

Out-of-court dispute settlement mechanisms for failures in content moderation

by **Federica Casarosa**

Abstract: Content moderation is at the core of online platform activities. Many platforms allow users to post content that may or may not comply with the terms of service or that may violate national laws. In order to avoid these violations, online platforms have started to monitor content both *ex post* and *ex ante*. However, mistakes may still (frequently) happen.

In order to allow users to effectively contest decisions and compel platforms to restore content or accounts after erroneous decisions, online platforms should provide adequate due process mechanisms to appeal and seek redress. The EU has addressed this point by including specific provisions in the recently adopted

Digital Services Act (DSA). In particular, Article 21 provides that complaints against online platforms can also be resolved through out-of-court dispute settlement mechanisms provided by certified bodies.

After analysing the role of online platforms in content moderation, this essay focuses on the types of dispute resolution mechanisms envisaged in the DSA. Assessing, on the one hand, the proposed criteria for effective out-of-court dispute settlement bodies according to the principles of fairness, accountability, independence and transparency and, on the other hand, the shortcomings that emerge from the certification mechanism defined in the DSA.

Keywords: Online Dispute Resolution; Digital Services; Content Moderation; Certification; Harmonisation

© 2023 Federica Casarosa

Everybody may disseminate this article by electronic means and make it available for download under the terms and conditions of the Digital Peer Publishing Licence (DPPL). A copy of the license text may be obtained at <http://nbn-resolving.de/urn:nbn:de:0009-dppl-v3-en8>.

Recommended citation: Federica Casarosa, Out-of-court dispute settlement mechanisms for failures in content moderation, 14 (2023) JIPITEC 391 para 1

A. Introduction

1 Content moderation is at the core of online platform activities.¹ Every platform that provides a hosting service online allows users to post and disseminate content that may or may not comply with the

* Part-time professor at the Centre for Judicial Cooperation, EUI. Visiting Fellow at Mazaryk University. This contribution is based on research activity carried out in the framework of the DG Justice-supported project the e-Justice ODR scheme (GA n. 101046468).

1 Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press, 2018).

platform's terms of service or, even worse, may be in contrast with national laws applicable in the user's country or in the country of establishment of the hosting platform. In order to avoid violations, online platforms have started to control content both *ex post*: by reducing and minimising the dissemination of unlawful content, and *ex ante*: by employing a preventive mechanism able to screen and eventually hamper uploads of content even before they are published.²

2 See James Grimmelman, 'The Virtues of Moderation' (2015) 17 *Yale Journal of Law & Technology* 42. Note that automated content filtering has been used since the first years of internet development, as many tools have been deployed to analyse and filter content and among them the most

- 2 Content moderation, which is more and more frequently carried out using technical tools that include artificial intelligence systems, was not originally part of the legal obligations involved in the services provided by online platforms. In fact, online platforms were (and still are) not required to verify the content available on their platforms, as, for instance, the editors of online newspapers are. Nonetheless, as we will see, it has become a norm as a result of both economic decisions and incentives provided by policymakers.
- 3 However, content moderation is definitely not free from flaws depending on the ability of technological tools to recognise the substance of the content analysed (e.g. whether or not it qualifies as hate speech or aggressive expression) and also the context in which the content is expressed (e.g. a quotation from another person, a joke or a verbal attack). If these factors are not correctly evaluated then mistakes may occur which have a subsequent effect on the choice of the online platform to remove or disable access to the content.
- 4 What happens if content is wrongly removed? Many examples can be recalled, including the decision by Facebook to remove the well-known photo of the so-called ‘Napalm girl.’³ When public outcry points out the mistake it is easy to restore the status quo. Eventually this may also help technology improve, as in the above example the algorithm used by the social platform learned that the specific photo was not to be deemed to be pornography.
- 5 However, several less dramatic cases may emerge, leaving users to decide whether it is worth starting a quarrel with an online platform over why content has been removed. It must be acknowledged that some platforms have already started to provide different forms of resolution. For instance, the Facebook Oversight board employs a procedure applicable to a selected number of complaints⁴ and

common and well known are those for spam detection or hash matching. For instance, spam detection tools identify content received at one’s email address, distinguishing between clean emails and unwanted content on the basis of certain sharply defined criteria derived from previously observed keywords, patterns and metadata. See Thamarai Subramaniam, Hamid A. Jalab and Alaa Y. Taqa ‘Overview of Textual Anti-spam Filtering Techniques’ (2010) 5 *International Journal of Physical Science* 1869.

- 3 See Hortense Goulard, ‘Facebook accused of censorship of ‘Napalm girl’ picture,’ 9 November 2016 <<https://www.politico.eu/article/norwegian-prime-minister-facebook-wrong-to-censor-vietnam-war-picture/>>.
- 4 Kate Klonick ‘The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free

the YouTube Content ID claim mechanism is an automated tool that is triggered by an inclusion of copyrighted material on the platform.

- 6 Recently the European Commission has addressed the issue of a reliable complaint handling mechanism that should reduce the negative impact of erroneous removals of content. The recently adopted Digital Services Act (DSA)⁵ addressed the point by including specific provisions. In particular, Art. 20 provides that for decisions on content removal, suspension of service and account termination the internet platform should make available a free internal electronic complaint handling mechanism. This should not only be automated but also have human oversight. Alternatively, Art. 21 provides that complaints against online platforms can also be resolved using out-of-court dispute settlement mechanisms provided by certified bodies. In order to verify if the solution proposed in Art. 21 DSA will be an efficient tool to resolve cases of erroneous decisions by online platforms we need to clarify which standards are adopted in the legislation.
- 7 This contribution will therefore first analyse the role of online platforms in content moderation (Section B). Subsequently, it will describe the type of dispute resolution mechanisms envisaged in the DSA (Section C), assessing on the one hand the proposed criteria for effective out-of-court dispute settlement bodies according to the principles of fairness, accountability, independence and transparency and, on the other hand, the shortcomings that emerge from the certification mechanism defined in the DSA. Conclusions follow.

