EU law (Article 3(h)). Those mechanisms shall be 'easy to access, user-friendly, and allow for the submission of notices exclusively by electronic means' (Article 16(1)). Thus, as mentioned above, the DSA prescribes that providers must allow for notifications to be received by electronic means.[34] The DSA in other parts imposes an obligation on recipients of the service to send any communications electronically (Article 12(1)). As such, both sides of the notice and action mechanism are effectively required to utilise electronic communication. In practice, the hosting providers are likely to use a range of options, including regular email submission, electronic forms, or more sophisticated application interfaces for repeated notifiers. The DSA foresees a role for standardisation in this space (Article 44), also to ensure that this obligation is affordable for the smallest hosting service providers, or for those services that only offer hosting as an ancillary service.

The obligation contained in Article 16 refers primarily to the establishment of a submission 'mechanism' that shall 'facilitate the submission of sufficiently precise and adequately substantiated notices' (Art 16(2)). Compliance with this Article will be met when the mechanism is in place under the conditions established, in particular allowing the submission of sufficiently precise and adequately substantiated notices, a standard derived from CJEU case law.[35] Such mechanisms should also be 'clearly identifiable, located close to the information in question and at least as easy to find and use as notification mechanisms for content that violates the terms and conditions of the hosting service provider'.[36] If such a system is not in place, it would not allow notifiers to provide the necessary information (e.g. making the submission of notices difficult or hiding the mechanism).[37] Such providers could be liable for violations of the DSA's due diligence obligations, despite the fact that they will retain their liability exemptions for the hosted content.

The language thus builds upon the elements that are required to confer actual knowledge of third-party illegal content upon hosting providers. The concept's goal is two-fold: to allow notifiers to identify what is illegal, and to allow them to explain reasons, including providing additional evidence, if necessary. The 'mechanisms' that would preclude the provision of evidence or constrain the explanation to avoid receiving effective knowledge are incompatible with Article 16. The procedure looks as follows (see Figure 11.2).

---

[34] Saulius Kaleda, Paul-John Loewenthal and Folkert Wilman, *Digital Services Act: Commentary* (forthcoming, OUP 2024) § 16 (regarding the knowledge standard).

[35] *L'Oréal SA and Others v eBay International AG and Others* ECLI:EU:C:2011:474.

[36] DSA, recital 50.

[37] This was the main problem identified in the first round of enforcement of German NetzDG, or Germany's Network Enforcement Law. In July 2019, the authority imposed a fine on Facebook of €2 million based on its flawed transparency report for the first half of 2018 and on its reporting form being 'too hidden' for its users, see German Federal Office of Justice, 'Federal Office of Justice Issues Fine against Facebook' (3 July 2019) <https://web.archive.org/web/20221219105930/https://www.bundesjustizamt.de/DE/ServiceGSB/Presse/Pressemitteilungen/2019/20190702_1.html> accessed 19 December 2022; also Patrick Zurth, 'The German NetzDG as Role Model or Cautionary Tale? – Implications for the Debate on Social Media Liability' (2021) 31 Fordham Intell Prop Media & Ent LJ 1084.

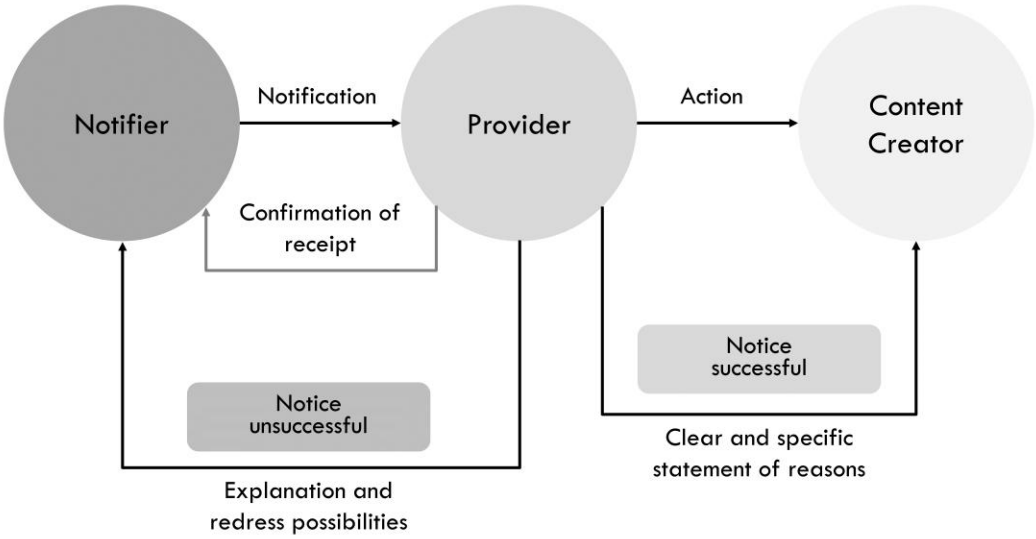


Figure 11.2 Notice and takedown choreography (Husovec)

11.2.2 The 'notice'

Article 16(2) DSA spells out that the notification mechanism shall be such 'to enable and to facilitate the submission of notices[38] containing all of' the following four basic elements: identification of the notified information, explanation of why it is illegal, notifier's name and contact details, and a good faith statement.

11.2.2.1 Identification of the illegal content

One of the most contentious parts of notification requirements concerns the identification of the information. The DSA requires that the notification includes 'a clear indication of the exact electronic location of that information, such as the exact URL or URLs, and, where necessary, additional information enabling the identification of the illegal content adapted to the type of content and to the specific type of hosting service' (Article 16(2)(b)). The provision thus identifies URLs as the primary method of reference in a notification, while not insisting on them as the only way. This is sensible given that some hosting services might not allow for identification by a URL, for instance when accessing content via an app.

Each notification should therefore include any means that identifies the illegal information in question. If the content is dynamic and without a URL, it could be a screenshot complemented by instructions (e.g. accounts or pages) on where to find it. The notification mechanism could also extract the location automatically from the content the user is reporting. Finally, if the notification concerns behaviour of an individual (e.g. repeated spamming) or groups (e.g. harassment or war propaganda) rather than a specific

[38] The obligation is not to have complete notices, but to establish a system that enables the gathering of all these elements.

piece of content,[39] the identification would primarily concern the actors, and then examples of their behaviour.

In any event, since any follow-up search efforts slow down the assessment by platforms and increase the cost of compliance, notifiers shall employ reasonable means to prepare the instructions concerning the location of information. The responsibility here also lies with the providers who should make it easier to pinpoint specific information for the notification. After all, Article 16(2) requires that platforms design their service to facilitate such notification. If they fail to do so, they might be expected to sort through notifications that are more difficult to locate. On the other hand, if providers offer a simple and accessible notification interface to locate the information, there is no reason why they should accept hard-to-process forms like screenshots with ambiguous written instructions. This is also emphasised by the reference to 'adaptation' to each service.

In any event, what is clear is that the notice mechanism is envisioned to identify specific pieces of illegal content, and not to receive generic instructions, such as those concerning all works of a given author or all posts by a relevant user. The provision does not allow for the so-called notice-and-staydown mechanisms, which are aimed at preventing the re-appearance of the same content. The location of a piece of content refers to the *existing* piece of content and not to future ones. It is possible, however, to notify multiple specific items of allegedly illegal content through a single notice.[40]

11.2.2.2 Explanation of why the content is considered illegal

Notices can only trigger actual knowledge of the illegality of the content 'where they allow a diligent provider of hosting services to identify the illegality of the relevant activity or information without a detailed legal examination' (Article 16(3)).

In order to do so, Article 16(2)(a) requires that the notifier provides 'a sufficiently substantiated explanation of the reasons why the individual or entity alleges the information in question to be illegal content'. The level of explanation will differ between various types of unlawful content. Content that is per se illegal will probably require little explanation. Content that is context-sensitive will require addressing the obvious reasons why the content is not justified. For instance, a notification claiming a violation of copyright must explain that the picture is protected, who owns its copyright, and why the use is not justified by exceptions or an existing license. Only with this information can the provider make a proper decision. Similarly, in the context of allegedly defamatory content of a factual nature, the notifier must explain why the facts are incorrect, and the content therefore illegal. If the evidence is necessary to substantiate these claims, the notifier should be allowed to upload it.

11.2.2.3 Notifiers' identification

While the DSA requires the self-identification of *notifiers* as a general rule (Article 16(2)(c)), this does not necessarily imply the self-identification of *victims*. In cases of certain sexual or child-related sexual offences,[41] the notifier does not have to identify themselves. Moreover, since the DSA allows victims to be

---

[39] See examples provided by Douek (n 13) 540; Camille François, 'Actors, Behaviors, Content: A Disinformation ABC' [2019] Working paper of Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression <https://www.ivir.nl/publicaties/download/ABC_Framework_2019_Sept_2019.pdf> accessed 29 August 2023.

[40] DSA, recital 50.

[41] Art 16(2) DSA refers to Articles 3-7 of Directive 2011/93/EU, which covers the following offences: sexual abuse, sexual exploitation, child pornography, solicitation of children for sexual purposes, incitement, aiding and abetting, and attempt, see

represented by various organisations, they do not have to file in their own name even in the context of other unlawful content (Article 86). In addition, many organisations can file notifications even without consulting or knowing the victim because the ability to file notifications is *not* preconditioned upon any right of action. Only the notification of illegal content whose assessment is context-dependent and usually requires additional evidence (e.g. the truth defence in the context of defamation) might be indirectly constrained since the notification cannot succeed without such evidence. In contrast, content that is per se illegal, such as child abuse material or terrorist content, can be notified by anyone. This means that the self-identification of the victim is not required in many other cases, not just those which are explicitly exempted.

11.2.2.4 Statement of good faith

Each notification shall be accompanied by 'a statement confirming the bona fide belief of the individual or entity submitting the notice that the information and allegations contained therein are accurate and complete' (Article 16(2)(d)). This requirement is presumably inspired by the US Digital Millennium Copyright Act (DMCA).[42] The providers themselves might use the statements when they decide to act against the notifiers, for instance, to collect damages for the cost of compensation paid to content creators for wrongful takedowns or restrictions. The statement might also be used by content creators, other users, or their organisations in situations where they seek redress for wrongful restrictions.

11.2.3 What is illegal content?

The definition of illegal content was debated in negotiations until the very end. 'Illegal content' is defined as 'any information that, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with Union law or the law of any Member State which is in compliance with Union law, irrespective of the precise subject matter or nature of that law'.[43] The notion of illegality thus covers the illegality of content or activity related to content. The former covers typical scenarios of terrorist content, hate speech, and copyright infringements, while the latter covers grooming, stalking or spam. Even situations where content creators sell goods or offer services without a license should fall under illegal content because the resulting information is not compliant with consumer law.

The DSA, through the definition of 'illegal content', de facto opens the doors to the applicability of the EU Charter to all digital content disputes. By extension, the CJEU is competent to decide whether a national law is in violation of the EU Charter (and very likely of the European Convention on Human Rights as well). As a result, the CJEU determines the outer limits of what is 'legitimately' illegal content on the national level. In this special area, it thus gains a role similar to the European Court of Human Rights, which has always been in charge of the human rights review of all national rules on content. The DSA places the CJEU in the same position through a piece of secondary EU law. Therefore, for instance, any national rule prohibiting the depiction of homosexual relationships, or prescribing disclaimers, can be reviewed for its compatibility with the EU Charter. If it is legitimate, it must be enforced through DSA's mechanisms. If it violates the EU Charter, it may not be. While in theory this is clear, in practice, it will be arduous to find a legal route through which the courts, and eventually the CJEU can review such rules (Chapter 3).

