

CASE NOTES

Shadowbanned on X: The DSA in Action

Jacob van de Kerkhof¹ and Catalina Goanta

Utrecht University Faculty of Law Economics and Governance, Utrecht, Netherlands

Corresponding author: Jacob van de Kerkhof; Email: jj.w.vandekerkhof@uu.nl

Abstract

Small claim. Regulation 861/2007 (European Small Claims Procedure Regulation). Unfair term in terms of service agreement. Breach of contract through shadowbanning. Infringement on Articles 12 and 17 of Regulation 2022/2065 (Digital Services Act).

Keywords: Content moderation; Digital Services Act (DSA); European small claims procedure regulation; consumer law; social media

I. Legislation

Article 4(1) Regulation 861/2007 of the European Parliament and of the Council of 11 July 2007 establishing a European Small Claims Procedure, *OJ L 199, 31.7.2007, p 1-22*; Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts, *OJ L 95, 21.4.1993, p 29-34*; Articles 12 & 17, Regulation 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, *OJ L 277, 27.10.2022, p 1-102*.

II. Facts

The applicant has a premium account on social media platform X (owned by Twitter).¹ He has posted repeatedly on issues of EU policy. On 11 October 2023 he posted on the privacy implications of the Proposed Regulation to Prevent and Combat Child Sexual Abuse. On 13 October 2023 he posted an article he had written in a newspaper on the European Commission misleading citizens with disinformation campaigns and illegal advertisements. Alerted by a third party, the applicant found out that his post of 13 October 2023 was not findable on X. On 13 October 2023, the applicant emailed X questioning whether there was a search ban on his account and whether it could be lifted. In a reply on 15 October 2023, X notified the applicant that it was reviewing his request. On 14 November 2023, X notified the applicant that his account had been sanctioned by automated content moderation mechanisms analysing posts associated with child sexual exploitation, which may result in a temporary account-level restriction. On 12 January 2024, X notified the applicant that his post of 11 October 2023 was the reason for the temporary restriction on his account. and that, after review, the restriction had been lifted and there were currently no restrictions on the applicant's account. Applicant had in the meantime, on 24 October

¹ It is important to differentiate between X, the platform offered by Twitter, and Twitter, the legal entity that owns X. Twitter is party to the procedure.

2023, requested the Amsterdam District Court to decide on this matter through a small claims procedure under Regulation 861/2007.

III. Judgement

The decision deals with three main points: admissibility of the procedure under the European Small Claims Regulation (ESCR), violation of Article 17 of the Digital Services Act (DSA) and violation of Article 12 DSA. Via a small claims procedure as laid down in Article 4(1) of the ESCR, the applicant requested a declarative judgement on X having breached his user agreement, as well as having acted unlawfully by violating the DSA. The applicant requested X to send him a statement explaining the reasons for the restrictions imposed on his account, pursuant to Article 17 DSA and requested X to appoint a point of contact for recipients of their service as per Article 12 DSA. Finally, the applicant requested damages from Twitter for the period during which his account had been restricted, claiming that the restricted visibility and searchability of his account had resulted in a violation of his user agreement. The damages claimed amounted to \$1.87.²

The Amsterdam District Court agreed with the applicant that the case was admissible under the small claims procedure. Twitter had raised that the case would be too complex for such a procedure, as well as that the monetary interest in the case would supersede €5,000, firstly because a declarative judgment has an undetermined value, and secondly because the case could affect the outcome in other cases and thus cost Twitter a lot more. The Court disagreed; the small claims procedure is not only for simple claims, it is intended for claims under €5,000. The monetary interest of this individual request is below that amount, and the suitability for a small claims procedure is irrespective of potential impact on other cases.³ If a requested declarative judgment is, during examination, likely to supersede the value of €5,000, the Court will find it inadmissible under this procedure. During examination, this was not the case.⁴

The Court granted the applicant a declarative judgment that Twitter had breached the applicant's user agreement. Twitter had objected that even though visibility of his account was restricted, the most important features of the service were still usable, and therefore the agreement was not breached.⁵ Further it claimed that, under the terms and conditions, Twitter was not held to provide unlimited functionality to the applicant, including unlimited visibility of all user-generated content. Twitter reserves the right, under its terms and conditions, to limit content visibility as it sees fit.⁶ This led the Court to an *ex officio* examination of those terms and conditions under Directive 93/13/EC (Unfair Contract Terms Directive, UCTD).⁷ The District Court found that the terms and conditions relied upon by Twitter were indeed unfair. The fact that Twitter can unilaterally suspend provision to paid services without accountability for any reason and without notification is considered unfair in the context of UCTD's "grey" list of unfair terms (UCTD Annex point k).⁸ As a result, Twitter cannot rely on this provision in this case, as unfair provisions are not binding on the consumer. Further, the fact that the suspension of services occurred through automated means in trying to prevent the spread of child sexual abuse material cannot justify Twitter's

² Amsterdam District Court, 5 July 2024, *Applicant v Twitter*, ECLI:NL:RBAMS:2024:3980, para 2.

