Edited by
João Pedro Quintais

From the DMCA to the DSA

A Transatlantic Dialogue on Online Platform Regulation and Copyright

Verfassungsbooks
ON MATTERS CONSTITUTIONAL

DOI: 10.17176/20240429-081042-0
ISBN Print: 978-3-759825-95-7

Verfassungsbooks

Verfassungsblog gGmbH
Elbestraße 28
12045 Berlin
verfassungsblog.de
info@verfassungsblog.de

Cover design by Carl Brandt

This symposium is possible due to the support of the Institute for Information Law (IViR) at the University of Amsterdam. João Pedro Quintais' work on this symposium received funding from the Dutch Research Council (NWO) under the VENI grant scheme (project "Responsible Algorithms: How to Safeguard Freedom of Expression Online", grant number: VI.Veni.201R.036).

Edited by
João Pedro Quintais

From the DMCA to the DSA

A Transatlantic Dialogue on Online Platform Regulation and Copyright

Verfassungsbooks
ON MATTERS CONSTITUTIONAL

Contributing Authors

Niva Elkin-Koren

Niva Elkin-Koren is a Professor at Tel-Aviv University Faculty of Law and a Faculty Associate at the Berkman Klein Center for Internet and Society at Harvard University. She is the Director of the Chief Justice Meir Shamgar Center for Digital Law and Innovation at Tel Aviv University.

Giancarlo Frosio

Giancarlo Frosio is Professor of Intellectual Property and Technology Law and Director of the Global Intellectual Property and Technology (G-IPTech) Centre at the School of Law of Queen's University Belfast.

Christophe Geiger

Christophe Geiger is Professor of Law at the Luiss Guido Carli University in Rome (Italy) and President of the International Association for the Advancement of Teaching and Research in Intellectual Property (ATRIP).

Eric Goldman

Eric Goldman is Associate Dean for Research and Co-Director of the High Tech Law Institute at Santa Clara University School of Law, located in California's Silicon Valley.

Rachel Griffin

Rachel Griffin is a PhD candidate and lecturer in law at Sciences Po Paris.

Natali Helberger

Natali Helberger is Distinguished University Professor of Law and Digital Technology, with a special focus on AI at the University of Amsterdam, and affiliated with the Institute for Information Law (IViR).

Martin Husovec

Martin Husovec is an Associate Professor of Law at The London School of Economics and Political Science (LSE).

João Pedro Quintais

João Pedro Quintais is an Associate Professor at the Institute for Information Law (IViR) at the University of Amsterdam.

Eleonora Rosati

Eleonora Rosati is Professor of Intellectual Property Law at Stockholm University and Of Counsel at Bird&Bird.

Pamela Samuelson

Pamela Samuelson is the Richard M. Sherman Distinguished Professor of Law and Information at the University of California, Berkeley. She is also Co-Director of the Berkeley Center for Law & Technology.

Sebastian Schwemer

Sebastian Schwemer is Associate Professor and Director of the Centre for Information and Innovation Law at the University of Copenhagen. He advised the European Commission on the DSA proposal.

Martin Senftleben

Martin Senftleben is Professor of Intellectual Property and the Director of the Institute for Information Law (IViR) at the University of Amsterdam. He works as an Of Counsel at Bird & Bird in The Hague.

Erik Stallman

Erik Stallman is an Assistant Clinical Professor of Law and the Associate Director of the Samuelson Law, Technology & Public Policy Clinic at Berkeley Law.

Rebecca Tushnet

Rebecca Tushnet serves as the Frank Stanton Professor of First Amendment Law at Harvard Law School.

Jennifer Urban

Jennifer M. Urban is a Clinical Professor of Law at the University of California, Berkeley, where she is Director of Policy Initiatives for the Samuelson Law, Technology & Public Policy Clinic and Co-Director of the Berkeley Center for Law and Technology.