European Parliament and Council Directive 2011/93/EU on combating the sexual abuse and sexual exploitation of children and child pornography [2011] OJ L335/1, arts 3-7.

[42] DMCA (1998), 17 USC § 512(c)(3)(A)(v).

[43] Recital 12 DSA informs the broad interpretation that should be made on what can constitute 'illegal content'.

The DSA does *not* create any explicit mechanism through which such considerations can enter the application of the law by the DSCs, providers, users, or ODS bodies. Thus, for instance, providers assessing notices, or ODS bodies hearing appeals, do not have an explicit mandate to scrutinise the underlying law from the perspective of its validity. After all, it is not easy to create such a mechanism in the context of delegated enforcement by private parties. The main resolution would seem to be to presume the lawfulness of such national rules and challenge them ex post through national judicial systems.

In most cases the content will be 'legitimately illegal' in one Member State but not necessarily in another. Consider examples of prohibitions of the promotion of abortion services, religious blasphemy, different shades of laws about defamation, rules on disinformation, advertising of sugary drinks, or depiction of cruel behaviour on animals, etc. The DSA, as the E-Commerce Directive (ECD) before it, does not explicitly limit the geographical implementation of content moderation decisions.[44] However, the constitutional requirement of strict targeting of the measures, reiterated by CJEU case law[45] and the DSA itself (Recital 51), means that the over-implementation of local restrictions can be problematic. Over-implementation often diminishes the rights of other users 'who are using the provider's services in order to lawfully access information'.[46] The situation is, however, complicated by the fact that service providers have a right to conduct their business as they see fit, thereby they are in principle allowed to set their own house rules that can contractually broaden the scope of inacceptable content (e.g. to prohibit nudity).

The legal complexity that arises in a global environment is likely to incentivise the removal of content based on terms and conditions.[47] A similar situation applies if the basis for enforcement is local orders, for which the DSA also provides specific requirements on their geographic scope.[48]

11.2.4 The 'action'

Articles 16(1) and (2) do not *oblige* hosting service providers to act upon notices. The notice may or may not give rise to actual knowledge or awareness for the purposes of Article 6, depending on whether they allow a diligent provider of hosting services to identify the illegality of the relevant activity or information without a detailed legal examination. If the notice does give rise to actual knowledge, the hosting service provider will lose the protection conferred by the liability exemption. However, its liability (in case it failed to act) will have to be judged on the basis of the national or European law establishing positive liability, not on the basis of the DSA.

During negotiations, it was discussed whether creating an *obligation to remove* content was necessary or appropriate (as is the case under much more narrow German law),[49] but the question was deliberately

---

[44] Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* ECLI:EU:C:2019:821.

[45] Case C-401/19 *Republic of Poland v European Parliament and Council of the European Union* ECLI:EU:C:2022:297.

[46] ibid, para 56.

[47] Ben Wagner reports that the average level of content deleted by online platforms is 10.9 % and, within this, 4.33 % was deleted for reasons of legal compliance, while 95.67 % was deleted owing to terms of service violations: Ben Wagner, 'Digital Election Observation: Regulatory Challenges around Legal Online Content' (2020) 91(4) The Political Quarterly 739.

[48] Chapter 8.

[49] Germany's Network Enforcement Law, Netzwerkdurchsetzungsgesetz (NetzDG) v 01.09.2017 (BGBl. I S 3352), requires networks to remove or disable access to *manifestly illegal content* to be removed or blocked within 24 hours. Content which is *non-manifestly illegal* can be removed or blocked within 7 days. A similar law (Avia law) was passed in France, but it was declared unconstitutional by the French Constitutional Court, due to its interference on freedom of expression, Conseil Constitutionnel, 18 Juin 2020, Décision n° 2020-801 DC.

decided in the negative.[50] It was feared that doing so would incentivise the over-removal of lawful content, as providers would err on the side of caution and immediately remove the content in cases of doubt to avoid infringing such an obligation.[51] Moreover, the DSA is a horizontal measure, and thus creating such an obligation could have a draconic effect in some areas where the loss of the exemption should not immediately lead to liability. It would also change Article 16's ethos to more of a content-regulating rule.

After receiving the notice, the provider shall issue a *confirmation of receipt* to the notifier, if its contact details were provided with the notice (Article 16(4)). Once the provider decides on the action to be taken, the notifier shall be updated 'without undue delay' about the outcome, including the redress possibilities that the notifier has (Article 16(5)).

The DSA has two requirements for the provider's decisions:

- Each assessment of a notification on the basis of the fact that the content is *illegal*[52] shall be 'timely, diligent, non-arbitrary and objective' (Article 16(6)).
- Each decision leading to a relevant content restriction must be accompanied by 'clear and specific statement of reasons' for the decision (Article 17(1)); this requirement applies to decisions based on illegality and/or terms and conditions.

There is no time specification of what 'timely' means. Given the broad scope of the obligation (all hosting service providers), it was left open as to which service provider would be expected to act. All these questions are merely informed by Recital 51, which clarifies that, to be aligned with the fundamental rights guaranteed under the Charter of all parties concerned:

> …any action taken by a provider of hosting services pursuant to receiving a notice should be strictly targeted, in the sense that it should serve to remove or disable access to the specific items of information considered to constitute illegal content, without unduly affecting the freedom of expression and of information of recipients of the service. Notices should therefore, as a general rule, be directed to the providers of hosting services that can reasonably be expected to have the technical and operational ability to act against such specific items. The providers of hosting services who receive a notice for which they cannot, for technical or operational reasons, remove the specific item of information should inform the person or entity who submitted the notice.

Failing that, the provider's interference with the freedom of information of those users would be unjustified in the light of the objective pursued.

This recital covers two important and interrelated issues. Firstly, any action must be interpreted to be limited geographically to what is strictly necessary: if the content is illegal in one Member State, global removal is not strictly targeted. Secondly, if a hosting service provider cannot act with the necessary level of granularity against a particular item to achieve strict targeting because it can only act at a higher level up in the infrastructure (as it could be the case for instance for webhosting services or cloud infrastructure services), it is not expected to 'have the technical and operational ability to act against' a specific item. This results in an 'unserved notice', which, however, should not lead to a loss of the liability exemption. In that

---

[50] *Interview with Irene Roche Laguna* (n 31).

[51] Whether Member States are allowed, after the adoption of the DSA, to impose such obligation in national law, is discussed in Chapter 17 of this book.

[52] If the decision is taken on the basis of the provider's terms and conditions, Article 14(4) still requires providers to 'act in a diligent, objective and proportionate manner in applying and enforcing the restrictions', but it does not require the action to be 'timely'.

case, the provider must inform the notifier about the situation. This is what has been called the 'subsidiarity principle', whereby an action should be as close to the content as possible given the circumstances, and technical and operational ability to act.

Another special case concerns decentralised versions of content moderation on services which split their content moderation between centralized (operated by the provider's legal team) and decentralized forms (driven by the community or left to operators of add-on services). The DSA presupposes that moderation is mostly centralised in the hands of employees or contractors of the provider. However, some services, such as Reddit or Wikipedia, are known to split their content moderation.

The DSA's rules apply only to those channels of content moderation that are instructed by the provider. Thus, if a community moderates its content independently without significant involvement of the provider, it remains outside of the remit of DSA rules. For this reason, for instance, an independent community might not need to issue a statement of reasons. That being said, notifications or orders that are received based on the DSA's recognised channels (orders or notifications), must be resolved in compliance with DSA rules. This means that the DSA effectively pushes community content moderation to areas where the content is found *proactively*, that is, without an external notification or order, such as following the community's own monitoring efforts.

On the other hand, if the community is only an extension of the provider and lacks any real independence from it, the actions of volunteers can be attributable to the provider and thus remain governed by the DSA. The key test for this lies in the words 'undertaken by providers of intermediary services' (Article 3(t)).

11.2.5 Relationship with liability exemptions

While observance of Article 16 is clearly related to standards that govern liability exemptions, the relationship between compliance with conditions for liability exemptions on one hand, and due diligence obligations on the other, is more complex and fundamental.

For instance, the provider might comply with due diligence obligations and process a notification in a procedurally sound manner, but still make the wrong decision substantively. In such case, there is no violation of the DSA due diligence obligations, but there might be liability for underlying content. Alternatively, the provider acts upon a notification and avoids any liability for the users' content, it might do so in a way that is procedurally in violation of DSA due diligence obligations (e.g., by failing to provide a confirmation to the notifier or to give advance notice prior to suspension). Therefore, the due diligence obligations, even those in the context of Article 16, which are based on interrelated standards, are best seen as a separate layer of obligations, with their own goals and enforcement mechanisms (Chapter 9).

While Article 16 does not contradict the case law of the CJEU on the hosting liability exemption, neither does it modify such case law in any way. A failure of providers to meet obligations imposed by Article 16 does not trigger any automatic consequences in the context of the liability exemption envisaged by Article 6 DSA.

Functionally, the most important consequence of the notice is that, if it is *valid*, it might trigger the acquisition of the provider's actual knowledge or awareness about the illegality of content (Art 16(3)). This strips the provider of the liability exemption – but only for that notified piece of content: the knowledge is limited only to what is specifically notified and does not extend to other information. That being said, if the provider uncovers other illegal information while examining the notice, it shall act on those as well. In

practice, however, it might be difficult to obtain evidence on the extent of the provider's knowledge and subsequent omission in this case (Chapter 7.4).

The Commission's proposal for the DSA initially intended to introduce an implied presumption of knowledge.[53] It suggested that just by 'ticking all the boxes' on the prescribed content of the notice, such notice would automatically trigger actual knowledge. This was not in line with CJEU case law. In the final version, the text has been revised. A complete notice (containing all elements under Article 16(2)) will *not* automatically trigger actual knowledge. The inverse is also true. Even a notice missing the identity of the notice provider, or a good faith statement could still trigger actual knowledge.[54] However, the notifier omitting such information would fall short of DSA's requirements, which could be taken into account when considering what damages he or she is owed.[55]

Article 16(3) DSA stipulates what a notice requires to give rise to actual knowledge or awareness for the purposes of Article 6 in respect of a specific item of information: it must 'allow a diligent provider of hosting services to identify the illegality of the relevant activity or information without a detailed legal examination'. This provision codifies the standard of manifest illegality developed by the CJEU during the legislative process (Chapter 7.4). It was heavily debated during negotiations, raising the questions of: what precisely is 'manifestly illegal', and why should actual knowledge be limited to 'manifestly illegal content'?

As explained in Chapter 7.4, CJEU in *YouTube and Cyando* established that a notice 'must contain sufficient information to enable the operator of that platform to satisfy itself, without a detailed legal examination, that that communication is illegal and that removing that content is compatible with freedom of expression'.[56] This was subsequently confirmed in *Poland* judgment. In the Council, several Member States insisted[57] on explicitly adopting this high threshold for notices after the CJEU published the landmark *Poland* judgment[58] which reiterated *YouTube and Cyando*. As a result, Recital 53 now reads as follows:

---

[53] The proposed Article 14(3) stated that 'notices that include the elements referred to in paragraph 2 shall be considered to give rise to actual knowledge or awareness for the purposes of Article 5 in respect of the specific item of information concerned', Commission, 'Proposal For A Regulation Of The European Parliament And Of The Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC' COM(2020) 825 final, art 14(3).

[54] Kaleda, Loewenthal and Wilman (n 34) § 16.