³ *Ibid.*, para 4.

⁴ *Ibid.*, para 5.

⁵ *Ibid.*, para 7.

⁶ *Ibid.*, para 8, with reference to Terms and Conditions under. 2.a.

⁷ *Ibid.*

⁸ *Ibid.*, para 9.

breach of contract, since the litigious post did not contain any such material itself.⁹ Twitter was therefore found to be in breach of contract. In light of the €85 per year subscription costs the claim does not supersede the value of €5,000, and is therefore admissible under this procedure.¹⁰

The Court subsequently examined whether Twitter acted unlawfully by violating Article 17 DSA. That Article requires providers of hosting services to provide a clear and specific statement of reasons to any recipient whose (access to) user-generated content is being restricted. Twitter contested that the restrictions placed on the applicant's account did not fall under the scope of Article 17, arguing that it was not only the specific post which was not visible, but the applicant's entire account.¹¹ The Court found, however, that a shadowban falls within the scope of Article 17(1)(a), and therefore a statement of reasons had to be provided. Twitter claimed that it had notified the applicant three times about the restriction: on 15 October 2023, 14 November 2023 and 12 January 2024. The question is whether those notifications fulfilled the requirements of Article 17(3), meaning that they provided information on the nature of the restriction, on the facts and circumstances that led to it, on whether the decision was automated, on the legal and contractual provisions relied upon, and on the available means of redress. The Court found that the information provided on the nature of the restriction, the facts and circumstances, the legal provision relied upon and the means of redress were all missing in Twitter's communication with the applicant.¹² Even in later communications, the information was too vague to enable the applicant to seek effective redress. In this regard, Twitter raised that the requirements of Article 17(3) impose on it a disproportional burden, in light of the sheer volume of user-generated content with which it deals. The Court found that especially for larger platforms like X compliance with the DSA is highly relevant, and therefore it cannot follow the disproportionality argument.¹³ The DSA is especially intended to overcome inherent risks to the service that larger platforms provide, and has a layered system that especially targets very large online platforms (>45mil active users in the EU per month *ex* Article 33(1)). Raising the size of the platform as a reason not to comply with the DSA is therefore counterproductive and not in the spirit of that regulation. However, the Court finds that since the applicant has learned about the reasons behind his restriction required by Article 17 through the course of these proceedings, he has lost legitimate interest in a declaratory judgement.¹⁴ The Court therefore cannot grant a declaratory judgement on Twitter's violation of Article 17 DSA.¹⁵

Finally, the court considered the applicant's claim concerning the violation of Article 12 DSA (obligation to provide a point of contact for recipients of intermediary services). Twitter argued that it provides a Help Center reachable through an e-mail address, which the applicant had used and through which the platform had responded in a timely manner.¹⁶ The Court underlined that the intention of the DSA is to enable swift and efficient communication between users and providers of online platforms. Recital 43 lists a number of means that could be used to do so, but Twitter has failed to adequately deploy any of those means.¹⁷ The fact that, while timely, the email of 15 October 2023 did not contain any useful information due to it being a standard reaction, led the Court to conclude that Twitter had not complied with its duty to offer users a point of contact

⁹ *Ibid*, para 10.

¹⁰ *Ibid*, para 11.

¹¹ *Ibid*, para 14.

¹² *Ibid*, para 17.

¹³ *Ibid*, para 18.

¹⁴ *Ibid*, para 19.

¹⁵ *Ibid*, para 21.

¹⁶ *Ibid*, para 24.

¹⁷ *Ibid*, para 25.

providing swift and efficient communication.¹⁸ The Court found that Twitter had indeed violated Article 12, but it could not grant applicants' request beyond providing applicant individually with a contact point. Applicant had requested that Twitter end its violation of Article 12 in a general sense, but such would go beyond the scope of the interest of the applicant. Twitter must therefore provide applicant with a point of contact with a periodic penalty on non-compliance.¹⁹

To summarise, Twitter has breached the user agreement, and must pay \$1.87 in damages. Twitter is further ordered to provide the applicant with a point of contact *ex* Article 12 DSA under penalty payment of €100 per day with a maximum of €100,000 for non-compliance. Further requests are denied.