B. The role of online platforms in content moderation

- 8 The starting point to understand the role and obligations of online platforms regarding content moderation is the Directive on electronic commerce 2000/31/EC,⁶ which will be applied at least until 2024.⁷ In its Art. 14, this directive classifies online

Expression’ (2020), *The Yale Law Journal* 129, 2450.

- 5 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).
- 6 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (‘Directive on electronic commerce’).
- 7 As will be explained later, the new legal framework provided

platforms as Information Society Service Providers (ISSP), and in particular as hosting providers.⁸ Hosting providers are only exempted from liability for the content they store if they have neither actual knowledge of illegal activity or information nor awareness of facts and circumstances from which illegal activity and information are apparent. Only if they obtain such knowledge or awareness are hosting providers obliged to act expeditiously to remove or disable access to the information through a notice and take-down procedure. As a result, hosting providers are treated as pure passive and neutral actors that should not interfere in the storage and transmission of online content.⁹ The Directive on electronic commerce goes even further and in Art. 15 it excludes an obligation to *ex ante* monitor content. This article has been further clarified in recent CJEU case law.¹⁰ In its *Glawischnig-Piesczek v. Facebook* decision, the court affirmed that there is no violation of the prohibition of a monitoring obligation in Art. 15(1) of the Directive on electronic commerce even if a national court orders a platform to prevent the publication of “information with an equivalent meaning.”¹¹

- 9 However, seeing hosting providers as pure passive intermediaries is now an outdated vision of their role. Hosting providers still distribute user content and facilitate user interactions, although they are now more and more able to intervene in the experience that users have of their online activities.¹²

in the proposed Digital Services Act will depend on the date of its adoption by EU bodies. After it is published, its rules will apply 15 months after its entry into force or from 1 January 2024, whichever is later. See European Commission, ‘The Digital Services Act: ensuring a safe and accountable online environment,’ available at <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en>.

- 8 According to Art 14, hosting “consists of the storage of information provided by a recipient of the service”.
- 9 Directive 2000/31/EC, Recital 42.
- 10 CJEU, Case C-18/18, *Eva Glawischnig-Piesczek v. Facebook Ireland Limited* [2019] ECLI:EU:C:2019:821.
- 11 Case 18/18 (n 11) para 46. See Federica Casarosa ‘When the algorithm is not fully reliable: the collaboration between technology and humans in the fight against hate speech,’ in Oreste Pollicino and Andrea Simoncini (eds.) *Constitutional Challenges in the Algorithmic Society* (2021, Cambridge University Press) 298.
- 12 Note that the ‘active hosting providers’ qualification has also been developed in national jurisprudence. See the analysis of Italian jurisprudence on this point in Federica Casarosa

Online platforms now provide a wide-ranging set of services including online advertising platforms, marketplaces, search engines, social media, creative content outlets, application distribution platforms, communication services, payment systems and platforms for the collaborative economy.¹³ Although from a technical perspective, each of the above-mentioned cases has specific characteristics, from a substantial perspective delivery of these services allows online platforms to steer and control what users may disseminate.

- 10 How is this control exercised? The immediate answer is content moderation. As mentioned above, content moderation aims to verify if content hosted and stored on a platform is in line with its internal rules and conditions and with the applicable laws and regulation. This monitoring, which is exercised both *ex ante* and *ex post*, is not without consequences in terms of the choices available to users and also the ability of users to express themselves on online platforms.¹⁴ The literature has highlighted that pre- and post-publishing moderation activities have strong impacts on the exercise of users’ freedom of expression rights.¹⁵ This has also been confirmed by

‘Copyright Infringing Content Available Online – National Jurisprudential Trends’ in Agusti Cerrillo i Martínez, Miquel Peguera, Ismael Peña-López, et al. (eds.) *Challenges and Opportunities of Online Entertainment. Proceedings of the 8th International Conference on Internet, Law & Politics. Universitat Oberta de Catalunya, Barcelona 9-10 July, 2012* (UOC-Huygens Editorial, 2012) 61.

- 13 For a taxonomy of the activities provided by online platforms, see European Parliament *Liability for online platforms* (2021, European Union publications : Brussels) IV.
- 14 Content moderation, although mostly interpreted as a form of monitoring of comments, posts and speech in general, can cover all the types of content that are shared on an online platform, such as, for instance, copyrighted material in the form of text, audio, video or also goods, as is clarified in CJEU, Joined cases C-236/08 to C-238/08, *Google France SARL and Google Inc. v Louis Vuitton Malletier SA* (C-236/08), *Google France SARL v Viaticum SA and Luteciel SARL* (C-237/08) and *Google France SARL v Centre national de recherche en relations humaines (CNRRH) SARL and Others* (C-238/08), 23 March 2010, ECLI:EU:C:2010:159.
- 15 Enguerrand Marique and Yseult Marique, ‘Sanctions on digital platforms: Balancing proportionality in a modern public square’, *Computer law & security review* 36 (2020), 1; Luc von Danwitz, ‘The Contribution of EU Law to the Regulation of Online Speech’, 27 *MICH. TECH. L. REV.* 167 (2020), 185; Kate Klonick, ‘The New Governors: The People, Rules, and Processes Governing Online Speech’, 131 *Harv. L. Rev.* 1598 (2018); Orla Lynskey, ‘Regulation by Platforms: the Impact on Fundamental Rights’, in Luca Belli and Nicolo Zingales (eds), *Platform regulations: how platforms are*