[55] In liability systems around the world, it is an established principle that if victims contribute to their own damage by failing to exercise due care, the person who is otherwise liable will face less or no liability. This principle of comparative negligence was famously formulated by Lord Ellenboroug in 1809 who said that: 'One person being in fault will not dispense with another's using ordinary care for himself.' (*Butterfield v Forrester* (1809) 11 East 60 at 61, 103 ER 926 at 927). See more generally, Martin Turck, 'Contribution Between Tortfeasors in American and German Law - A Comparative Study' (1966) 41 Tul L Rev 1; Ernest Turk, 'Comparative Negligence on the March' (1950) 28 Chicago-Kent Law Review 189; Giuseppe Dari-Mattiacci and Eva S Hendriks, 'Relative Fault and Efficient Negligence: Comparative Negligence Explained' (2013) 9(1) Review of Law & Economics 1. Also Article 8.1.1 of the Principles of European Tort Law (PETL): 'Liability can be excluded or reduced to such extent as is considered just having regard to the victim's contributory fault and to any other matters which would be relevant to establish or reduce liability of the victim if he were the tortfeasor', taken from the European Group on Tort Law, 'Principles of European Tort Law (PETL)' <http://www.egtl.org/PETLEnglish.html> accessed 4 September 2023.

[56] Joined Cases C-682/18 and C-683/18 *Frank Peterson v Google LLC and Others and Elsevier Inc v Cyando AG* ECLI:EU:C:2021:503, para 116.

[57] *Interview with Irene Roche Laguna* (n 31).

[58] *Poland* (n 45), paras 53, 91.

> Where a notice contains sufficient information to enable a diligent provider of hosting services to identify, *without a detailed legal examination, that it is clear that the content is illegal*, the notice should be considered to give rise to actual knowledge or awareness of illegality.[59]

This means that, as a consequence, the provider is not expected to act expeditiously if the underlying notification relates to content whose illegal nature is not clear despite the notification. This is particularly relevant in cases of defamatory content whose illegality might be difficult to establish (e.g. it is not clear if a review of a restaurant is true or false based on what parties to the dispute claim).

A special situation concerns compliance with what I have termed 'illegitimate' national content rules; that is, when a provision of national law is in violation of the EU Charter, but it forms a basis for an order or notification. The DSA provides for a hosting liability exemption as long as providers act upon *manifestly* illegal content that they became *aware* of. This means that if the non-compliance with the illegitimate national rule is evident, providers are expected to act if they wish to preserve the liability exemption – although they only need to act locally. However, since they are only required to act under the illegitimate national rules and not under the DSA, they might also decide to take the legal risk and not comply with such requests.

In such cases, whenever the Member State, or an individual in that Member State, would try to enforce the rule against the provider in courts or through the authorities, the provider can invoke the defence of the rule's non-compliance with the EU Charter. If successfully established, the liability exemption would be preserved because the content was never 'illegal' according to the DSA's definition (Article 3(h)). The same would be true for potential violation of due diligence obligations that rest on 'illegality' of content.

While Article 6 DSA (similarly to Article 14 ECD) conditions the liability exemption to the hosting provider on acting 'expeditiously' upon obtaining actual knowledge or awareness, Article 16(6) requires hosting providers to take their decision in respect of a notice in a 'timely' manner. Despite some calls to define what expeditiously or timely means[60] and thereby establish specific timeframes, the DSA leaves the notion undefined on purpose.

Pre-DSA, some countries had experimented with explicit timeframes.[61] The two notable examples are Germany and France.[62] The French law, known as Avia law, was declared unconstitutional by the French *Conseil Constitutionnel*, which called for extreme caution against timeframes that were too tight.[63] Post-DSA, experimentation with explicit time-frames is no longer possible due to pre-emption (Chapter 17).

---

[59] DSA, recital 53 (emphasis mine).

[60] Several amendments in the European Parliament, that did not make it to the final report, called for a 24 hours deadline, or even '30 minutes where the illegal content pertains to the broadcast of a live sports or entertainment event', European Parliament, 'IMCO Report on the proposal for a regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC' Report A9-0356/2021, Opinion of the Committee on Legal Affairs, 419, amendment 111 (new Article 5(1)(a). In the EU Council, Germany attempted to reproduce its national law NetzDG into the DSA, thereby including a 24 hour deadline for removal, Presidency of the Council of the European Union, 'Digital Services Act: Consolidated Comments on Chapter 3 and Respective Recitals' (2021) Document WK 5155/2021 REV 2, 140.

[61] Chapter 17 on the DSA's relationship with national laws.

[62] Chapter 17 for detailed debate.

[63] The law imposed on certain operators of online platforms, under penalty of criminal sanction, an obligation to remove or make inaccessible within twenty-four hours illegal content because of its manifestly illegal nature, Conseil Constitutionnel, 18 Juin 2020, Décision n° 2020-801 DC, para 18.

As a result, whether the action was taken 'expeditiously' (condition of the liability exemption) or 'timely' (a requirement for compliance with Article 16 DSA) will need to be assessed on a case-by-case basis.[64] The Member States are *not* allowed to prescribe specific time limits in their national legislation due to the DSA's pre-emption that requires ad hoc balancing. Only the EU legislature can specify the period, as it did in the Terrorist Content Regulation (Chapter 18). In this context, it is remarkable that the creative sector, for instance, who would naturally be interested in expeditious action against copyright infringements, openly argued *against* the introduction of specific time limits for the take down of illegal content, as 'any indication of specific time limits would be short-sighted and make the DSA obsolete very quickly in light of the continuous and rapid technological developments'.[65]

11.3 Step 2: Statement of reasons (Article 17)

An important added-value of the DSA is its ability to extend and improve on existing rules that apply in sector-specific contexts (like the platform-to-business relationships)[66] or industry practices based on the US DMCA's 'counter-notices' for copyright notices.[67] Article 17(1) is an example of such added-value and has plays a crucial role in the DSA because it defines the due process obligations applicable to crux of the entire choreography: the content moderation decision itself. The scope is broad because it applies to restrictions based on *illegality* or is deemed *incompatible with terms and conditions.* This leaves out only actions by providers that do not qualify as individual restrictions, such as demotions as a result of the prominence given to paid content.

11.3.1 Scope of individual explanations

Article 17(1) requires all hosting service providers to provide a 'clear and specific' statement of reasons to any affected recipients of the service whose information has been made subject to any of the following restrictions:

> (a) any restrictions of the visibility of specific items of information provided by the recipient of the service, including removal of content, disabling access to content, or demoting content;
>
> (b) suspension, termination or other restriction of monetary payments;
>
> (c) suspension or termination of the provision of the service in whole or in part;
>
> (d) suspension or termination of the recipient's accounts.

The scope of Article 17 is narrower than the definition of content moderation in Article 3(t). To recall, the DSA's definition of content moderation covers any measure aiming, in particular, at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions. What Article

---

[64] Folkert Wilman, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US* (Edward Elgar Publishing 2020), para 2.74.

[65] International Federation of the Phonographic Industry (IFPI) and others, 'DSA: A Missed Opportunity and a Step Backwards' (2021) para 2 <https://www.ifpi.org/wp-content/uploads/2021/10/Creative-sector-in-Europe.-A-letter-on-the-DSA_October-2021.pdf> accessed 4 September 2023.

[66] Article 4 P2BR already establishes that 'where a provider of online intermediation services decides to restrict or suspend the provision of its online intermediation services to a given business user in relation to individual goods or services offered by that business user, it shall provide the business user concerned, prior to or at the time of the restriction or suspension taking effect, with a statement of reasons for that decision on a durable medium'.

[67] DMCA (1998) 17 USC § 512(g).

17(1) does is spell out what these measures can represent: those relating to a) visibility of content, b) its monetisation, and c) suspension or d) termination of service or account. The table below, inspired by Eric Goldman's taxonomy of remedies,[68] illustrates how various typical remedies in content moderation are likely to be classified from the perspective of Article 17 DSA (see Table 11.1).

Content visibility restrictions (Art 17(1)(a))	Monetisation restrictions (Art 17(1)(b))	Service restrictions (Art 17(1)(c))	Account restrictions (Art 17(1)(d))	Content moderation outside Article 17
Removing content	Forfeiting earnings	Blocking access to a service	Blocking accounts	Shaming
Suspending content	Suspending access to earnings	Suspending rights on a service	Suspending accounts	Community service
Relocating content	Disabling income on some content	Reducing speed on a service	Erasing accounts	Attach context or warning
Redacting content	Reducing earnings	Limits puts on notifications	Blacklisting registration	Apology
Shadow banning or Blacklisting	Fines	Limits on internal complaints		General changes to recommender systems

Table 11.1

The Commission's original proposal had been limited to decisions 'to remove or disable access to specific items of information',[69] but this was eventually revised to the current iteration with its arguably far wider scope. This broad scope is likely to cause difficulties in determining what precisely falls under each of the classifications, but the most problematic tensions will likely revolve around what constitutes visibility restrictions under Article 17(1)(a) as almost anything could be described as such.

Consider when a platform makes changes to its recommender algorithm to start downranking a particular category of content (e.g. nudity). Such restriction clearly affects visibility. However, the question is whether it must be individually explained to everyone affected by it. It definitely will need to be disclosed by online platforms if it constitutes one of the main parameters for such recommendations (Article 27). The DSA does not have an elegant answer, but the solution likely resides in the words 'any restrictions of the visibility of *specific* items of information'.[70] If a platform redesigns its recommender system that targets specific items of information, an individual explanation is in order, even if the item is only part of a class (e.g. importing a blacklist). However, if the changes are aimed at problems without having specific items of information in mind (e.g. downranking violent content), only general disclosure under transparency obligations are owed to the public (Articles 15(1)(e) and 27).

Similarly, labelling arguably is not a visibility restriction unless the original content is simultaneously restricted (e.g. labelling and subsequently blurring an image of violent content before showing it to users).

[68] Eric Goldman, 'Content Moderation Remedies' (2021) 28(1) Michigan Technology Law Review 1.

[69] Commission, 'COM(2020) 825 Final' (n 53), art 15(1). Both the Council and the European Parliament agreed that this was too restrictive and extended the obligation to other kinds of content moderation decisions.

[70] Emphasis mine.

The usage of the Commission's database of the statement of reasons suggests that at least some companies understand labelling as a visibility restriction, although the majority consider it outside of Article 17.[71]

In the coming years, Article 17(1)(a) is likely to cause major difficulties in interpretation, and regulators and judges will need to flesh out the distinction. Applying Article 17(1)(a) too broadly can seriously strain platform development and design, while applying it too restrictively can undermine the goal of the legislature – that is, to ensure the explainability of decisions. Without such judicial or legislative clarification, the costs of uncertainty in the interim can be fairly large since this obligation applies to all hosting services, including micro businesses. The Commission's efforts to implement a database for these decisions made by online platforms (Article 24(5)) might be able to give some limited guidance to companies.

Furthermore, Article 17(4) provides that the statements of reasons shall be 'clear and easily comprehensible and as precise and specific as reasonably possible under the given circumstances'. These statements must reasonably allow the recipient of the service concerned to effectively exercise the redress possibilities that the DSA foresees – internal or external appeals, or general judicial redress. Nothing in the provision prohibits the statement of reasons to be issued by an automated system. In fact, it is expected that the bigger the platform, the more likely it is that it will automate the process. Yet, even with this qualification, Article 17(1)(a) can still prove to be rather burdensome in practice, if it is not meaningfully curtailed by soft law.