IV. Comment

This case decided by the Amsterdam District Court against Twitter is one of the first cases in the Netherlands where a user requests a declaratory judgment on a platform's compliance with the DSA. It provides an insight into the functioning of the small claims procedure in the context of the DSA, the difficulty for applicants to complain about account restrictions, and the interplay between the DSA and other EU legal instruments such as the Unfair Terms Directive.

V. The small claims procedure

Firstly, this case presents a procedural peculiarity, as the applicant's requests were made under the ESCR.²⁰ The ESCR was adopted to provide a simple, expedient and proportionate way to claim small (below €5,000) consumer and commercial damages. Requesting compliance with the DSA, a very systemic instrument, could end up costing significantly more than €5,000, therefore excluding application of a very targeted instrument such as the ESCR. The applicant had also requested declaratory judgments on contract breach and wrongdoing by Twitter with an undetermined monetary value, which could also exceed the scope of the Regulation. The question is whether the ESCR is indeed the right instrument to use for a systemic legal framework such as the DSA. The DSA is risk-based regulation, in which the mitigation of risks for the general population using intermediary services on the internet is achieved through several procedural safeguards and transparency obligations.²¹ The ESCR is more focused on individual swift relief, which is not necessarily inherent to the systemic nature of the DSA. In situations such as these there is an avenue available in Article 53 DSA, to complain about service provider's failure to comply with the DSA with the national Digital Services Coordinator that might be more suited, because the DSC can order platforms to address user's concerns on a more structural base rather than an individual one. In this case however, the District Court was indulgent in allowing the applicant's requests: even though it is indeed possible for declaratory judgements to exceed the established value of €5,000, it would be impossible to request damages under the small claims procedure against large corporations if the potential for judicial precedent is added to the value of the proceedings. This would stand in the way of providing simple and expedient relief in the cases of small civil claims, the exact purpose of that Regulation.

¹⁸ *Ibid*, para 26.

¹⁹ *Ibid*, para 27.

²⁰ Regulation (EC) 861/2007 of the European Parliament and of the Council of 11 July 2007 establishing a European Small Claims Procedure.

²¹ Giovanni De Gregorio and Pietro Dunn, "The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age" (2022) 59 *Common Market Law Review* 473.

It remains to be seen whether the small claims procedure will be used more often, since it does not necessarily align with the systemic nature of the DSA well. This raises a question about subsidiarity: are there not less burdensome and more suitable avenues for users of intermediary services to address non-compliance with the DSA? Should a judicial authority determine whether such avenues should be exhausted before agreeing to admit a small claims or even regular procedure? For example: if users disagree with restrictions *ex* Article 17(1), they should be able to rely on an internal complaint-handling system provided by the provider of the online platform *ex* Article 20. Should this complaint system be exhausted before relying on a small claims proceeding in order to better ensure the simple and expedient procedure foreseen by that Regulation? Next to that, the abovementioned complaint mechanism of Article 53 DSA with the national digital services coordinator is also better suited, because concerns can be addressed directly with the provider of the intermediary service without judicial intervention, but with a more structural result. It could be argued that a requirement of subsidiarity in these cases would prevent a possible over-use of small claims procedures to rectify content moderation decisions, especially since such decisions can effectively be taken at the level of the provider of the online platform. However, as is visible in this case, it can be poor communication from the platform provider that can stand in the way of effective redress, and therefore a principle of subsidiarity could obstruct access to justice for users. That latter sentiment also resonates with the fact that application of the DSA is intended to be without prejudice to judicial review, and therefore the option for users to complain with a judicial authority should not be restricted too much.

VI. Article 17 of the Digital Services Act

One of the key underpinnings of the DSA is that informing users of restrictions imposed on them or the content they have uploaded enables them to seek effective redress against such restrictions. This includes decisions about removal, suspension or termination of account services, but also any restriction on the visibility of specific items of information, such as demoting content or “shadowbanning.” As argued by Leerssen, the duty to inform users about content moderation could provide insight into restrictions that have previously been opaque.²² In this case, the District Court clarified the scope of visibility restrictions as meant in Article 17 DSA, and what information should be provided in such cases. There are some caveats to this. The first being obvious: even though a platform should notify its user that a restriction is being placed, they can fail to do so. In that event, the user must find out on their own that a restriction is being placed on their account.²³ While this may be easier in cases of content removal, it is significantly more difficult for cases of “shadowbanning.” Here, users would have to rely on information provided by third parties, e.g. noticing that their account was not findable. In many cases, however, users may never find out that their content is being restricted. Sometimes, a restriction in visibility may just be the result of the recommender system rather than a sanction, and in some cases, a platform’s search engine may just be defective, as in the case of X itself.²⁴ It can furthermore be difficult to gather evidence of a visibility restriction. In a parallel case, the applicant requested access to so-called “blacklists”, under Article 15 of the General Data Protection Regulation.²⁵ Twitter denied such lists existed in the first place; therefore