cases brought before national and European courts. The European Court of Human Rights (ECtHR) in the *Cengiz and Others v Turkey* case emphasised that online platforms such as Facebook, Twitter and YouTube provide an “unprecedented” means of exercising freedom of expression online.¹⁶ This has also been confirmed more recently in *Delfi v Estonia*, in which the Strasbourg court affirmed that the internet is an “unprecedented platform for the exercise of freedom of expression.”¹⁷ Similarly, national courts have acknowledged that social networks can be equivalent to public spaces,¹⁸ although the internet may lead to “inexpensive, easy, and instantaneous means whereby unscrupulous persons or ill-motivated malcontents may give vent to their anger and their perceived grievances against any person.”¹⁹ In order to cope with these risks, it is possible to affirm that when online platforms design the moderation rules they are contextually providing their own balancing of the rights and freedoms of users on the platform itself.²⁰

11 From the point of view of online platforms, content moderation rules must strike a balance between the protection of free speech online and business interests. Clearly, platforms are eager

regulated and how they regulate us (2017, FGV Direito Rio), 83.

16 ECtHR, *Cengiz and Ors v Turkey* Apps. nos. 48226/10 and 14027/11 (ECtHR, 1 December 2015), para 49.

17 ECtHR, *Delfi AS v Estonia* App. no. 64569/09 (ECtHR, Grand Chamber, 16 June 2015), para 110. For a more detailed analysis of the jurisprudence of the ECtHR on freedom of expression with specific application to social media, see Lorna Woods ‘Social Media Jurisprudence: The European Court of Human Rights’ in Federica Casarosa and Evangelia Psychogiopoulou (eds.) *Social Media and National Courts In Europe: A Fundamental Rights Perspective* (Routledge, forthcoming 2023), 48.

18 See Italian Court of Cassation decision no. 37596/2014, in which the Court affirmed that Facebook is to be considered a place open to the public as it constitutes a ‘virtual’ place open to access by anyone using the network. For more, see Federica Casarosa and Concetta Causarano, ‘Social Media Before Higher Courts In Italy: A Thorough Adaptation of Existing Rules and Protection of Constitutional Rights Online’ in Casarosa and Psychogiopoulou (n 18), 170.

19 See Mr Justice Peart’s opinion in *Tansey v Gill* [2012] IEHC 42, as quoted by Elisabeth Farries in ‘Social Media, Fundamental Rights and Courts: An Irish Perspective’ in Casarosa and Psychogiopoulou (n 18), 152.

20 Giovanni De Gregorio and Oreste Pollicino, ‘The European Constitutional Road to Address Platform Power’ *VerfBlog*, 31 August 2021 <<https://verfassungsblog.de/power-dsadm-03/>>; Klonick (n 5).

to attract and retain users, not only in terms of the numbers of individuals registered but also in terms of content that circulates on the platform. Only if the users feel – relatively – free to express their opinions on platforms will they participate and indirectly contribute to its growth. However, in order to enhance users’ perceptions that they are part of a network of like-minded people, the online platform may promote the visibility of selected content, leading to a proliferation of so-called ‘filter bubbles.’²¹ The ability to decide what users may or may not get in contact with has been acknowledged as a concentration of power in the hands of online platforms, which has triggered a wide academic debate regarding the legitimacy and effectiveness of pre- and post-moderation activities, not only considering the standards applied but also considering the technical tools applicable to such activities.²²

12 The development of technology has also impacted the ability of online platforms to scan and identify suspicious content. Several studies have highlighted the increased adoption of artificial intelligence tools for content moderation.²³ The advantages of these technologies are of course lower costs, rapidity of analysis and, presumably, a high rate of correct evaluation of content. However, the effectiveness of the technology is limited by its ability to accurately analyse and classify content in its own context. The ability to parse the meaning of a text is highly relevant when making important distinctions in ambiguous cases such as, for instance, when differentiating between contemptuous speech and irony. For this task, the industry has now increasingly turned to machine learning to train its programs to become more context sensitive.²⁴

21 The concepts of ‘echo chambers’ and ‘filter bubbles’ were identified as risks in internet communication since early 2000 by Cass Sunstein and Eli Pariser respectively. See CR Sunstein, *Republic.com* (Princeton University Press, 2001); and Eli Pariser *The Filter Bubble: What the Internet Is Hiding from You* (Penguin, 2011).

22 De Gregorio and Pollicino (n 20); David Kaye, “Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression” Human Rights Council, A/HRC/38/35, 6 April 2018 <https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/38/35>; Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Public Affairs, 2019).

23 Emma Llanso (2019), Platforms want centralized censorship. That should scare you. *Wired*, 18 April. Available at: <<https://www.wired.com/story/tumblr-porn-ai-adult-content/>>; Tarleton Gillespie ‘Content moderation, AI, and the question of scale,’ *Big data and society* (2021).

24 Christoph Krönke, ‘Artificial Intelligence and Social Media,’

- 13 It should not be a surprise that the development of technology has not always delivered the expected results. For instance, Appelman, Quintais and Fahy highlight that content moderation systems fail to safeguard freedom of expression in particular in cases of speech by minority and marginalised groups, black activist groups, environmental activist groups and other activists.²⁵
- 14 These cases cannot be qualified as mere mistakes as the level of automation adopted by content moderation mechanisms allows online platforms to assess millions of posts (be it in textual or graphical representation) every week and even very low error rates can equate hundreds of thousands of mistakes every week.²⁶ Moreover, the biases that may – consciously or not – be embedded in the automated content moderation mechanism may lead to a risk of over-broad censorship.²⁷
- 15 In order to allow users to effectively contest decisions and compel platforms to restore content or accounts after erroneous decisions (so called ‘put-back’), online platforms should provide adequate due process mechanisms to appeal and seek redress, either by an internal complaint handling mechanism or by an out-of-court dispute settlement mechanism.²⁸

in Thomas Wischmeyer and Timo Rademacher (eds.) *Regulating Artificial Intelligence* (Springer, 2019).