Content

João Pedro Quintais

From the DMCA to the DSA

A Transatlantic Dialogue on Online Platform Regulation and Copyright

For most of the early 21st century, EU law on online intermediaries was sparse, with no comprehensive harmonization of intermediary liability. The centerpiece of the legal framework was the 2000 e-Commerce Directive[1], which contained mere conditional liability exemptions, or "safe harbors", for certain types of intermediary services involving claims for damages (mere conduit or access, caching, and hosting), as well as a prohibition on the imposition by Member States on intermediary service providers of general monitoring obligations (Arts.12-15 e-Commerce Directive). Under this regime, intermediaries may still be required to take measures against the infringement of third party rights, because it remains possible to subject intermediaries to injunctions in regards to intellectual property rights, and duties of care (van Hoboken et al. 2018[2], Wilman 2021[3]). The interpretation of this constellation of provisions is complex and far from settled (see e.g. Angelopoulos 2020[4]). It is sufficient here to state that the development of case law from the Court of Justice of the EU (CJEU) in interpreting specific subject matter rules to extend the reach of harmonized EU law to online intermediaries, like in the context of intellectual property, led to increasing push towards additional regulation of online platforms.

This push has been justified around a somewhat blurry concept of legal, societal, political and even moral "responsibility" of online platforms (Helberger, Pierson and Poell 2018[5]; Taddeo and Floridi 2017[6]). The potential result, could "represent a substantial shift in intermediary liability theory", signaling a "move away from a well-established utilitarian approach toward a moral approach by rejecting negligence based intermediary liability arrangements", practically leading to a "broader move towards private enforcement online" (Frosio and Husovec 2020[7]).

In Europe, this state of affairs has led to a deluge of new "platform regulation" legislation in the past years, featuring the adoption of rules on terrorist content online, video-sharing platforms, copyright content-sharing platforms and – in what is the centerpiece of this push – horizontal rules for all online intermediaries in the Digital Services Act[8] (DSA) (Buiten 2021[9], Farrand 2019[10]). The DSA – which came into force on 17 February 2024 – takes a novel regulatory approach to intermediaries by imposing not only liability rules for the (user) content they host and moderate, but also separate due diligence obligations for the provider's own role and conduct in the design and functioning of their services (Husovec and Laguna 2022[11], Wilman 2022[12]; Hoboken, Quintais, Appelman et al. 2023[13]). The main target of these obligations are Big Tech companies, namely very large online platforms and search engines. They are subject to the largest set of obligations, including on due process and risk assessment and mitigation. These obligations extend to algorithmic moderation systems and the effect of their services on users' fundamental rights. This legislative push has also featured non-binding instruments like Codes of Conduct, Memoranda of Understanding and Recommendations on hate speech online[14], counterfeited goods[15], disinformation[16], and piracy of live events[17] (Quintais, Appelman, Ó Fathaigh 2023[18])

The US platform regulation story is different. It is undeniable that most of the largest and most successful internet intermediaries – at least in the Western world – originate from the U.S. Authors like Kossef causally link this fact to the U.S. legal landscape (Kossef 2022[19]), in particular to Section 230 of the Communications Decency Act (CDA) – passed in 1996 – which immunizes online platforms for liability arising from significant amounts of user-generated content.

Importantly, Section 230 contains a number of exceptions, such as for the enforcement of federal criminal law, copyright law, and electronic communications privacy laws. The copyright law exception is found in the 1998 Section 512 of the US Digital Millennium Copyright Act[20] (DMCA). Section 512 sets out a notice-and-take-down system that caching, hosting and linking platforms must comply with in order to qualify for the safe harbors. This regime directly influenced the design of the intermediaries' "safe-harbors" in the e-Commerce Directive and, as Sag notes, has also influenced the shape of online copyright enforcement online, leading to the implementation of "DMCA-plus" private agreements between rightsholders and large commercial platforms "in the shadow of those safe harbors" (Sag 2018[21]). These have ultimately resulted in automated copyright content moderation systems, including sophisticated filtering tools like YouTube's Content ID[22] or Meta's Rights Manager[23].