11.3.2 Exceptions

The DSA introduces three *exceptions* to the obligation to issue a statement of reasons.

The first exception concerns situations 'where the information is deceptive high-volume commercial content' (Article 17(2)). The exception clearly targets spam bots or computational propaganda that relies on automation (Recital 55). However, it is rather narrow in its definition because if the content is deceptive and high-volume, but not commercial, such as war propaganda, the exception usually would not apply.

The second exception concerns decisions made based on orders from national authorities (Article 9).[72] This makes sense, as Article 9(5) already obliges providers of intermediary services to inform the recipient of the service, including a statement of reasons 'explaining why the information is illegal content, by reference to one or more specific [legal bases]' (Article 9(2)). Such orders can delay or limit these explanations. As explained in Chapter 8, the use of Article 9 is not mandatory for the national authorities of Member States. However, if national authorities do comply with Article 9 when issuing the orders to providers, Article 17 does not apply. Thus, whenever the providers comply with requests of authorities, they must issue a statement of reasons, either based on Article 9(2) or Article 17. The benefit of using Article 9(2) for national authorities, if they comply with its safeguards, is that the provision gives them more flexible mechanisms which can be adopted by national law compared to strong pre-emptive effect of Article 17.

The final third exception is a situation where the electronic contact information of the affected recipient of the service is not available (Article 17(2)). This is usually when the content creator who uploaded the content did not have to register, or the registration ceased to exist.

---

[71] See https://transparency.dsa.ec.europa.eu/statement-search (upon closer look, all such notifications only come from Facebook).

[72] In some areas, like the Terrorist Content Regulation, there are specific rules on this.

11.3.3 Who is owed what explanations?

The statement of reasons shall be given only to 'any affected recipients' (Article 17(1) and (2)). Due to the broad notion of a recipient (Article 3(b)), the obligation extends not only to affected content creators, but theoretically also those who are 'using' to the service.

Consider a situation of registered users of a video-sharing service, who can no longer view the video that was removed. The DSA certainly was not written with them in mind; however, they qualify as recipients of the service, and depending on the context, they can be 'affected' by the imposed visibility restriction. If their details are known to the provider, this would suggest that the provider must also provide another statement of reasons to them (e.g. by keeping the link to a video as a tombstone with an explanation of why the content was removed or, if applicable, interlinking with the EU Commission's centralised database according to Article 24(5)). This is a possible, albeit somewhat unintended, reading of Article 17. On the other hand, an *a contrario* reading of Article 32, whereby providers of online marketplaces need to inform buyers when a product of service happens to be illegal, could suggest that Article 17, despite its broad language, was intended to cover *only* content creators and notifiers who are affected.

The legal obligation under Article 17 also clears away any doubts about the legitimacy of the processing of personal data that the statement of reasons might entail. For instance, if the search engines, with respect to their links, were to be considered hosting services, they would have an obligation under Article 17(1) DSA to issue a statement of reasons to the affected websites (e.g. news publishers) whose content they are delisting under the GDPR's right to be forgotten.[73] A similar requirement already applies to search engines delisting or de-ranking 'corporate websites' under the Platform to Business Regulation (P2BR).[74] The DSA and P2BR respectively create legal bases for processing of personal data in such statements of reasons.

Finally, if the notice fails to lead to a relevant content restriction, the *notifier* should equally receive notice of such an important development by reason of Article 16(5). This information must allow the notifier to exercise redress against the decision. The DSA does not explicitly equate it with the framework of the statement of reasons and an unsuccessful notifier cannot be 'affected' by visibility or other restrictions. However, Articles 16(5), 20(5), and 21(1) respectively expect that the next steps of each party are informed by the possibilities for redress and the appeal rights in each case do not differ. Thus, given that appeal is only possible if the grounds are known to the affected party, arguably, Article 16(5) should be interpreted in line with Article 17.

With regards to *what* information should be included in the statement of reasons, article 17(3) specifies the minimum contents as follows (see Table 11.2)

When required	Minimum contents of the provider's statement of reasons
Always	Specification of the type of restriction: visibility-restriction, account-restriction, monetisation-restriction or service-restriction

[73] As a consequence, this would then mean that the data protection law has to recognise such notifications under the legal basis of 'legal requirement' under Art 17 GDPR.

[74] Art 5(4) P2BR: 'Where a provider of an online search engine has altered the ranking order in a specific case or delisted a particular website following a third party notification, the provider shall offer the possibility for the corporate website user to inspect the contents of the notification.' It must be noted, however, that the existence of this obligation in the data protection context seems to be disputed by some DPAs without much explanation. The Swedish Administrative Court of Appeal in Gothenburg in November 2021 ruled that Art 5(4) P2BR does not require a search engine to notify webmasters in the right to be forgotten context, see Swedish Administrative Court of Appeal, *Google LLC* [2021] Case No. 2232-21.

Always	'[T]he facts and circumstances relied on in taking the decision', such as whether it resulted from a notification or own-initiative investigations
Always	'[C]lear and user-friendly information on the redress possibilities' available to the recipient of the service (internal or external appeals mechanisms)
If necessary	The identity of the notifier
If relevant	The duration of the restriction
If relevant	The geographical scope of the restriction
If relevant	Information on the use made of automated means
Always	Either: a reference to the legal ground relied on and explanations as to why the information is considered to be illegal content on that ground
	Or: a reference to the contractual ground relied on and explanations as to why the information is considered to be incompatible with that ground

Table 11.2

11.3.4 Disclosure of the identity of the notifier

During negotiations, it was strongly debated whether the service provider should disclose the identity of the notice provider to the content provider.[75] On the one hand, such a measure could discourage rogue or abusive notices, for fear of legal consequences. On the other hand, such measures could precisely discourage the exercise of the right to report illegal content, for fear of retaliation (e.g. women reporting online harassment). Furthermore, given that Article 16 does not oblige notice providers to disclose their identity, the service provider may not even have such information. The resulting text of Article 17 is a compromise between these two concerns.

Article 17(3)(b) limits the obligation to disclose the identity of the notifier to the affected content provider only when it is 'strictly necessary'. This also follows from Recital 54, which states that 'where this information is necessary to identify the illegality of the content, such as in cases of infringements of intellectual property rights'. The rationale is thus to disclose the identity of the notifier (who does not have to be the same person as a victim) only if it is *functionally necessary* for the content provider to be able to defend himself or herself, for instance by challenging the illegal nature of the content. Article 17 also provides a legal basis for the necessary processing of personal data, as required under the GDPR.

For instance, to defend herself against a claim of trademark infringement, the content provider needs to know who claims to own what rights as some legal defences depend on it; to defend herself against a claim of defamation, the content provider will need to know whose reputation is claimed to be affected. On the other hand, to assess a notification about child sexual abuse images, it is often irrelevant who reported it. It is therefore the applicable law regulating illegality that determines whether the disclosure of identity is functionally required or not. If a defence to illegality can be made in the abstract without knowing the identity of the notifier (e.g. the picture is not sexually explicit), disclosure is not strictly necessary.

The lack of disclosure in the context of the decision, however, does not prejudice the potential recourse that the content provider might have against the service provider to disclose the identity of the notifier.[76]

[75] *Interview with Irene Roche Laguna* (n 31).

[76] Chapter 8.

Such disclosure is usually based on national law, although in some areas, such as intellectual property law, it may be harmonised in the EU.[77]

11.3.5 Notification of suspected crimes (Article 18)

Article 18 provides that all hosting providers who become aware of information 'giving rise to a suspicion' that some criminal offices took place, must notify the law enforcement or judicial authorities of the 'Member State(s) concerned'. If the concerned Member State is unclear, they must inform either the authorities of their Member State of establishment or the Member State of their legal representative, Europol, or both. If the providers inform only one of these entities, their obligation under Article 18 should be fulfilled too.

Hosting providers may learn about suspected relevant crimes when carrying out their risk mitigation strategies, or voluntary own investigations, but most typically such information arises in the context of their content moderation processes Consequently, this means that providers, in addition to its standard assessment, must now also include a flagging procedure within their content moderation process for notices that raise the suspicion of criminal offences. Where they suspect a relevant crime was or is likely to be committed, they must notify authorities. Moreover, they must assess the information, such as video or text, to determine which state is concerned. This is particularly important if the crime is only attempted, or ongoing. However, some crimes, such as distribution of terrorist content, can be equally time sensitive even after they are completed. DSA provides a number of factors that can be used by hosting providers to determine which authorities to notify: (a) the state where the crime was or is likely committed, (b) the place where the offender or (c) victim likely resides or is located.

Along with a notification, they must retain and transmit all relevant information at their disposal concerning the suspected relevant crime. The DSA does not specify what information could be relevant. Recital 56, however, mentions that this includes 'where relevant, the content in question and, if available, the time when the content was published, including the designated time zone, an explanation of its suspicion and the information necessary to locate and identify the relevant recipient of the service'. Providers are effectively under obligation to retain moderated content in many circumstances for the reinstatement purposes. Article 18 equally implicitly requires retention of the suspected moderated content.

The interesting question is whether Article 18 compels the providers to always notify authorities about the identity of the suspected criminal and victim. Recital 56 speaks of the identity of the recipient of the service, which a term that could cover both a victim and suspected perpetrator. However, in some countries, disclosure of identity can require some legal processes to be followed, such as issuance of orders by competent authorities. The proactive provision of information by Article 18 seems to circumvent these processes where they exist, as it gives the law enforcement information about identity without the authorities starting any process on their own. DSA says little about this concern. Recital 56 only explains the following:

> This Regulation does not provide the legal basis for profiling of recipients of the services with a view to the possible identification of criminal offences by providers of hosting services. Providers of hosting services

[77] Martin Husovec, *Injunctions against Intermediaries in the European Union: Accountable but Not Liable?* (Cambridge University Press 2017).

should also respect other applicable rules of Union or national law for the protection of the rights and freedoms of individuals when informing law enforcement authorities.

It is unclear how companies will use implement this provision, particularly when potential crimes are of different kinds, and the legislation in the Member States can substantially differ not only for what is labelled as crime, but also for what processes to follow procedurally.

The relevant crimes that trigger this obligation are criminal offences 'involving a threat to the life or safety' of persons, regardless whether imminent or already committed. Recital 56 details that this definitely includes offences specified in the Directives on combating human trafficking,[78] on combating the sexual abuse and sexual exploitation of children and child pornography,[79] or on combating terrorism.[80] These include:

- Directing or assisting a terrorist group
- Public provocation to commit a terrorist offence
- Recruitment and providing or receiving training for terrorism
- Organising or otherwise facilitating travelling for the purpose of terrorism
- Terrorist financing
- Offences concerning sexual abuse, such as engaging in sexual activities with a child
- Offences concerning sexual exploitation, such as recruiting or forcing a child to participate in prostitution or pornographic performances
- Offences concerning child pornography, such as acquisition, possession, dissemination, transmission or knowingly obtaining access to child pornography
- Solicitation of children for sexual purposes
- Offences concerning trafficking in human beings, such as recruitment, transportation, transfer, harbouring, exploitation or reception of persons

The national law can specify other relevant crimes if they meet the threshold of 'involving a threat to the life or safety' of persons. The Commission's proposal to limit the relevant crimes only to 'serious' criminal offences was opposed by the European Council.[81] The final version thus refrains from using the term.