- 25 Naomi Appelman, João Pedro Quintais and Ronan Fahy, ‘Using Terms and Conditions to apply Fundamental Rights to Content Moderation: Is Article 12DSA a Paper Tiger?’ *VerfBlog* 1 September 2021 <<https://verfassungsblog.de/power-dsa-dma-06/>>. See also the case of Google’s AI tool aimed at detecting toxic comments, which according to studies often classifies comments in African-American English as toxic. See Jonathan Vanian, ‘Google’s Hate Speech Detection A.I. Has a Racial Bias Problem’, *Fortune*, 16 August 2019 <<https://fortune.com/2019/08/16/google-jigsaw-perspective-racial-bias/>>.
- 26 Nicolas Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (2019, Cambridge University Press); Daniel Holznagel, ‘The Digital Services Act wants you to “sue” Facebook over content decisions in private de facto courts’ *VerfBlog* 24 June 2021 <<https://verfassungsblog.de/dsa-art-18/>>.
- 27 The UN Special Rapporteur on freedom of expression has criticised these content moderation systems for their overly vague operating rules, inconsistent enforcement and over-dependence on automation, which can lead to over-blocking and pre-publication censorship. See also Kaye (n 23).
- 28 Marta Cantero Gamito, ‘Regulation of online platforms,’ (2021) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3971076>; Giancarlo Frosio, ‘Why Keep a Dog and Bark Yourself? From Intermediary Liability to

As was mentioned above, this last option is one of the most interesting innovations in the recent proposal for a Digital Services Act. Unfortunately, as we will describe in the next sections, the out-of-court dispute resolution mechanism envisaged in the proposed legislation does not meet expectations as it not only refrains from providing more detailed standards related to the due process guarantees but it also leaves issues with addressing the certification mechanisms that should apply to out-of-court dispute resolution providers.

C. The out-of-court dispute resolution mechanism envisaged in the Digital Services Act

- 16 On 16 December 2020 the European Commission published two linked proposals addressing the governance of digital data, namely the Digital Services Act (DSA) and the Digital Markets Act (DMA). Both proposals were already envisaged in the European Digital Strategy “Shaping Europe’s Digital Future”²⁹ and were aimed at promoting fundamental rights in digital services and promoting technological innovation through the establishment of common rules for digital service providers in the European single market and beyond.³⁰ The final text of the DSA was adopted on 19 October 2022.³¹

- 17 The DSA aims to provide a dedicated horizontal regulatory framework for online platforms with rules on digital services in order to prevent unfair

Responsibility’ (2018) 26 *Oxford Int’l J. of Law and Information Technology* 1; Marten Schultz ‘Six Problems with Facebook’s Oversight Board. Not Enough Contract Law, Too Much Human Rights,’ in Judith Bayer, Bernd Holznagel, Paivi Korpisaari and Lorna Woods (eds.) *Perspectives on Platform Regulation* (2021, Nomos Verlagsgesellschaft mbH & Co. KG) 145; Amy Schmitz ‘Expanding Access to Remedies through E-Court Initiatives’ (2019), 67 *Buffalo Law Review* 89.

- 29 European Commission, “Shaping Europe’s Digital Future” (European Commission, February 2020) <https://ec.europa.eu/info/sites/info/files/communication-shaping-europes-digital-future-feb2020_en_4.pdf>.
- 30 European Commission, “Proposal for a Regulation on a Single Market For Digital Services (Digital Services Act)” COM(2020) 825 final (European Commission, December 2020), p 2, <<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en>>.
- 31 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).

practices by online intermediaries and to reduce the power of gatekeepers.³² Although the European Commission presented the DSA as an act reshaping the European rules on platform governance, its provisions addressing internet service provider liability cannot be qualified as innovative. Instead, the act makes an effort to integrate the Court of Justice's interpretation of the rules on liability.³³

- 18 The DSA follows the same distinction provided in the Directive on electronic commerce between mere conduit, caching and hosting services. In the last category the DSA includes a subcategory of online platforms that are defined as operators bringing together sellers and consumers such as online marketplaces, app stores, collaborative economy platforms and social media platforms. Online platforms can only benefit from the liability exemption contained in Art. 6(1) DSA if the following conditions are met:

(a) the online platform does not have actual knowledge of the illegal activity or illegal content and, regarding claims for damages, is not aware of facts or circumstances from which the illegal activity or illegal content is apparent; and

(b) on obtaining such knowledge or awareness, it acts expeditiously to remove or to disable access to the illegal content.³⁴

- 19 Moreover, online platforms will not lose the benefits of the liability exemption even when carrying out “voluntary own initiative investigations or other activities aimed at detecting, identifying and removing, or disabling access to illegal content,” as is affirmed in Art. 7. Accordingly, the DSA confirms that online platforms may autonomously perform content moderation activities regarding information stored and transmitted through their platforms without a need to receive prior permission from judicial or other competent authorities.

- 20 However, the DSA introduces an additional step that addresses the procedure that online platforms should follow when content is removed or disabled. Art. 17 DSA provides that users should be informed of the removal of their content or disablement of access to it at the latest by the time of the decision,

providing not only a statement of the fact, but also a clear and specific statement of the reasons that led to the platform's decision. Users should also receive information on the redress possibilities available to the recipient of the service in respect of the decision, particularly through internal complaint-handling mechanisms, out-of-court dispute settlement and judicial redress.