²² Paddy Leerssen, “An End to Shadow Banning? Transparency Rights in the Digital Services Act between Content Moderation and Curation” (2023) 48 *Computer Law & Security Review* 105790.

²³ Sarah Myers West, “Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms” (2018) 20 *New Media & Society* 4366, 4374.

²⁴ Even its owner Elon Musk thinks so, see: <https://x.com/elonmusk/status/1589022495189127169>

²⁵ Amsterdam District Court, 4 July 2024, *Applicant v Twitter*, ECLI:NL:RBAMS:2024:4019

no access could be provided. The Amsterdam District Court found that, even though one can reasonably assume that such lists exist, the applicant had failed to provide sufficient evidence as to their existence. Lacking sufficient access to evidence about the existence of restrictions, it is hard to enforce rights under Article 17 DSA.

A second point is that, after learning about the restriction imposed on his account, the applicant also found out that he had been subject to automated decision-making, deployed by X against child sexual abuse material. From this, one could deduce that Twitter violates another DSA provision, namely Article 24(5) on the DSA Transparency Database. The Database stores all content moderation by all online platform providers, including X. It has been shown elsewhere that compliance with the transparency database varies across platforms, and X fails to comply on a regular basis, due to underreporting its content moderation decisions.²⁶ In the meantime, X prides itself on an “artisanal” approach to content moderation, which always ensures a human in the loop. It only uploads content moderation decisions in the transparency database, which did not rely on automated detection or automated decision-making. In this case, Twitter relied on the fact that the applicant had been wrongly restricted by automated means, as a defence against liability for breaching applicant’s user agreement. In doing so, its non-compliance with Article 24(5) DSA, since it has not uploaded any automated decisions. A motivated Digital Services Coordinator or the European Commission could address these forms of non-compliance through DSA enforcement mechanisms.

VII. The interplay of the DSA and provisions and principles of the EU consumer acquis

The judgment also raises several questions with respect to the relation between the DSA and other instruments pertaining to the consumer *acquis*, such as the UCTD and the Digital Content Directive.

Before tackling this relationship, it is essential to point out that social media terms of services (ToS) are complex and confusing. Platforms often unilaterally change ToS without informing users and fail to provide a clear overview of how they may differ across jurisdictions. According to X’s own statements,²⁷ no less than 18 versions of its ToS have been applicable between 2008 and 2024. Some of these versions are more complex than others. Its most recent ToS, in force as of 29 September 2023, consist of one version applicable to users residing “outside the European Union, EFTA States, or the United Kingdom, including if you live in the United States,” and another one applicable to users residing “in the European Union, EFTA States, or the United Kingdom.” Apart from these general ToS, Twitter also has a plethora of additional ToS applicable to different digital products it has been monetising for revenue. One of these products is the “Paid Service,” subscribed to by the applicant, which has its own terms of service. The most recent version of these terms (version 7)²⁸ dates from 13 May 2024.²⁹ We assume this to be the version analysed by the Court, although no clarifications are given in the judgment.

²⁶ Rishabh Kaushal and others, “Automated Transparency: A Legal and Empirical Analysis of the Digital Services Act Transparency Database,” *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2024) <<https://dl.acm.org/doi/10.1145/3630106.3658960>> accessed 27 September 2024; Daria Dergacheva and others, “One Day in Content Moderation: Analyzing 24 h of Social Media Platforms’ Content Decisions through the DSA Transparency Database” (Center for Media, Communication and Information Research 2023) <<https://platform-governance.org/2023/one-day-in-content-moderation-by-social-media-platforms-in-the-eu/>> accessed 27 September 2024.

²⁷ See <<https://x.com/en/tos/previous>>.

²⁸ See <<https://legal.x.com/en/purchaser-terms/previous.html>>.

²⁹ See <<https://legal.x.com/en/purchaser-terms.html>>.