The other relevant crimes that potentially come under Article 18 are incitement to violence,[82] stalking of protected persons, such as journalists, or crimes of kidnapping, hostage-taking, murder, or war crimes generally. The category is thus potentially very broad and almost unworkable. Given the ever-changing landscape of such crimes on the national level, and the number of Member States, it might be in the best interest of victims for law enforcement authorities, along with Europol, to try to standardise what qualifies as the relevant types of crimes in different countries and try to prioritise them. Otherwise, the

[78] European Parliament and Council Directive 2011/36/EU on preventing and combating trafficking in human beings and protecting its victims, and replacing Council Framework Decision 2002/629/JHA [2011] OJ L101/1

[79] European Parliament and Council Directive 2011/93/EU on combating the sexual abuse and sexual exploitation of children and child pornography [2011] OJ L335/1.

[80] European Parliament and Council Directive (EU) 2017/541 on combating terrorism [2017] OJ L88/6.

[81] General Secretariat of the Council, 'Note on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC - General approach' (2021) Document 13203/21, 41 (Amendment 42a)

[82] Council Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia by means of criminal law [2008] OJ L328/55

providers might simply over-burden the authorities by the sheer volume of notifications. The authorities often have limited resources to act. Practically, it makes the most sense to notify about upcoming threats where authorities can avert the harm to victims, or where speedy action allows authorities to preserve and act upon the evidence. Notification about past crimes is only as helpful as are the capabilities of local law enforcement authorities to process them and potentially prosecute the offenders.

11.4 Step 3: Internal appeals (Article 20)

One of the DSA's main objectives is to establish a safe, predictable, and trusted online environment in which fundamental rights of users are effectively protected. This entails empowering users through Article 20 (and Article 21)[83] to contest certain decisions of online platform providers concerning the illegality of content or its incompatibility with the terms and conditions that negatively affect them.[84]

In the case of users as notifiers, if the notice was not successful, that is, where the notifier is dissatisfied with the provider's decision, the notifier can avail itself of the internal and external appeals envisioned by the DSA (Articles 20 and 21). As explained above, notifiers will receive, under Article 16, all necessary information regarding the follow-up given to its notice by the provider, including possibilities for redress.

Similarly, in the case of users as content creators, where restrictions were imposed by the provider against the content creator's content perhaps because a notice was successful, they can similarly avail themselves of the same appeal procedures under Articles 20 and 21.

Such agency granted to users is much more needed in the case of medium and larger platforms which are also often the most used platforms. This might explain why the obligations under Articles 20 and 21 apply exclusively to them, and do not bind hosting services which are not online platforms, or small or micro firms. Once these DSA processes become part of general expectations, smaller companies might adapt these mechanisms voluntarily to keep up with the rest of the industry.[85]

11.4.1 Appeal scenarios

Given the myriad of paths to redress for both content creators and notifiers, it is useful to delineate five paths to redress in the content moderation process:

- The content creator's content is restricted on the provider's own initiative.
- The content creator's content is restricted due to an order.
- The notifier is fully successful with its notice.
- The notifier is fully unsuccessful with its notice.
- The notifier is partly unsuccessful with its notice.

Information about redress possibilities accompanies all these decisions (Article 16(5), Article 17(3)(f), Article 9(5)). However, the redress available is not always the same. Following the DSA, there are generally three possible types of redress, which are not mutually exclusive nor mutually dependent:

- Internal appeals, or so-called complaint handling system (Article 20);
- External appeals, or so-called Out-of-court dispute settlement mechanism (ODS) (Article 21); and

[83] Discussed fully below in Chapter 11.5.

[84] DSA, recital 58.

[85] But even where they decide to use them voluntarily, they are not subject to advanced due obligations about how to operate them and cannot be fined for breach with those obligations.

- Judicial redress, in accordance with the laws of the Member State concerned (Recital 59).

The DSA itself regulates the first two but does not establish the availability of judicial redress for individuals (Chapter 19). Notifiers and content creators might, however, still have a cause of action against providers, including those providers who do not have to offer redress mechanisms (see Table 11.3).

Outcome / Procedural step		Content creator's content restricted on provider's own initiative or orders	Notifier is fully **successful**	Notifier is fully **unsuccessful**	Notifier is **partly unsuccessful**
All hosting providers	Level 2: Statement of reasons	Yes, provided to the content creator (Art 17; 9(5))		Explanation for the decision in feedback received (Art 16(5))	Provided to content creator (Art 17(3)(f)) and notifier (Art 16(5)) respectively
Only mid-size and bigger online platforms	Level 3: Internal appeals	Provided to the content creator under Art 20 (except for orders under Art 9(2))		Provided to the notifier (Art 20)	Both the notifier and content creator can make use of internal appeals simultaneously
	Level 4: External appeals (ODS)	Obligatory to submit to ODS initiated by content creator under Art 21 (except for orders under Art 9(2))		Obligatory to submit to ODS initiated by notifier (Art 21)	Both the notifier and content creator can make use of ODS simultaneously
All providers	Level 5: Judicial redress	Always available under national law			

Table 11.3

The content creator's content is restricted on the provider's own initiative

In the case where the content creator's content is restricted following the provider's own investigation, they must receive a statement of reasons (Article 17), which includes information on the redress possibilities (Article 17(3)(f)). As there is no notifier within the meaning of Article 16, any internal and external appeals procedures filed by the affected content creator proceed without the presence of any notifier.

The content creator's content is restricted due to an order

If the provider adopts a relevant content restriction based on an *order* issued by authorities and providers issue a statement of reasons under Article 17, the usual appeals procedure under Article 20 is available. However, this will not be the case if the authority decided to rely on Article 9(2).[86] If Article 9 is used, an explanation is still given, but only on the basis of Article 9(5). On the basis of Article 20(2), the affected content creator cannot use internal appeals against decisions based on Article 9 orders and must instead challenge the order itself. While the DSA does not discuss this option, an argument can be made for the availability of internal appeals, as follows: Although Article 20(2) intentionally omits Article 9 orders, such

[86] Note, however, that only Member State authorities can rely on Article 9; authorities of non-Member States cannot.

decisions are not excluded from the scope of Article 20(1). Thus, the matter remains somewhat open. On the one hand, it is understandable that the legislature did not want to offer the ability to make appeals in response to well-drafted legal orders according to Article 9, as these are issued by authorities that are the right forum for any contestation. On the other hand, the providers could be over-implementing such orders on their own initiative, which is a decision that cannot be reversed by the authorities and thus deserving of an internal appeal avenue. Accordingly, the possible middle-ground is to allow such internal appeals only after the authorities at stake issue a clarification as to the proper scope of their orders or modify their orders to accommodate the complainant. After such clarifications by authorities, the decision of the provider is arguably no longer based on Article 9 orders.[87] This does not seem anyhow drastic given that other orders than those complying with Article 9 continue to be reviewable under Article 20. For them, Article 17 applies without any changes, because there is no carve-out similar to Article 17(5).

The situation is even more complex for external appeals. Although Article 21(1) references Article 20(1) for the reviewable subject matter, it does not even implicitly limit it by the types of explanations given along with the decision (as does Article 20(2)). The question thus is whether Article 9 orders could be reviewed before ODS bodies, when they cannot be reviewed in internal appeals. This, in my view, seems internally inconsistent. In any case, ODS bodies cannot review the order itself. While the same is true for orders that are not based on Article 9, there the authority is not bound to explain the redress possibilities against the decision itself. Thus, in my view, the argument seems stronger that Article 21 equally cannot be used to review decisions of platforms that implement Article 9 orders.

Notifier is successful

In the case where the notice is successful, the content creator's content is restricted. The notifier is informed about the decision (Article 16(5)), and the affected content creator receives the statement of reasons (Article 17), which includes information on the redress possibilities (Article 17(3)(f)). If the platform qualifies as a medium or larger online platform, the provided information must include information about internal and external appeals; otherwise, it need only mention judicial redress, but can mention any other voluntary systems, if available. In this scenario where the notifier is fully successful, it is the affected content creator who can avail herself of these additional options to contest decisions.

Notifier is unsuccessful

In the case where the notice is unsuccessful, the notifier too can avail itself of the internal and external appeals, as confirmed by Articles 20(1) and 21(1). While the unsuccessful notifier is not issued with the statement of reasons according to Article 17, the feedback it receives on the decision on the basis of Article 16 should include sufficient explanation which will facilitate any subsequent use of internal and external appeals. As explained above, Article 16(5) only speaks of 'providing information on the possibilities of redress' and does not mention any provision of reasons for the decision, but one can easily imagine that without a relevant explanation of reasons for refusal, appeals could hardly be properly used. For this reason, I have argued above that Article 16(5) should be given equivalent meaning as Article 17. Given the sanction mechanism implicit in Article 21, online platforms are likely to be incentivised financially to better explain their decisions to discourage the notifiers from pursuing external appeals.

Notifier is partly unsuccessful

[87] The question, though, remains on how to measure time in such cases.

In the case where the notice is only partly successful (e.g. only part of the content is taken down), both the notifier and the affected content creator can make use of the internal and external appeals simultaneously. The DSA does not foresee this situation explicitly, but it is likely that if two actions are pending in parallel, the providers could merge them.[88]

11.4.2 Scope and outcomes

Each relevant decision can be reviewed within six months following the day on which the recipient of the service, including the notifier, was informed about the relevant decision (Article 20(2)). The appeal must be handled electronically and proceed at no direct cost to the complainant (Article 20(1)). It can concern the relevant content restrictions based on illegality, or incompatibility with the terms and conditions of the provider (Article 20(1)). The relevant content restrictions, following Article 20(1), are the same as the ones indicated in Article 17 (visibility, account, monetisation, and service restrictions). As a result, the internal appeals do not for instance cover decisions about disclosure of information about users, or other typical remedies, such as labelling, attachment of warnings, or other contextual information.

Internal appeal systems must be 'easy to access, user-friendly, and enable and facilitate the submission of sufficiently precise and adequately substantiated complaints' (Article 20(3)). The internal reviews must be undertaken 'in a timely, non-discriminatory, diligent and non-arbitrary manner' (Article 20(4)). The DSA specifies that internal appeals should allow full reviewability, that is, reversal of any possible decision that was initially made, including initial inaction or application of a disproportionate remedy. If the provider finds that it made an error in its initial assessment, it shall 'reverse its decision ... without undue delay' (Article 20(4)). Such 'reversibility' implies that the providers may only disable disputed content at the start and can only remove it following the expiration of the period for internal appeals. During this period, they have a retention obligation to assure 'reversibility'.

The internal review by a provider shall result in a 'reasoned decision' (Article 20(5)) accompanied by information about further available remedies, including ODS appeal and judicial remedies. The DSA does not further specify the content for such a reasoned decision; however, the content required for the statement of reasons is likely applicable by analogy. Unlike the initial content restriction decisions and the related statements of reasons, the 'reasoned decisions' issued by providers following the internal appeal process must be overseen by humans (Article 20(6)). Thus, the expected minimum requirement is a proper response to complainants' raised objections. The DSA also requires that the human experts overseeing internal appeals are 'appropriately qualified' (Article 20(6)). It is possible that future codes of conduct will try to specify what types of qualifications and conditions for decision-making can be expected of persons handling such internal appeals (Chapter 20).