- 21 The inclusion of a specific rule addressing the availability of out-of-court dispute settlement mechanisms is not new in EU legislation, as several other recent European interventions have included a set of similar provisions. For instance, Art. 13 of Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services³⁵ requires “providers of online intermediation services and organisations and associations representing them to, individually or jointly, set up one or more organisations providing mediation services [...] for the specific purpose of facilitating the out-of-court settlement of disputes with business users arising in relation to the provision of those services.” Similarly, Art. 17(9) of Directive 2019/790 on copyright and related rights in the Digital Single Market³⁶ specifies that online content-sharing service providers shall provide an effective and expeditious complaint and redress mechanism, which is qualified as an out-of-court redress mechanism in cases of disputes between rightsholders asking for content removal and platforms. Another example comes from Directive 2018/1808 amending the Audio-visual Media Services Directive,³⁷ Art. 28b provides for out-of-court redress for the settlement of disputes between users and video-sharing platform providers.³⁸

- 22 The DSA identifies a more detailed architecture

32 Miriam Buiten ‘The Digital Services Act: From Intermediary Liability to Platform Regulation’ (2021) <<https://www.ibm.com/cloud/learn/machine-learning>>.

33 Caroline Cauffman and Catalina Goanta ‘A New Order: The Digital Services Act and Consumer Protection’ *European Journal of Risk Regulation* (2021) 12(4), 758.

34 See CJEU, C-324/09, *L'Oréal SA and Others v eBay International AG and Others*, [2011], ECLI:EU:C:2011:474, part. Para 113.

35 Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services.

36 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

37 Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities.

38 Jörge Wimmers ‘The Out-of-court dispute settlement mechanism in the Digital Services Act – A disservice to its own goals’ 12 (2021) JIPITEC 421

addressing the selection of potential conflicts that may emerge on the basis of decisions by online platforms. The cases are listed in Art. 20 DSA and include

- “(a) decisions whether or not to remove or disable access to or restrict 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.”³⁹
- 23 When affected with these types of decisions by the platform, Art. 21 DSA requires online platforms to indicate a certified out-of-court dispute resolution provider that can solve the dispute. The article goes further and relies on the lawful behaviour of online platforms in the procedure (“engage in good faith”) and recognises the decisions of the out-of-court dispute resolution providers as binding.⁴⁰

- 24 Art. 21(3) identifies a set of due process guarantees that the out-of-court dispute resolution provider should ensure in order to be certified. The provision lists the following elements:

- “(a) it is impartial and independent, including financially independent, of providers of online platforms and of recipients of the service provided by providers of online platforms, including of individuals or entities that have submitted notices;
- (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;
- (c) its members are remunerated in a way that is not linked to the outcome of the procedure;

³⁹ These cases can also be solved using internal complaint-handling mechanisms, as Art. 20 DSA requires online platforms to provide users with an internal complaint-handling system “for a period of at least six months following the decision.” Where a complaint contains sufficient evidence that the information is not illegal and not incompatible with the terms and conditions of the provider, the provider shall reverse the decision.

⁴⁰ Note that Article 21 only refers to online platforms.

(d) the out-of-court dispute settlement that it offers is easily accessible, through electronic communications technology and provides for the possibility to initiate the dispute settlement and to submit the requisite supporting documents online;

(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;

(f) 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.”

- 25 Moreover, in order to increase the incentives for users to submit their complaints, Art. 21 (3) specifies that if the final decision of the out-of-court dispute resolution provider results in favour of the user, the latter will receive a refund from the online platform covering the fees and expenses incurred.

I. The due process guarantees

- 26 Although several criticisms have already emerged in the literature addressing the doubts and ambiguities regarding the subjective and material scope of out-of-court dispute resolution mechanisms,⁴¹ this contribution focuses on the due process guarantees that allow the out-of-court dispute provider to be certified.

- 27 Although Art. 21(3) DSA does not elaborate in detail on the elements listed, we can interpret these elements on the basis of applicable criteria in the existing literature.⁴²

⁴¹ Wimmers (n 38); Holznapel (n 27).

⁴² See also the Manila Principles on Intermediary Liability, a joint declaration by a group of civil society organisations. These provide some minimal guidelines on what a legitimate decision-making process should include. Most relevantly, the Manila Principles require that users be given an opportunity to appeal decisions to restrict content, and these processes should be as transparent as possible without harming the privacy rights of individuals. These procedural safeguards are the hallmark of legitimate decision-making. Under the standards of the rule of law, rules must be clear, well known and fairly applied, and they must represent some defensible vision of the common good. See Suzor (n 27); Pablo Cortes, *The Law of Consumer Redress in an Evolving Digital Market - Upgrading from Alternative to Online Dispute Resolution* (2017, Cambridge University Press); Jie Zheng, *Online Resolution of E-commerce Disputes - Perspectives from the European Union, the UK, and China* (2020, Springer)