A similar sector specific path was followed in Europe based on the combined application and CJEU interpretation of direct liability rules for communication to the public of copyrighted works and the "safe-harbors" in the e-Commerce Directive, and national (non-harmonized) rules on secondary liability under national law. The latest development in this legislative story has seen the EU adopting a highly complex special regime for "online content-sharing service providers" (OCSSPs) in Art. 17 of the Copyright in the Digital Single Market Directive[24] (CDSMD). This provision applies to OCSSPs that host and provide public access to copyrighted content. This regime is unique in that it imposes direct liability on OCSSPs, sets aside the application of the hosting safe-harbor, and imposes its own special liability exemption mechanism, featuring best efforts obligations to obtain licenses, and implement measures for

notice and takedown, notice and stay down, and preventive filtering (see Quintais et al., 2022[25], 2024[26]; and COM/2021/288 final[27]).

Also, in the U.S. there is significant pressure to reform these legal regimes. For the moment, efforts to implement a solution similar to Art. 17 CDSMD have largely failed (Samuelson 2021[28]), in part due to skepticism surrounding its adoption (e.g. Bridy 2019[29]) and its roll out in Europe, which has already included a challenge on its validity on fundamental rights grounds (Case C-401/19 – *Poland v Parliament and Council*[30]; Quintais 2022[31], Husovec 2023[32]). Section 230 CDA, on the other hand, has faced much more persistent frontal attacks – including in ongoing US Supreme Court litigation (see e.g. Funk et al 2023[33], Rozenshtein 2023) and calls for reform with bipartisan support, even if on different grounds (see e.g. Anand et al, 2021[34], Jurecic 2022[35], Perault 2023[36]).

Against this background, a group of European and American scholars convened in 2023 to discuss the potential benefits and risks of the EU's new approach in its transatlantic context. They debated the DSA's potential to lead to a new EU/U.S. consensus or even EU influence on US platform regulation and liability debates (see Urban 2023[37]). The first meeting in the US led to the publication of a special issue[38] on the topic in the Berkeley Technology Law Journal. The second workshop in Amsterdam gave rise to this blog symposium.

The contributions to this symposium come from leading academics in the EU and U.S., often in collaboration with each other. They can be divided into two larger themes. A first set of contributions considers transversal issues of platform regulation in the EU and U.S., namely those of consistency (Rebecca Tushnet), due process (Eric Goldman and Sebastian Felix Schwemer), fundamental rights (Christophe Geiger and Giancarlo Frosio; Martin

Senftleben) and the potential "Brussels Effect" of the DSA (Martin Husovec and Jennifer Urban). A second set of contributions zooms in on key regimes, critically assessing rules on trusted flaggers (Eleonora Rosati), human in the loop (Rachel Griffin and Erik Stalman), access to data for researchers (Niva Elkin-Koren), and transparency (Pamela Samuelson and Natali Helberger).