Since one of the main claims in the proceedings relates to breach of contract, determining the scope and content of contractual obligations between Twitter and the applicant is of paramount importance. Since it is not disputed by Twitter that restrictions were placed on Applicant's account,³⁰ the court makes an analysis of the legality of these restrictions. This assessment of unfairness is a historical win for consumer protection, as it targets platform power in the form of unchecked discretion over the provision of its services. Instruments such as the UCTD are based on general clauses (e.g. fairness), which have long been criticised in European private law,³¹ but which are increasingly showing their relevance in the context of protecting consumers in the digital economy. Most notably, the UCTD is important because the assessment of unfairness can be made by judges *ex officio*.³² In this case, the unfairness of the terms was not raised by the applicant, who focussed instead on data protection and DSA violations. This raises the issue of the interplay between competing consumer protection instruments, some of which may only be familiar to few consumers and legal counsels. The court's exploration of unfairness *ex officio* therefore shows how the UCTD can help address this lack of familiarity by giving judges additional opportunities to protect consumers. From this perspective, it can be argued that the UCTD, famously called a "sleeping beauty" by Micklitz and Reich,³³ can witness a revival. As the terms listed in the UCTD Annex still require a judicial assessment, this judgment can represent a precedent for addressing the unfairness of unilateral action in contracts concluded between consumers and social media platforms. The scope of judicial assessment raises further questions. Looking at the applicable "Paid Service" terms, which apply simultaneously with other standard terms, we can see a myriad of additional potentially unfair terms, such as liability exemption clauses, which have been the very reason why the UCTD came into existence.³⁴ The judgment of the Amsterdam District Court seems to imply that the *ex officio* review has been carried out only on the terms invoked by the applicant. One could ask: why stop there instead of evaluating all the applicable terms? Unfortunately, this would be almost impracticable in light of the intricate web of applicable ToSs.

Equally relevant for unilateral changes to consumer contracts for the provision of digital content or services is the Digital Content Directive (DCD), an instrument which, in spite of its broad applicability to services such as social media subscriptions, remains entirely obscure. According to Article 11 DCD, corroborated by Articles 5, 7 and 8 DCD, it can be argued that failing to provide consumers with digital content that is contractually fit for purpose (e.g. amplified as promised under the ToS), triggers the liability of the service provider. In the case in comment, it is well known that a paid X subscription increases content visibility as compared to standard accounts.³⁵ The applicant's account being shadowbanned, was no longer suitable to fulfil its purpose. In addition, Article 19 DCD indicates that if any modifications occur to the terms of service, the consumer must be given a valid reason for these modifications, the modifications must not cost the

³⁰ Para 6.

³¹ Lucinda Miller, "After the unfair Contract Terms Directive; Recent European Directives and English Law" (2007) 3(1) *European Review of Contract Law* 88.

³² Charlotte Pavillon, "Ignorance is Bliss – How Ex Officio Control Became the Raison d'Être of the UCTD" (2024) 32(3) *European Review of Private Law* 519.

³³ Hans-W. Micklitz & Norbert Reich, "The Court and Sleeping Beauty: The Revival of the Unfair Contract Terms Directive (UCTD)" (2014) 51(3) *Common Market Law Review* 771–808.

³⁴ Clause 3 under General Terms reads: "TO THE FULLEST EXTENT ALLOWED UNDER APPLICABLE LAW, THE X ENTITIES' MAXIMUM AGGREGATE LIABILITY FOR ANY NON-EXCLUDABLE WARRANTIES IS LIMITED TO ONE HUNDRED US DOLLARS (US\$100.00)."

³⁵ Lance Whitney, "X Premium Explained: What You Get and How to Use a Paid Twitter Account" (PC Mag, 22 March 2024) <https://www.pcmag.com/explainers/what-is-x-premium-plus-subscription-how-much> accessed 27 September 2024.

consumer any extra money, and, most importantly, the consumer must be notified of the modification. Although the DCD has received its own share of criticism,³⁶ and it is not yet entirely clear what remedies it leads to other than termination – which may not always be relevant or desirable – it is still an instrument worth testing in the context of social media platforms.

This judgment shows the manifold legal issues that arise when the DSA is applied in practice, particularly showing its potential complementarity and overlap with the consumer *acquis* and data protection legislation.

VIII. Reference

Applicant v Twitter International Unlimited Company, Amsterdam District Court (The Netherlands), Judgment of 5 July 2024, ECLI:NL:RBAMS:2024:3980

³⁶ Katarzyna Wiśniewska & Przemysław Pałka, “The impact of the Digital Content Directive on online platforms’ Terms of Service” (2023) 42 Yearbook of European Law 388.