- 28 The first element defined in Art. 21(3)(a) DSA is the impartiality and independence of the body vis-à-vis online platforms and users. In this case, independence can be evaluated by means of the membership rules that are applied by the out-of-court dispute settlement body. On the one hand, members of the body should have terms of office long enough to ensure the independence of their actions; on the other hand, members should disclose any circumstances that may, or may appear to, affect their independence or create a conflict of interest. An additional element related to independence is the availability of adequate financial and human resources to carry out their functions effectively.⁴³ This is an important issue as the financial resources of an out-of-court dispute settlement body may come from the fees allocated to the parties for the settlement procedure as mentioned also by letter (c). Therefore, it may be possible that in order to attract as many cases as possible there is a risk of preferring the positions of claimants in order to incentivise their participation.
- 29 The second element defined in Art. 21 (3)(b) is the necessary expertise, which requires knowledge and competence not only concerning the legal rules applicable to the case at stake but also concerning the terms and conditions that may have triggered the decision of the online platform. This very general requirement should be framed according to the object of the dispute settlement procedure. Therefore, the members of the out-of-court dispute settlement body will have a keen understanding of the law and its application when balancing conflicting fundamental rights. Accordingly, out-of-court dispute settlement bodies should select their members from trained lawyers who are familiar not only with the applicable Union and national laws but also with the relevant case law.⁴⁴
- 30 The elements defined in Art. 21(3) (d) and (e) are connected to the provision of a settlement procedure that is easily accessible by users and that does not require them to make a high investment of time and resources. Out-of-court dispute settlement bodies may adopt several features in order to ease their accessibility by users. These clearly include the fee adopted in order to cover the cost of the procedure. However, this point is further specified in Art. 18(3), in which a clear preferential treatment for users is defined. If an out-of-court dispute settlement body decides the dispute in favour of the user then it is the online platform that bears all the costs (including any fees and procedural expenses) suffered by the user. However, if the decision is in favour of the online platform then the fees and procedural costs are allocated to each party.
- 31 Other practical elements can include the availability of sample documents able to clarify the type of information required, the availability of a (free or paid) online expert advisor and the possibility to select the type of case documentation to provide if, for instance users are only asked to fill in a template online or if the provider allows paper filings to be automatically converted into online forms. Additionally, the technological tools used for resolution of the dispute can be adapted to the preferences of the users, including mediation, blind bidding, videoconferencing, chat rooms etc. The selection of such tools may impact the ability of users to access the settlement body.⁴⁵
- 32 The last element introduced in Art. 21(3)(f) is transparency and fairness of procedure. Out-of-court dispute settlement bodies should ensure that all the steps that lead to the decision are transparent and fair.⁴⁶ For instance, there must be clear rules on the procedure to select the person in charge of deciding the dispute, the factual circumstances that the deciding body will take into account and how the documentation will be handled and stored. Moreover, attention should be paid to the power of investigation that may be allocated to the out-of-court dispute settlement body and the power of the parties to contest the results of the investigation. Finally, out-of-court dispute settlement bodies should indicate the standards that would apply regarding evaluation of content as unlawful, in particular if they rely on the national or international provisions addressing the exercise of freedom of expression online.
- 33 This element is also relevant to avoid the risk of out-of-court settlement body shopping leading to a race to the bottom. If procedures are uniformly assessed according to the criteria of fairness and transparency there will be fewer opportunities for different

236 ff; Loïc Cadet, Burkhard Hess, Marta Requejo Isidro (eds.) *Privatizing Dispute Resolution – Trends and limits* (2019, Nomos);

43 Kristina Irion, Wolfgang Schulz and Peggy Valcke, *The Independence of the Media and Its Regulatory Agencies: Shedding New Light on Formal and Actual Independence Against the National Context* (2014, Intellect, Bristol UK / Chicago USA).

44 Compare with the requirements provided by the Directive 2013/11/EU of the European Parliament and of the Council of 21 May 2013 on alternative dispute resolution for consumer disputes, in particular article 6. See the analysis in Cortes (n 42) 107.

45 Cortes (n 42) 254.

46 Compare also with the analysis of Caroline Daniels, 'Alternative Dispute Resolution for European Consumers: A Question of Access to and Standards of Justice', in Cadet, Hess, Requejo Isidro (n 42) 257, part. 287.

quality standards across Europe. Then, users may select any certified out-of-court settlement body in any Member State.

- 34 As was clarified above, Art. 21 (3) DSA only provides a short list of elements that should be evaluated. However, more detailed standards can be identified. The following may act as recommendations for a checklist that can guide evaluation of the due process guarantees.

Impartiality and independence	<ul style="list-style-type: none"> - Disclosure of financial resources - Duration of membership - Absence of conflict of interests
Expertise	<ul style="list-style-type: none"> - Selection of members (interview evaluation, prior experience, etc.)
Accessibility, efficiency and cost-effectiveness	<ul style="list-style-type: none"> - Availability of an advice portal (free or paid) providing sample legal documents and online expert advice - Case documentation (only online filings or both online and paper filings, according to the preferences of the users; paper filings are manually converted into online forms or automatically converted) - Availability of technological tools (blind bidding; videoconferencing; chat room) - Cost of procedure (initial fee)
Transparency and fairness	<ul style="list-style-type: none"> - Structure of the deciding panel (single member or more than one panel) - Selection of the members of the deciding panel (by parties; by internal allocation based on alphabetical order, least number of pending cases, language of both parties, area of expertise) - Case management (information included in each case; people able to access case files; order of appearance of cases in the list of cases) - Classification of cases (differentiation based on type of defendant; differentiation based on type of claim) - Possibility to merge cases (based on predetermined conditions; based on decisions by the parties) - Access to case files by the deciding panel and parties (all available documents; only to documents not marked as internal notes by a party) - Availability of online hearing (open to the public, on restricted access) - Possibility to opt out from the procedure

- 35 These features can provide a starting point for the evaluation by a certification body identified in the DSA, but again the rules defined in the proposed legislation are not well detailed and disregard other important issues, including security of the dispute settlement platform,⁴⁷ the guarantees in case of use of artificial intelligence tools,⁴⁸ etc. Moreover, the listed criteria leave too much room for national adaptation

47 Fahimeh Abedi, John Zeleznikow, Chris Brien, 'Developing regulatory standards for the concept of security in online dispute resolution systems', *Computer law & security review* 35 (2019) 1.