João Pedro Quintais

References

1. 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), 178, http://data.europa.eu/eli/dir/2000/31/oj/eng (last visited Mar 24, 2024).
2. Content and Technology (European Commission) Directorate-General for Communications Networks et al., *Hosting Intermediary Services and Illegal Content Online: An Analysis of the Scope of Article 14 ECD in Light of Developments in the Online Service Landscape: Final Report*, (2019), https://data.europa.eu/doi/10.2759/284542 (last visited Apr 12, 2024).
3. Folkert Wilman, *The Eu's System of Knowledge-Based Liability for Hosting Service Providers in Respect of Illegal User Content – between the E-Commerce Directive and the Digital Services Act*, 12 Journal of Intellectual Property, Information Technology and E-Commerce Law (2021), https://www.jipitec.eu/issues/jipitec-12-3-2021/5343 (last visited May 10, 2024).
4. Christina Angelopoulos, *Harmonising Intermediary Copyright Liability in the EU: A Summary*, (2019), https://papers.ssrn.com/abstract=3685863 (last visited Mar 24, 2024).
5. Natali Helberger, Jo Pierson & Thomas Poell, *Governing Online Platforms: From Contested to Cooperative Responsibility*, The Information Society, https://www.tandfonline.com/doi/full/10.1080/01972243.2017.1391913 (last visited Apr 12, 2024).
6. The Responsibilities of Online Service Providers, (Mariarosaria Taddeo & Luciano Floridi eds., 2017), https://link.springer.com/book/10.1007/978-3-319-47852-4 (last visited Apr 16, 2024).
7. Giancarlo Frosio & Martin Husovec, *Accountability and Responsibility of Online Intermediaries*, *in* The Oxford Handbook of Online Intermediary Liability (Giancarlo Frosio ed., 2020), https://papers.ssrn.com/abstract=3451220 (last visited Apr 12, 2024).
8. 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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).
9. Miriam C. Buiten, *The Digital Services Act: From Intermediary Liability to Platform Regulation*, 12 Journal of Intellectual Property, Information Technology and E-Commerce Law (2022), https://www.jipitec.eu/issues/jipitec-12-5-2021/5491 (last visited May 10, 2024).

10. Benjamin Farrand, *The Ordoliberal Internet? Continuity and Change in the Eu's Approach to the Governance of Cyberspace*, 2 European Law Open (2023), https://www.cambridge.org/core/journals/european-law-open/article/ordoliberal-internet-continuity-and-change-in-the-eus-approach-to-the-governance-of-cyberspace/D79BB48E33BA37EA48457301C9910CCB (last visited Apr 12, 2024).

11. Martin Husovec & Irene Roche Laguna, *Digital Services Act: A Short Primer*, (2022), https://papers.ssrn.com/abstract=4153796 (last visited Apr 12, 2024).

12. Folkert Wilman, *The Digital Services Act (DSA) – an Overview*, (2022), https://papers.ssrn.com/abstract=4304586 (last visited Apr 12, 2024).

13. Putting the Digital Services Act Into Practice: Enforcement, Access to Justice, and Global Implications, (Joris van Hoboken et al. eds., 2023), https://papers.ssrn.com/abstract=4384266 (last visited Apr 12, 2024).

14. 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 (last visited Mar 24, 2024).

15. 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 (last visited Mar 24, 2024).

16. European Commission, 2022 Strengthened Code of Practice on Disinformation, (2022), https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation (last visited Mar 24, 2024).

17. Recommendation on Combating Online Piracy of Sports and Other Live Events, Shaping Europe's Digital Future, (2023), https://digital-strategy.ec.europa.eu/en/library/recommendation-combating-online-piracy-sports-and-other-live-events (last visited Mar 24, 2024).

18. João Pedro Quintais, Naomi Appelman & Ronan Ó Fathaigh, *Using Terms and Conditions to Apply Fundamental Rights to Content Moderation*, 24 German Law Journal (2023), https://www.cambridge.org/core/journals/german-law-journal/article/using-terms-and-conditions-to-apply-fundamental-rights-to-content-moderation/B30B9043D1C6F14AE9C3647A845E6E10 (last visited Apr 12, 2024).

19. Jeff Kosseff, *A User's Guide to Section 230, And a Legislator's Guide to Amending It (Or Not)*, 37 Berkeley Technology Law Journal (2022), https://doi.org/10.15779/Z38VT1GQ97 (last visited Apr 12, 2024).

20. 17 U.S. Code § 512 - Limitations on liability relating to material online, https://www.law.cornell.edu/uscode/text/17/512 (last visited Apr 24, 2024).

21. Matthew Sag, *Internet Safe Harbors and the Transformation of Copyright Law*, 93 Notre Dame Law Review, https://scholarship.law.nd.edu/ndlr/vol93/iss2/2 (last visited May 25, 2024).