The DSA does address the question of how to resolve conflicts between multiple internal or external appeals, or internal and external appeals. Article 21(2) only provides that an online platform may refuse to engage 'when the dispute concerning the same information and the same grounds of alleged illegality or incompatibility of the content has already been resolved'. Therefore, in the case of multiple internal appeals, it is up to a provider to organise its own systems. The more serious question concerns the conflict between external appeals decisions, as these can be issued by different ODS bodies. However, since online platforms are not effectively bound by these decisions, and only face pressure to observe them in general,

[88] The problem, however, is that either side can file an internal complaint within 6 months, and an ADR review without time limitation. Thus, two review requests do not have to meet in time.

these exceptional circumstances should be easily resolvable by implementing the decision the provider finds more convincing. Internal appeals can be initiated even when a statement of reasons was not issued.[89]

11.5 Step 4: External appeals (Article 21)

The DSA considers private rulemaking that affects third-party content a potential source of significant private power over the society and market. Because providers of digital services come in different shapes and sizes, the DSA only constrains such power gradually. All providers must disclose their rules (Article 14(1)), and those who store content (hosting providers) must explain their individual content moderation decisions (Article 17). Mid-sized companies that distribute the content publicly as their main functionality (online platforms) must also allow internal appeals and engage in external dispute resolution (Articles 20 and 21). These tiered procedural rules constrain the private power of companies when exercising their discretion and commercial freedom.

Any firm with a superior bargaining position that unilaterally sets rules for its business partners and consumers, cannot remain impartial when interpreting such rules, especially if it remains affected by the outcomes of such disputes. The temptation to bend the rules to the desired outcomes is too high. Any hidden changes affect the legal certainty. And without legal certainty, business partners and consumers build their livelihoods and lives on top of unstable fundaments. The carpet can be pulled from under them at any point without advance notice. The DSA does not cement the rules — it would be foolish to do so — but it tries to bring some *predictability* to the inevitable changes.

Since internal appeals of content disputes are handled by the companies behind the platforms and are thus not entirely independent, the DSA supplements the system by external review of decisions through means of ODS (Article 21). ODS systems were known in the digital ecosystem for years. Domain name disputes, for instance, are decided by swift and efficient alternative dispute resolution (ADR) systems since 1999.[90] They help domain name authorities to tackle the problem of abusive registration on a global scale. Similarly, there are ADR systems in consumer law, such as in the aviation industry in which airlines submit their disputes to impartial decision-making concerning the compensation for delayed or cancelled flights.[91]

ODS bodies are not courts, not even 'de facto' courts. The DSA's system is probably better described as a system of second opinions that providers cannot easily ignore and must pay for if they lose.

Article 21 DSA has a three-fold goal. First, it provides a *credible remedy* to the affected individuals who can complain to an independent party and have their cases quickly reviewed by an expert who issues a second opinion. Second, because successful applicants are reimbursed by providers, the existence of ODS creates

---

[89] This is also confirmed by Article 20(2) DSA.

[90] All ICANN-accredited domain name registrars must follow the Uniform Domain-Name Dispute-Resolution Policy (UDRP) which was adopted by Internet Corporation for Assigned Names and Numbers (ICANN) in 1999. See ICANN, 'Uniform Domain-Name Dispute-Resolution Policy' (*ICANN*, 24 October 1999) <https://www.icann.org/resources/pages/help/dndr/udrp-en> accessed 4 September 2023.

[91] See e.g. Centre for Effective Dispute Resolution (CEDR), 'CEDR Aviation Adjudication Scheme Rules' <https://www.cedr.com/wp-content/uploads/2021/10/Aviation-Adjudication-Rules-Nov-2020-v2.pdf> accessed 29 August 2023. For further background in EU law in general, Pietro Ortolani, 'The Digital Services Act, Content Moderation and Dispute Resolution' <https://papers.ssrn.com/abstract=4356598> accessed 4 September 2023. Some examples of similar ADR bodies are France's FEVAD, Signal Conso and ReclameICI, Spain's OCU, Portugal's electronic consumer complaints book, Luxembourg's Médiateur de la consommation, Belgium's French-speaking Test Achat and Dutch-speaking Test Aankoop, Germany's Reklamation24, Italy's Altroconsumo, UK's PissedConsumer.com, Austria's Ombudsstelle, the EU ODR platform, and the European Consumer Centres etc.

a *financial incentive for providers to make fewer mistakes* in their own internal review. Third, because ODS bodies are independent of providers, they provide for a credible separation of decision-making powers from the profit-orientated leadership of companies.[92]

The first two goals were proposed and empirically tested in the academic literature by Fiala and I before it became part of Article 21 DSA.[93] The papers' idea was that the rational bias toward over-blocking documented during the past two decades can be mitigated by giving the affected parties the ability to correct the mistakes by an outsider, and thereby inflict small costs onto providers for their initial mistakes. If many individuals use their newly gained rights and try to improve their situation, the cost for the provider becomes significant. Thus, in turn, they will work to change the flawed internal processes that give rise to ODS disputes in the first place. The ultimate goal of the ODS is thus better internal content moderation with fewer mistakes.

11.5.1 Who can file an external appeals and when?

Every 'recipient' of the platform service addressed by the relevant content restrictions (visibility, account, monetisation, and service) can file an external appeal. This includes content creators, notifiers of the content, and potentially affected readers. They are entitled to file an external appeal even if they failed to initiate an internal appeal or did initiate it, but it is still pending (Article 21(1)). Naturally, they can equally file an external appeal if the internal appeal process has been already finalised. The DSA does not place any time limit to initiate such an ODS procedure.

Unless states decide to subsidise some ODS bodies, the majority will be run as fee-charging projects. Thus, the existence of fees will impose an informal structure: free internal appeals are relied upon first, and fee-based external appeals are relied upon next.

Platforms have an obligation to inform their users about ODS in their statement of reasons and internal appeal decisions. According to Article 21(1), they should even facilitate access to such bodies on their own interface. Ideally, the providers would provide a simple link through which affected parties can refer their case to one of the ODS bodies whose systems are interconnected with those of providers. This is particularly important given that the DSA also requires providers to be forthcoming with various kinds of specialist independent non-profit organisations of users (Article 86(1)).

The choice of an ODS body is with the complainant. However, ODS bodies must be experts in a particular area of dispute resolution for which they have received certification. Thus, while the choice is with the complainant, it is constrained by the availability of expert bodies in the area of the dispute. The certification in one EU state is valid across the EU.

In terms of geography, the DSA does not limit the relief to European citizens or inhabitants (Article 3(b), Article 21(1)). Thus, a successful Nigerian Youtuber who has a following in the EU can use the ODS system vis-à-vis decisions of online platforms that are regulated in the EU (Article 3(e)). The DSA does not provide a hard limit on which EU non-residents can use the ODS system, and benefit from its compensation scheme. The CJEU will need to clarify to what extent 'an EU connection' is even a relevant criterion for individuals invoking the DSA rights in some cases. It is likely that the providers might protest the unrestrained use of

[92] Douek argued for separation within companies, Douek (n 13) 587 ('platforms should be required to put a wall between those concerned with the enforcement of content moderation rules … and those whose jobs is measured against other metrics, such as product growth and political lobbying').

[93] Fiala and Husovec (n 17).

ODS bodies for non-EU individuals because it increases their costs. The DSCs and ODS bodies should therefore think about how they inform the non-EU complainants about the compensations of fees and costs.

11.5.2 What decisions are subject to ODS appeals?

ODS shall be available for all *relevant* decisions by *online platforms*.

All online platforms are obliged to submit to ODS (Article 21 DSA). Online platforms are services that store and distribute content to the public and are operated by at least mid-sized companies (that is, 50 employees, or €10 million turnover).[94]

In practice, most disputes will concern very large online platforms (see Table 11.4).

	Company	Digital Service
Social media	Alphabet	YouTube
	Meta	Facebook
	Meta	Instagram
	Bytedance	TikTok
	Microsoft	LinkedIn
	Snap	Snapchat
	Pinterest	Pinterest
	Twitter	Twitter
App stores	Alphabet	Google App Store
	Apple	Apple App Store
Wiki	Wikimedia	Wikipedia
Marketplaces	Amazon	Amazon Marketplace
	Alphabet	Google Shopping
	Alibaba	AliExpress
	Booking.com	Booking.com
	Zalando	Zalando
Maps	Alphabet	Google Maps

Table 11.4

Other non-VLOP platforms that are subject to ODS if they meet the mid-size company criterion are, for instance: BeReal, Reddit, Telegram (groups), Viber (groups), Airbnb, Apple Books, Vinted, Allegro, Cdiscount, Leboncoin, Roblox, eBay, Tripadvisor, Trustpilot, Gutefrage, Heureka, Skyscanner, PornHub, OnlyFans, Spotify Podcasts, DailyMotion, Github, Discord and Tumblr.

Very large search engines (VLOSEs) are not explicitly subject to ODS; however, their situation can be complicated by the fact that various features of search engines can constitute platforms. Thus, disputes revolving around the right to be forgotten or copyright infringements could be in theory subject to Article 21 DSA, but the interpretation is far from being settled (Chapter 7). The Commission's database suggest that search companies so far do not issue statements of reasons in delisting or other cases.

[94] Chapter 9.

The disputes can concern businesses or consumers. The relevant decisions are defined very broadly as 'decisions taken by the provider of the online platform on the grounds that the information provided by the recipients constitutes illegal content or is incompatible with its terms and conditions:

(a) decisions whether or not to remove or disable access to or restrict the visibility of the information;
(b) decisions whether or not to suspend or terminate the provision of the service, in whole or in part, to the recipients;
(c) decisions whether or not to suspend or terminate the recipients' account;
(d) decisions whether or not to suspend, terminate or otherwise restrict the ability to monetise information provided by the recipients'.[95]

Although the relevant decisions are widely construed, some decisions are nevertheless not subject to Article 21 DSA. For instance, typical disputes between consumers and marketplaces about the quality of performance, such as services, goods, or their delivery, are not subject to external appeals. The typical disputes between traders and marketplaces, in contrast, are more likely to be subject to them. Removal of their offerings, deduction of penalties, hiding of offers, or suspension of accounts all can qualify.

On services like Wikipedia that rely on substantial community content moderation, only moderation that is conducted by the provider itself is relevant. Unless the community is specifically instructed by the provider, community decisions are unlikely to be attributable to the provider (see Chapter 11.2.4).

11.5.3 Who counts as an ODS expert?

In the ODS certification process, national DSCs must assess the expertise of the candidate bodies. According to Article 21(3)(b): 'it has the necessary expertise in relation to the issues arising in one or more particular areas of illegal content, or in relation to the application and enforcement of terms and conditions of one or more types of online platform, allowing the body to contribute effectively to the settlement of a dispute'. In addition, Article 21(3)(e): 'it is capable of settling disputes in a swift, efficient and cost-effective manner and in at least one of the official languages of the institutions of the Union'. Thus, the expertise concerns three main areas: language, illegal content, and terms and conditions.

It is inevitable that ODS bodies will delimit their expertise into sub-areas because no single body can engage experts with track records covering all the areas of law or platforms.

The two main types of ODS bodies are likely: (1) *area-specific ODS bodies* that focus on specific types of illegal content, such as hate speech, child abuse material, or copyright infringements; (2) *platform-specific ODS bodies*, such as Facebook-related, Twitter-related, etc. In addition, the ODS bodies might further narrow down their expertise through the language of the disputes (e.g. Slovak, Finish, etc.). In theory, ODS bodies could limit their expertise further. However, they risk having the fee compensations being challenged by the platforms whenever their expertise would be overstepped only a little. This is because they benefit from the fee mechanism only within the scope of their certification.

In the certification process, the national DSCs will be mostly looking into the composition of the experts who are engaged by the ODS candidate bodies. An ODS body that claims to specialise in hate speech must be able to demonstrate that its experts have a track record of assessing such subject matter. The DSA does not require legal expertise. However, given that the key activity will consist in resolving interpretation disputes, some subject matter expertise might be required, even though legal education might not be

[95] DSA, art 20(1)

necessary. Thus, experts working at NGOs that were previously notifying hate speech on specific platforms could probably submit their work experience as relevant. National DSCs should, however, be careful whom they assess to be 'experts'.