48 Hibah Alessa, 'The role of Artificial Intelligence in Online Dispute Resolution: A brief and critical overview', *Information & Communications Technology Law*, 31:3, (2022) 319.

which may run contrary to the objective of a fair and harmonised level of protection of users' rights in out-of-court dispute resolution mechanisms.

II. Certification of out-of-court dispute settlement bodies

- 36 According to Art. 21(3) DSA, in order to receive certification the existence of the list of elements that address the due process guarantees should be evaluated by the Digital Services Coordinator of the Member State where the out-of-court dispute settlement body is established.⁴⁹ The DSA sets up a certification mechanism that is fully allocated to the national authority designated by the Member States for consistent application of the DSA. However, the provisions in the DSA only indicate that Digital Service Coordinators should appraise each and every (general) requirement provided in Art. 21(3) and then the Commission of the list of certified bodies.⁵⁰ Regardless of the specificities that emerge from the fact that the out-of-court dispute settlement bodies address conflicts that may concern the freedom of expression of users, the certification process should be well-defined in order to ensure that users of the certified out-of-court dispute settlement can rely on evaluation given by the certification body.

- 37 It is evident that the certification mechanism described in the DSA lacks any additional specification in terms of the definition of applicable standards, the type of evaluation, the geographical scope of the certification scheme and the duration of the certification appraisal. This is a lost opportunity which cannot be justified by lack of knowledge or expertise, as in many other legislative interventions the Commission has engaged in a more structured description of the certification mechanism.

- 38 In EU law, there are several areas where certification mechanisms were adopted and have flourished. The Commission has fostered the use of certification in

49 The Digital Service Coordinator is defined in Article 49 DSA as the national competent authority in charge of verifying the application and enforcement of the DSA in each Member State.

50 Art. 21(8) DSA provides that "Digital Services Coordinators shall notify to the Commission the out-of-court dispute settlement bodies that they have certified in accordance with paragraph 3, including where applicable the specifications referred to in the second subparagraph of that paragraph, as well as the out-of-court dispute settlement bodies the certification of which they have revoked. The Commission shall publish a list of those bodies, including those specifications, on a dedicated website that is easily accessible, and keep it up to date."

the digital market by adopting the Cybersecurity Act (CSA)⁵¹ and has also included certification schemes in the General Data Protection Regulation.⁵² More recently, the Proposal for an Artificial Intelligence Act⁵³ describes the structure for certification of artificial intelligence systems. In all these cases the Commission has defined in a more or less elaborate manner⁵⁴ a certification procedure involving several actors and stakeholders that contribute to the definition of the standards adopted, and a detailed structure of actors in charge of accreditation and certification. The most developed is the certification mechanism defined in the CSA, which provides a clear preparatory phase for the definition of the standard and an equally detailed guidelines for the evaluation of the compliance with the standards. Accordingly, it would be the most suitable point of comparison to achieve not only consistency, but also reliability of the certification mechanism itself. In the following, the DSA procedure will be compared with the more detailed procedure described for cybersecurity certification.

- 39 In general terms, a certification scheme should involve at least two phases, a conformity assessment and an attestation of conformity, the latter being a statement that the underlying process, product or person complies with a set of pre-defined requirements that are identified on the basis of the objectives and reach of each certification scheme.
- 40 The certification scheme of the CSA is defined in a centralised process started by the European Commission involving both the ENISA and relevant stakeholders in the field which aims to achieve

the most updated level of information security.⁵⁵ The result of this process is the adoption by the Commission of the certification scheme, which may include different assurance levels (basic, substantial or high)⁵⁶ that take into account the resilience of the ICT product, process or service in the face of potential security threats based on either past experience or potential vulnerabilities.

- 41 This level of detail regarding the procedure is absent in the DSA. Although the criteria for the certification scheme are already listed in the DSA, as explained in the previous section, there are several sub-criteria that the Digital Service Coordinators may identify in order to operationalise each item in the list provided. The different approaches that may emerge at the national level may run the risk of different safeguards being provided to the users of the out-of-court dispute settlement bodies.
- 42 Another step in the procedure is the conformity assessment. Art. 58 CSA requires each Member State to designate one (or more) financially and institutionally independent authorities to oversee the enforcing of rules included in European cybersecurity certification schemes and monitoring the compliance of ICT products, services and processes with the requirements of the European cybersecurity certificates. Accordingly, the certification authorities enjoy both investigative and enforcement powers allowing them to carry out investigations (i.e. audits) of conformity assessment bodies, European cybersecurity certificate holders and issuers of EU statements of conformity to verify their compliance,⁵⁷ and in cases of infringement to impose penalties in accordance with national law.⁵⁸

51 Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), OJ L 151, 7.6.2019.

52 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016.

53 Proposal for a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final.

54 Federica Casarosa 'Cybersecurity certification of Artificial Intelligence: a missed opportunity to coordinate between the Artificial Intelligence Act and the Cybersecurity Act,' *Int. Cybersecur. Law Rev.* (2022).

55 The process is qualified as a centralised one as on the basis of the Union rolling work programme the European Commission defines the strategic priorities for cybersecurity certification schemes. Accordingly, the Commission requests ENISA to draft a candidate scheme, setting very specific goals and requirements such as the subject matter and scope of the scheme, the types or categories of ICT products, systems and services covered, the purpose of the scheme and references to technical standards and specifications etc. These requirements are to be strictly followed by ENISA in order to ensure coherence and uniformity of the certification scheme structure, taking into account differences that may clearly emerge depending on the scope, sector and context of each scheme. See Article 47-49 CSA. For a more detailed analysis of the CSA certification scheme see Casarosa (n 54).