22. YouTube Help, How Content ID works, https://support.google.com/youtube/answer/2797370?hl=en&ref_topic=9282364&sjid=17105169569832105410-EU (last visited Mar 24, 2024).

23. Meta Rights Manager, https://rightsmanager.fb.com/ (last visited Mar 24, 2024).

24. 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, https://eur-lex.europa.eu/eli/dir/2019/790/oj (last visited Mar 24, 2024).

25. João Pedro Quintais et al., *Copyright Content Moderation in the EU: An Interdisciplinary Mapping Analysis*, (2022), https://papers.ssrn.com/abstract=4210278 (last visited Mar 24, 2024).

26. João Pedro Quintais et al., *Copyright Content Moderation in the European Union: State of the Art, Ways Forward and Policy Recommendations*, 55 International Review of Intellectual Property and Competition Law, https://doi.org/10.1007/s40319-023-01409-5 (last visited Mar 24, 2024).

27. Communication from the Commission to the European Parliament and the Council, *Guidance on Article 17 of Directive 2019/790 on Copyright in the Digital Single Market*, COM/2021/288 final, (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021DC0288 (last visited Mar 24, 2024).

28. Pamela Samuelson, *Pushing Back on Stricter Copyright ISP Liability Rules*, 27 Michigan Technology Law Review (2021), https://repository.law.umich.edu/mtlr/vol27/iss2/4 (last visited May 10, 2024).

29. Annemarie Bridy, *The Price of Closing the Value Gap: How the Music Industry Hacked EU Copyright Reform*, 22 Vanderbilt Journal of Entertainment & Technology Law (2020), https://scholarship.law.vanderbilt.edu/jetlaw/vol22/iss2/4/ (last visited May 10, 2024).

30. Judgment of 26 April 2022, C-401/19, *Poland v Parliament and Council*, ECLI:EU:C:2022:297.

31. João Pedro Quintais, *Between Filters and Fundamental Rights: How the Court of Justice Saved Article 17 in C-401/19 – Poland V. Parliament and Council*, Verfassungsblog (2022), https://verfassungsblog.de/filters-poland/ (last visited Mar 24, 2024).

32. Martin Husovec, *Mandatory Filtering Does Not Always Violate Freedom of Expression: Important Lessons from Poland V. Council and European Parliament*, 60 Common Market Law Review (2023), https://doi.org/10.54648/cola2023007 (last visited Mar 24, 2024).

33. Allie Funk, *Q&A: Section 230 Is at the Supreme Court. Here's Why That Matters for Free Expression*, Freedom House (2023), https://freedomhouse.org/article/qa-section-230-supreme-court-heres-why-matters-free-expression (last visited Mar 24, 2024).

34. Meghan Anand et al., *All the Ways Congress Wants to Change Section 230*, Slate (2021), https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html (last visited Mar 24, 2024).

35. Quinta Jurecic, *The Politics of Section 230 Reform: Learning from Fosta's Mistakes*, Brookings (2022), https://www.brookings.edu/articles/the-politics-of-section-230-reform-learning-from-fostas-mistakes/ (last visited Mar 24, 2024).

36. Matt Perault, *Section 230 Won't Protect Chatgpt*, Lawfare (2023), https://www.lawfaremedia.org/article/section-230-wont-protect-chatgpt (last visited Mar 24, 2024).

37. Jennifer M. Urban, *Foreword*, 38 Berkeley Technology Law Journal (2023), https://btlj.org/wp-content/uploads/2024/01/0001_38-3_Urban.pdf (last visited May 10, 2024).

38. Berkeley Technology Law Journal, Volume 38, Issue 3, (2024), https://btlj.org/2024/01/volume-38-issue-3/ (last visited Mar 24, 2024).