Part of the expertise assessment could be how ODS bodies screen and train their own experts, including how they onboard new experts who join their ranks to keep the quality stable. Because it would not be practical for DSCs to reapprove certification every time the expert composition at ODS bodies significantly changes, one of the ways how to assess expertise would be to ask ODS bodies to come up with their *own criteria* that describe their present and future experts. This will force ODS bodies to think hard about how they define their expertise and provide DSC bodies with a simple checklist of criteria on which they potentially can re-assess or withdraw the certification.

11.5.4 What content rules apply in the ODS procedure?

The ODS body cannot have its own substantive rules for disputes. With every dispute, it must defer to the policies of a particular online platform. Therefore, the ODS bodies do not take away the rulemaking power of the platforms. Thus, in a dispute concerning hate speech, the contractual policy of a particular platform, along with state rules on hate speech, is determinative. The same applies to remedies. If an account suspension dispute concerns copyright infringement, the first source of law is the contractual arrangement between the platform and its users. Thus, if the policy contractually constrains some legal practices, that is the starting point. If policy only reiterates some legal concepts from the legal system, such as parody, copyright norms can be applied to determine the meaning.

The applicable content rules are one of the difficult issues that the industry will need to resolve in the coming years. Without a doubt, the dynamics under the ODS might change the drafting strategies of platform providers, who have thus far preferred to draft their own special policies. Post-DSA, they might decide to create industry-wide templates on specific issues to cut down on costs. In the absence of such templates, the idiosyncrasy of each of the platforms must be followed by ODS bodies.

One of the open questions of the ODS procedure is how much power it gives to the experts to declare some of the contractual arrangements to be *inapplicable* due to their conflict with the underlying legal rules. For instance, if EU consumer law considers some arrangements illegal, should an expert ignore them, or consider them relevant but refuse to apply them?

As will be explained in Chapter 12, review under Article 14(4) arguably allows the ODS bodies to set aside as invalid provisions which conflict with the underlying principles of fairness. However, the same argument does not apply to national laws stipulating illegality. The constitutionality review of legislation, even if inter partes, is reserved to the judiciary. The experts can always signal their views as obiter dictum, but their legitimacy is rather weak to to take on the role of arbiters of the constitutionality of national law - even if such decisions are not legally binding.

11.5.5 Is the ODS procedure binding?

ODS decisions are not contractually or otherwise legally binding for its parties (Article 21(2)). Instead of an arbitration model, the DSA opted for a non-binding mediation structure.[96] Providers must 'engage, in good

[96] While the Commission proposal established that online platforms 'shall be bound by the decision taken by the body', several Member States (in particular Finland) raised concerns as regards the constitutionality, under their respective legal systems, of a

faith, with the selected certified out-of-court dispute settlement body with a view to resolve the dispute' (Article 21(2)). Only if an identical dispute to the one at stake 'has already been resolved', can they refuse to 'engage' with an ODS body. An identical dispute is a dispute concerning the same information and the same grounds of illegality or incompatibility of content (Article 21(2)). ODS procedures do not affect judicial or other legal remedies that the eligible parties might have at their disposal (Article 21(1), third sentence).

In my view, the effect of ODS decisions is best described as 'systemically but not individually binding'. This means that parties can always disregard the individual outcomes, but they cannot systematically ignore the mechanism without an explanation. Systematic ignorance of ODS decisions, which cannot be justified, can lead to a violation of Article 21 DSA. Such providers cannot be said to 'engage in good faith'.[97] The DSA does not specify the cases in which non-compliance is justified; however, potential examples could be in my view jurisdiction overreach, abuse of the procedure or pending objections to the certification. These are objective reasons. But even if platforms disregard the decisions, they must pay for the procedure if they lose.

While it is true that nothing stops platforms from ignoring ODS decisions in individual cases, if such instances pile up, they can violate the DSA. For VLOPs, such refusals would be scrutinised very closely in their annual risk assessments, not to mention the impact that any refusals can have on the public relations and political dimensions. While a refusal to implement an ODS decision can be an important check on possible overreach by ODS bodies or the failure of DSCs to properly certify ODS bodies, if platforms have no justification and still refuse to engage, they risk violating Article 21(2) of the DSA.

Nevertheless, the force of the 'systemically but not individually binding' effect should not be overstated. Unlike in other areas (e.g. aviation), the substantive rules are not imposed by the legislature solely. Under the DSA, providers still formulate the terms and conditions of their service. Thus, they largely set the rules of the game, within the boundaries of what is made possible by the otherwise applicable legal system. Even if the decisions were binding, providers could have simply changed the underlying rules at their will. Subject to the applicable legal system, the platforms retain the monopoly on the formulation of the underlying rules for their services; however, they share the power to interpret their own rules with ODS experts (and courts).

In other words, providers still hold a pen on their terms and conditions, but going forward, they will cease to have a final word on the interpretation of their words in more cases.

Judicial or administrative authorities can always arrive at their own interpretation. But the costs, formality, and speed of such state procedures in many EU countries are unlikely to attract the same amount and type of disputes. This is very similar to how domain name disputes progress in many countries. Domain name ADR decisions are rarely challenged in courts.[98] Most often, only complex cases go to courts, while simpler cases are resolved by ADRs. It is even common for ADR bodies to have a special clause in their own

binding system, as it would be interpreted as impeding access to a judge, Commission, 'COM(2020) 825 Final' (n 53), art 18(1); Presidency of the Council of the European Union (n 60), 185, 191.

[97] Some colleagues argue that good faith efforts only extend to participation but not to outcomes (Saulius Kaleda, Paul-John Loewenthal and Folkert Wilman, Digital Services Act: Commentary (forthcoming, OUP 2024) § 21). I personally find it to be a too narrow reading which undermines the effectiveness of the mechanism. The provision speaks of 'engag[ing], in good faith, with the selected certified out-of-court dispute settlement body with a view to *resolving the dispute*'. Thus, it appears to me that the scope is broader.

[98] Annette Kur, 'UDRP: A Study by the Max-Planck-Institute for Foreign and International Patent, Copyright and Competition Law' (Max Planck Institute 2002) <https://www.zar.kit.edu/DATA/projekte/udrp_705937a.pdf> accessed 29 August 2023.

procedural rules, which provides that they would suspend self-execution of the ODS decision by a domain name authority, in anticipation of an application to the court within a particular time limit.[99]

11.5.6 Who qualifies as an ODS body?

ODS bodies utilised for the purposes of the DSA must be certified by national Digital Service Coordinators (DSCs). Each ODS body that fulfils the criteria is issued with a certification for a maximum period of five years. The certification decision is an administrative decision which should be subject to judicial review on a national level.[100] In the certification process, the ODS body must demonstrate the following:

a) Its impartiality and independence, including from providers, their users, and notifiers;
b) Necessary expertise in the subject matter area, such as area of illegal content, terms and conditions of specific services or platforms;
c) Its members' pay is not linked to the outcome of procedures:
d) It operates an electronic platform which can handle the disputes
e) It is capable of settling disputes in a 'swift, efficient and cost-effective manner' and in at least one of the official languages of the institutions of the EU; and
f) It adopted 'clear and fair rules of procedure that are easily and publicly accessible'.

Each certification decision specifies the expertise and language(s) of the ODS body. On an annual basis, ODS bodies report to their DSCs. They must report on their functioning; in particular, the report should at least cover the number of cases, their outcomes, length, and any encountered difficulties. National DSCs can request additional information. DSCs then use this information to compile bi-annual reports about the functioning of the ODS ecosystem, identifying best practices and making recommendations as to how to improve the system further (Article 21(4)). The list of eligible ODS bodies will be aggregated and published by the European Commission (Article 21(8)).

Since ODS bodies must be 'independent' from providers and their users (including content creators and notifiers), the existing ODS forums, such as Meta's Oversight Board, are unlikely to qualify unless they restructure their operations.[101] This is because the organisation behind it receives funding from the provider, and some of its members are appointed by it.[102] In contrast, the dispute settlement systems operated by World Intellectual Property Organisation, or even many other national bodies designated to operate domain name' dispute systems, are unlikely to have the same problem. These systems are not funded by the domain name authorities but by complainants. The funding thus does not come from any providers or notifiers. At the time of writing, WIPO Arbitration and Mediation Centre is preparing its copyright-focused system that it plans to certify in one of the Member States.

---

[99] Even in such a system, the period thus stipulated is obviously not binding upon the court. However, it gives a chance to the unsuccessful party to prevent the implementation of an ODS decision.

[100] See ECtHR's case law on the issue of the right to a fair trial under Article 6 ECHR, *Ramos Nunes de Carvalho e Sá v Portugal* App nos 55391/13, 57728/13, and 74041/13 (ECtHR, 6 November 2018), paras 177-181.

[101] For the same view, Aleksandra Kuczerawy, 'Social Media Councils under the DSA: A Path to Individual Error Correction at Scale?' in Matthias C Kettemann and Wolfgang Schulz (eds), *Platform://Democracy: Perspectives on Platform Power, Public Values and the Potential of Social Media Councils* (Leibniz Institut für Medienforschung | Hans-Bredow-Institut (HBI) 2023) <https://doi.org/10.21241/ssoar.86524> accessed 4 September 2023.

[102] Oversight Board, 'Governance' <https://www.oversightboard.com/governance/#governance> accessed 4 September 2023.

11.5.7 Who pays for ODS?

The innovative component of Article 21 is its payment structure. The payment structure of each ODS depends on who sets it up. Only two types of fees are possible: a) nominal fee, or b) no fee (Article 21(5)). The fee structure must be known upfront. For-profit ODS bodies will likely require a nominal fee, paid by the complainant (i.e. either the content creator or notifier). State-funded ODS bodies might require very low nominal fees or offer ODS at no cost to the complainants. The DSA envisages fees that are paid by complainants (content creators or notifiers) and providers.

The fees charged shall be reasonable and cannot exceed 'the costs incurred by the body'. This reference arguably means nothing else other than that the fee must reflect the costs of running the ODS. The original Commission's proposal was clearer in this respect when it elaborated that '[t]he fees charged by the body for the dispute settlement shall be reasonable and shall in any event not exceed the costs thereof'.[103]

Typically, each ODS has two main costs: (1) operational costs of running the digital tools and educating the pool of its experts; and (2) costs of experts who decide its individual cases. In domain name ADR bodies, operational costs usually account for 1/3 and the expert costs for 2/3 of the costs.

Since the fee structure determines the financial independence of the body and the potential bias of its experts, it should be reviewed as part of the certification process.

In theory, there are three potential fee structures: (a) the complainant pays a nominal fee, (b) the state pays for the entire system, or (c) the complainant and provider pay two nominal fees. Unlike the first and second options, the last option — provider participation fees — is likely to be incompliant with the rules about fee-shifting and independence of ODS bodies in many cases. Consider a case where an ODS body charges 20 euros to start a complaint and 200 euros for mere platform participation. Fee-shifting (Article 21(5)) in case of success is rendered useless because the punishment for a wrong decision is made irrelevant. Such an ODS body would have an incentive to hear many disputes to attract consumers even if it cannot help them. While it cannot be excluded that if such schemes can with additional safeguards, they could avoid one-sided incentives, the fee structure with provider-participation fees is inherently prone to abuse. The regulators should thus review it carefully and seek stronger justification for its existence. A fee structure that makes it irrelevant whether the provider wins or loses would go against the rules about fee-shifting and compensation in cases of abuse.