56 See Arts. 52 (6) and (7) CSA.

57 See Art. 58 (8) (b) Cybersecurity Act.

58 See Art. 58 (8) (f) Cybersecurity Act.

- 43 No similar provision is included in the DSA. On the one hand, the legislation relies on the resources and expertise of the Digital Services Coordinator at the national level: in this sense Art. 39 DSA provides a safety net. The article acknowledges that the Digital Services Coordinators should be independent in carrying out their tasks, and that the Member States should ensure that they have adequate technical, financial and human resources. On the other hand, nothing is stated about the powers of the body regarding evaluation of the certification schemes. There is no help in Art. 41 DSA on the powers of the Digital Services Coordinators, as it lists the powers of supervision and enforcement of the rules applicable to online platforms. No mention is made regarding supervision, investigation and sanctioning powers vis-à-vis out-of-court dispute settlement bodies.
- 44 Another interesting element is the validity of CSA-based certifications, which may vary but should never exceed the maximum of four years and requires a periodic review of the certification schemes adopted. Moreover, the certification scheme has a geographical scope that covers all EU Member States. This coverage is important not only as certification is recognised in any EU country market where the producer, manufacturer or service provider sells its product, process or service, but it also implies that certification can be obtained in any EU country regardless of the physical location of the requesting company.
- 45 The same cannot be said for the certification mechanism envisaged in the DSA. The out-of-court dispute settlement body can only be certified in the country where it is established. Although Art. 18(2) DSA acknowledges that the out-of-court dispute settlement body can provide its services in other EU languages, it is not expressly mentioned if the certification is to be recognised in other countries too. This is an evident lack of foresight as the attestation of conformity provided by the certifying body should allow services to be provided across Europe. It will be difficult for an out-of-court dispute settlement body to only provide its services on a country basis. Instead, it will aim to specialise in disputes emerging in some type of platforms (e.g. social networks or C2C marketplaces) in order to provide the service to any user regardless of nationality.
- 46 The certification mechanism applied to out-of-court dispute settlement bodies would create the conditions for offering transparency and increasing trust in the certified organisation, thus reducing the risks of fragmentation and differentiation in the standards applicable. This would be beneficial both for online platforms and for users. Given the complete absence of guidance regarding this certification process in the DSA, it will be up to each national legislator to fill the gaps in the EU

legislation so as to create *ad hoc* procedures and more detailed standards.

D. Conclusion

- 47 The increasing relevance of online platform activities in users' lives has relevant consequences for the ability of online platforms to gather information about preferences, opinions, and, more generally, knowledge about us. In fact, every platform can screen and potentially filter what is disseminated online by users both *ex ante* and *ex post*. This content moderation activity is increasingly reliant on algorithms and artificial intelligence systems. However, these tools are not fool proof. There are many studies that analyse if, when and to what extent these tools make mistakes, as the subsequent effect is removal or disabling of online content.⁵⁹
- 48 Of course, mistakes can occur. However, procedures that allow users to contest decisions of the online platform should be available. Internal complaint handling mechanisms are slowly emerging, but another promising alternative is out-of-court dispute settlement mechanisms that can be in charge of resolving disputes on content removal, suspension of service and account termination. The proposed legislation in the Digital Services Act shares this position and provides in its Art. 21 that users of online platforms shall be entitled to resolve the abovementioned types of disputes also through certified bodies providing their services in the EU.
- 49 The DSA provision not only pushes towards the creation of such out-of-court dispute settlement bodies but it also requires them to ensure due process guarantees, which are listed as the main criteria for certification of them. Although this is a commendable effort by the EU bodies to safeguard the position of users vis-à-vis the increasing power of online platforms, the provisions in the DSA run short of useful guidelines, which may hamper achievement of the objectives sought.
- 50 On the one hand, the list of criteria in Art. 21 is far from being immediately applicable and will require an effort by Digital Services Coordinators at the national level to operationalise the general elements into more practical features. Can each Digital Services Coordinator define its own criteria at the national level? It is more than probable that this issue will require coordination at the European

59 Article 19, 'The Social Media Councils: Consultation Paper' (2019) <<https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf>>; Stuart Benjamin 'Algorithms and Speech' 161 *University of Pennsylvania Law Review*, 1445-1493 (2013)

level. Otherwise, the harmonisation objective would be jeopardised.

- 51 Moreover, the certification mechanism provided in Art. 21 may also be qualified as a raw structure as the guidelines addressing the definition of applicable standards, the type of evaluation, the geographical scope of the certification scheme and the duration of the certification appraisal are very limited.
- 52 This seems to be a lost opportunity as out-of-court dispute settlement mechanisms will probably flourish as they are not only present in many recent legislative acts but they will also probably emerge more and more as an alternative way to resolve cases of user dissatisfaction.⁶⁰ In this context, certification may provide a very useful signal to users regarding due process guarantees and safeguard their position vis-à-vis platforms. Moreover, it is possible that users engaging in a copyright dispute may recognise among the out-of-court dispute settlement bodies one or more that have been certified according to the DSA procedure. In this case, certification may become an added value and steer the choice of users towards this provider.
- 53 What if the DSA certification mechanism (if improved and structured in a clearer way) also becomes the standard for bodies providing their services in other legal areas? Of course, this is clearly a step that will require further legislative interventions, but it may be possible that the path set by the DSA will lead in this direction.

60 Civil Justice Council's Online Dispute Resolution Advisory Group 'Online Dispute Resolution for Low Value Civil Claims' (2015) <<https://www.judiciary.uk/wp-content/uploads/2015/02/Online-Dispute-Resolution-Final-Web-Version.pdf>>.