Rebecca Tushnet

A Hobgoblin Comes for Internet Regulation

Recent laws in the US, along with the Digital Services Act[1] (DSA), seek to provide "due process" for individual content moderation decisions. Due process, understandably enough, often contains a component of treating like cases alike. It seems to follow, then, that if two relevantly similar users are treated differently, there is a problem of inconsistency, and that problem might be addressed by requiring more "due process" in the forms of appeals and clear rules and explanations of those rules to offenders. At least, the thinking goes, an appellate body can create coherent precedents and treat those who appeal consistently. And clearer rules are easier to apply; inconsistent applications should also be easier to detect than inconsistencies in the application of unclear rules.

But it is said that consistency is the hobgoblin of small minds.[2] In internet regulation, it is a damaging goal if taken as a mandate to make individual decisions uniformly consistent with each other. Evelyn Douek[3] has written about the need to focus on the overall system, not just the individual decisions that catch our attention, and Kate Klonick[4] has explained that this has always been part of serious thinking about content moderation. The DSA, more promisingly, suggests a focus on overall processes and does not treat errors as evidence of lawbreaking. By contrast, the Florida and Texas laws – currently enjoined pending Supreme Court review[5] – threaten platforms with large fines for each and every error.

Among many other things, Texas's HB20 prohibits large platforms from making editorial choices based on the "viewpoint" of the expression or user. Tex. Civ. Prac. & Rem. Code §§ 143A.001(1), 143A.002(a). It can be enforced either by the state or by individuals, and allows courts to impose "daily penalties sufficient to secure immediate compliance". § 143A.007. Similarly, Florida's S.B.7072

requires a "social media platform" to "apply censorship, deplatforming, and shadow banning standards in a consistent manner among its users on the platform". §501.2041(2)(b). The law does not define the phrase "consistent manner". On top of exposing violators to civil and administrative actions by the state attorney general, §501.2041(5), the law creates a private cause of action that allows individual users to sue to enforce the "consistency" mandate and authorizes awards of up to $100,000 in statutory damages for each claim, as well as actual damages, equitable relief, punitive damages, and in some cases attorneys' fees. §501.2041(6).

The problems of figuring out which content moderation cases are "relevantly similar" are well-known. Is breastfeeding "nudity"? What if it's posted with sexualizing prose? Should reporting on child abuse have extra leeway to describe what was done to a real victim? Should anti-Black speech be treated the same as anti-white speech? Is the term "Coke bottle" hate speech, given the uses that Brittan Heller[6] has explored? Is calling Bret Stephens a "bedbug"[7] the same as calling a group of people "bedbugs"?

Because of the fractal complexity of human communication, and its continuous evolution, no rule can both specify in advance what content is disallowed and also treat truly "like" cases – in terms of the harm they cause – alike. One goal must yield to the other.

But the problems of consistency are greater than that. Suppose we choose to prioritize having rules that can respond to new forms of identified abuses, and even new abuses if they appear. (The Texas and Florida laws suggest that the legislators, convinced that internet services discriminate against conservatives, would prefer rigidity instead, accepting new forms of abuse in order to prevent "censorship".)

Given the scale and variety of online communication on the largest services, it is impossible to expect more than the roughest consistency.

Appeals are likely to make the problems worse rather than better. Willingness to appeal content moderation decisions is not randomly distributed[8] (the Oversight Board writes about geographic origin but there is good evidence that other demographic factors also strongly affect willingness to make rights claims). Even if successful appeals lead to policy changes, that doesn't mean that previous removals will be revisited, or that the policy changes will be broad enough to treat analogous cases the same.

Other systems that operate at much smaller scales, but still with large numbers, have never been required to be consistent in this way. Consider teachers at state-run schools: They grade millions of students and even more student submissions. No one has ever suggested it is possible to constrain teachers so that an essay would receive exactly the same comments, and the same grades, from any teacher across a nation. Instead, rational school systems focus on processes for accrediting and evaluating teachers to make sure they are generally up to snuff. But two teachers can both be fine teachers even if they have very different views of what constitutes a good paper, and students' rights are not violated by this difference, as much as they may groan about it. Systems of federalism or localism mask some of this tolerance for inconsistency, as do doctrines of deference to decisionmakers on the ground. But it is not accidental that the most important critiques of these systems focus on their disparate impact by race, gender, disability, and other socially salient axes. Inconsistency and error alone are frustrating, but inevitable in human endeavors.