The fees charged have an important function.[104] Since the fee paid to ODS must be reimbursed in case of the complainant's success ('in favour of'),[105] this is a small punishment for the provider for its original mistake. At the same time, the risk of losing the fee paid when the complainant is not successful also discourages and filters abusive or meritless appeals, since only success is rewarded by the reimbursement. The compensation structure partly de-risks the dispute resolution for complainants by making sure that even if the provider prevails, the complainant will not be obliged to reimburse it for its costs; the only exception being an instance of the abusive complainant who 'manifestly acted in bad faith' (Article 21(5)). Such bad actors might be obliged to pay fees and other related reasonable expenses of *providers*.

---

[103] Commission, 'COM(2020) 825 Final' (n 53) 54.

[104] For background, Fiala and Husovec (n 17).

[105] DSA, art 21(5).

Any ODS body preparing its fee structure must consider: a) its costs, and b) the demand for its ODS services at distinct price points. The main cost for any ODS body is the labour of experts who prepare their decisions. To improve cost efficiencies, the ODS body can assist experts with technology, and flexible rules that do not require elaborate explanations. However, both practices, if stretched too far, can cost the ODS body its certification. At the same time, the higher the fees, the fewer disputes the ODS body is likely to attract because they must convince the future complainant that their services are worth it. As a result, the ODS bodies will need to navigate the trade-offs by tailoring their offerings. One way to do so would be to employ a tiered fee structure that charges according to the type of complainant and effectively cross-subsidises more lucrative disputes (e.g. e-commerce trader account terminations).

Ultimately, any optimisation of ODS operations must successfully juggle not only real costs and legally acceptable fee structures but also acceptance of decisions by providers and regulators who certify them. ODS bodies that tilt too much into one direction can lose their acceptance by providers, and thus become less interesting for the complainants who seek to correct the mistakes. If all providers refuse their decisions with proper justification, they can not only lose interested complainants but also certification.

On the other hand, the state-run free-of-charge ODS system runs a risk of being overwhelmed by complaints that have little prospect of winning, as their complainants do not face the risk of losing the fee. At the same time, such a system can be an important way to incentivise complaints by individuals who have no means, or whose complaints have low value to the complainants but have high social significance. In any case, the DSA also provides other ways to support the same goal. The state can support similar ODS cases through specialist non-profit independent organisations (Article 86) or trusted flaggers (Article 22), by bearing their fees. In my view, this approach is preferable because it allows better targeting of support.

Apart from the fee itself, Article 21(5) also foresees reimbursement of 'any other reasonable expenses that [the complainants] have paid' in case of the complainant's success. Some have interpreted this to mean that providers will be obliged to pay the full cost of legal representation.[106] In my view, such a reading is not very convincing.

Firstly, unlike in the pre-litigation phase, ODS does not require legal assistance. The point of ODS is to get an *expert view*. The fact that the complainant obtained expensive legal assistance cannot mean that such expense is necessary or reasonable given the merits of the case. Second, given that the DSA places clear limits on fees that can be charged by an ODS body, it would be inconsistent to allow costs that exceed the fees by several multiples to be reimbursed along with the fee. Any reasonable costs incurred by complainants must be reasonable given the forum (ODS), its costs and legal weight, and the complexity of the case. If the fee costs several hundred euros, the reasonable costs can hardly cost thousands. Thirdly, such 'reasonable costs' are not meant to provide a back door to damages caused by the original content moderation decision. Damages must be sought through judicial routes or negotiated settlements. This is again consistent with how similar systems operate in areas such as domain names.

ODS bodies have no foreseen role in estimating 'reasonable costs'. This leaves the system with several options: (1) ODS bodies take over the entire task of cost clearance (fees, and reasonable costs), (2) ODS

[106] Daniel Holznagel, 'A Self-Regulatory Race to the Bottom through Out-of-Court Dispute Settlement in the Digital Services Act' (*Verfassungsblog*, März 2022) <https://verfassungsblog.de/a-self-regulatory-race-to-the-bottom-through-art-18-digital-services-act/> accessed 4 September 2023. In my view, such interpretation could easily stumble upon constitutional limits and is thus less likely to be favoured by the courts.

bodies entirely leave it to the parties to collect compensation from each other, or (3) ODS bodies only clear the fees but not reasonable compensation. Unless DSCs express a strong preference, it will be up to ODS bodies to decide in which configuration they want to operate.

As mentioned, the complainants typically bear the fees charged by ODS bodies that charge a nominal fee, subject to possible reimbursement. However, an ODS body that does not get any assurances or deposits from the platform, risks not being able to collect reimbursements of fees and costs. If an ODS body leaves it to its complainants to collect such monies, this can undermine the attractiveness of ODS as a service. Successful complainants who were not reimbursed could complain to DSCs, for sure, but their position remains weaker than that of ODS bodies that aggregate many disputes.

It is therefore preferable that ODS bodies assume the role of complex reimbursement clearance. In the absence of specific agreements with platforms, they could request the platforms to put money corresponding to a nominal fee and capped reimbursement of reasonable costs into an escrow. Thus, as the procedure starts, ODS bodies would have complainants' and platforms' contributions secured, and only release the funds to the party depending on the success of the dispute. If the platform would refuse to put the money into escrow, the ODS body would still hear the disputes but warn the complainant about the situation.

In my view, the best outcome would be if the ODS body assumes the full role of fee reimbursement but caps the reasonable costs at the percentage of the original fee (e.g. 50%). This allows the ODS bodies to better adjust the expectations of the complainants because they know how much they can get reimbursed in the event of success. If the winning party thinks it is owed more, it can always try to seek such an amount bilaterally. ODS bodies could specify in their rules under which conditions they recognise such reasonable costs (e.g. engagement of an expert, evidence, or other expenses).

The question remains how to resolve cases where providers are unwilling to put the money into an escrow. The dilemma is as follows. If platforms lose the cases, they ought to pay to complainants. If ODS bodies leave it to the platform-complainant relationship, there remains the issue that this can undermine the confidence in the system. If ODS bodies absorb the loss and try to collect money from the platforms, they assume additional risks and can potentially undermine their own independence.

Lastly, the DSA does not specify how the fee reimbursement rules apply when success is only partial. Since Article 21 employs the phrase 'in favour of', the fee reimbursement should take place regardless of the amount of success. However, 'reasonable expenses' could be arguably adjusted for the scope of success.

11.5.8 What is a fair ODS procedure?

While the exact fashioning of the procedure is in the hand of ODS bodies, it must comply with the DSA and other applicable laws. Article 21(3)(f) states that: 'the out-of-court dispute settlement that it offers takes place in accordance with clear and fair rules of procedure that are easily and publicly accessible, and that comply with applicable law, including this Article'. Thus, the overarching criterion for procedural rules is that of fairness. There are very few explicit rules that ODS bodies must follow.

Their decisions must be made available to the parties within a reasonable period not exceeding 90 days after the receipt of the complaint. Only exceptionally, in very difficult cases, can the period be extended to 180 days (Article 21(4)). The process takes place entirely online. Other questions can be flexibly designed

by the ODS bodies. DSCs thus will need to assess the fairness of the proposed procedural rules holistically. Such rules clearly need to be part of the application for certification.

The DSA does not even specify who shall be invited to participate in such disputes. It is clear that the platform and the complainant must be part of the procedure. However, how about content creators whose account the notifier is contesting or a notifier whose notification content creator is contesting? The DSA does not require their presence. One can argue that it is in the platform's best interest to engage such parties because providers are in danger of losing the case and having to pay. That said, including such parties can also be very difficult for ODS bodies logistically. The simplest way to approach this would be for the ODS bodies to generally proceed without further parties, but incentivise the platforms to collect such views, and/or open up the process to such parties to comment on the submitted disputes.

However, ODS bodies are likely to view any additional procedural turns as a costly exercise. The DSA does not even ban secrecy, and thus the ODS bodies can easily opt for secret, or semi-secret decision-making that is cheaper to operate. This, per se, is not a problem, as long as ODS bodies retain some level of accountability for how they decide the disputes (e.g. through high-level summaries, and collection of detailed statistics that the DSCs will need). Some level of secrecy can be an enhancing feature for certain disputes, but it can also be counterproductive for other complainants, especially if they lose and are bound by strong non-disclosure rules. Preferably, the complainants should always have a choice if they want to have their procedure subject to secrecy or not.

11.5.9 What can an ODS body decide?

The ODS body can only decide to *undo* the platform's decision. It cannot award damages or order some other specific performance. While nothing stops ODS bodies from making recommendations, these have little legal consequences for the providers. Such recommendations can have, however, some legal value for regulators (e.g. to assess their compliance with Article 14). The findings of ODS bodies thus can also feed into the annual risk assessment. They can also file complaints before DSCs if they observe some problematic behaviour.

If ODS bodies take over the task of *cost clearance*, they can make decisions about the fee and reasonable cost compensation. They could also make decisions about the presence of bad-faith actors. If the money is put into escrow by all the parties, cost clearance is made much more efficient. Thus, the clearance model of ODS could be beneficial even to online platforms because they can easily enforce cost reimbursement against bad actors who try to abuse the ODS process.

11.5.10 What if an ODS body becomes a rogue?

The certification process should ensure that ODS bodies have sufficient expertise and independence. However, the DSCs should be aware that ODS bodies could try to abuse the system. For instance, one can imagine the possible emergence of ODS bodies that are ideologically biased and only hire experts who decide in a particular direction and then market themselves as such (e.g. right-wing, or left-wing friendly ODS bodies), or ODS bodies who encourage their experts to decide against providers, attract complainants, and collect their reimbursements.

The DSA places several safeguards in place to prevent this. The system of checks of balances relies upon ODS bodies, platforms and DSCs constantly looking over each other's shoulders to work.

First, DSCs are tasked to continuously review the independence of ODS bodies, and their internal fee and payment structure. They can revoke the certification at any point, following their own investigation or complaints of third parties (Article 21(7)). Second, DSCs should receive a steady stream of reports, including about how the cases are handled, so any bias can be detected rather early, especially when the DSC benchmarks results against those of other ODS bodies. Third, the providers who are obliged to 'engage in good faith' with ODS bodies might refuse to implement their decisions with the argument that the decisions are not impartial until a DSC decides about the pending investigation of an ODS body.

While the DSA leaves providers with no room for withholding the reimbursement of fees, if an ODS body loses its certification, it might arguably equally lose its due fees. The DSA does not deal with this situation. The reimbursement due in case of loss of certification might depend on the date from which the certification is lost. DSCs could, in theory, back-date their decisions to the point when the evidence of abuse exists to prevent rogue ODS bodies from reaping any benefit. Alternatively, it could fine the rogue ODS body and distribute the fines as compensation to the affected parties.

The DSA does not address the question of liability for damages under national law. However, if an ODS body runs a system in accordance with its rules and certification, it clearly cannot be liable for its functioning to the providers or other actors. However, if an ODS body loses such certification due to abuses of the process, the potential liability is probably not pre-empted by the DSA.

11.6 Conclusion

Articles 16 to 21 only focus on procedural fairness. Their goal is to impose minimum standards of due process on companies that annually handle billions of decisions about Europeans. The discussion has shown that while the DSA is marked by some imprecisions and potentially ambiguous language, the general direction of travel is very sensible. Obviously, whether and how the system works eventually, is an empirical question that I can hardly properly evaluate before the framework starts being used. One of the critical issues will be to educate and convince individuals to use their newly gained rights.