To take another example of a system that has to make hundreds of thousands of judgments on very different fact patterns every year, the US trademark registration system is unitary, and still gives itself cover for inconsistency by combining broad general principles and illustrative examples with a black-letter rule that each case is treated on its own merits. No applicant or opposer can succeed by showing that a similar trademark application was treated differently. Each application has its own unique context and evidentiary record. Since each application is reviewed by one of hundreds of trademark examiners, and there are hundreds of thousands of applications reviewed every year, there can be no other practice. Thus, when the Supreme Court invalidated bars on registering "disparaging", "scandalous", or "immoral" trademarks, it relied on the viewpoint-discriminatory nature of these bars. Some amici[9] highlighted the existence of inconsistencies – some applications including the term "MILF" were approved while others were rejected, and so on – and the Court alluded to this issue, but invalidating these bars because they could not be consistently applied would also endanger every other registration bar. It is equally impossible to be fully consistent about whether a term is descriptive as applied to the relevant goods or services, whether it is likely to cause confusion with another mark, and so on. Instead, we rely on general rules set forth in the Trademark Manual of Examining Procedure, which contains general rules and lots of examples, along with trained judgment – and we will never be totally satisfied with the results. As with trademarks, no map of content moderation can be as big as the territory. Of course there are and should be guideposts, but the fact that people disagree about applying those guideposts in particular situations doesn't mean that we've discovered an offense in need of remediation.

As Tarleton Gillespie has insightfully written,

> *Given the scale and the entire range of human communication, there is no such thing as a fully specified content policy: No guideline can be stable, clean, or incontrovertible; no way of saying it can preempt competing interpretations, by users and by the platform. Categorical terms like "sexually explicit" or "vulgar or obscene" do not close down contestation, they proliferate it: what counts as explicit? Vulgar to whom? All the caveats and clarifications in the world cannot make assessment any clearer; in truth, they merely multiply the blurry lines that must be anticipated now and adjudicated later. This is an exhausting and unwinnable game to play for those who moderate these platforms, as every rule immediately appears restrictive to some and lax to others, or appears either too finicky to follow or too blunt to do justice to the range of human aims to which questionable content is put.*
> (see Gillespie 2018[10] at 72-73)

Gillespie further explains that scale matters: "What to do with a questionable photo or a bad actor changes when you're facing not one violation but hundreds exactly like it, and thousands much like it, but slightly different in a thousand ways. This is not just a difference of size, it is fundamentally a different problem." Id. at 77. Social media posts have individualized contexts and records. As James Grimmelmann has noted[11], a post that decries eating Tide Pods and one that encourages eating Tide Pods can be indistinguishable to an outsider. As he says: "The difficulty of distinguishing between a practice, a parody of the practice, and a commentary on the practice is bad news for any legal doctrines that try to distinguish among them, and for any moderation guidelines or ethical prin-

ciples that try to draw similar distinctions." Of course, there are obvious rule violations, and situations where most people would have no trouble coming to a decision. But there are also constant pressures at the margins, and moderation itself contributes to those pressures as people try to get up as close to the line as they can without being banned, because borderline content gets more engagement[12]. The nearly harassing, the nearly inciting, the nearly nude all draw attention and encourage people to react. It is in this important area where there is no hope of true consistency, only of good training, diversity of moderators, and sampling for review.

We would better serve the human goals of due process by searching for patterns of disparate impact and looking for their causes. We should also, of course, aim to correct obvious errors. (These are often linked, as when automatic screening prohibits name-strings that correspond both to English slurs and real people's names, usually of non-English origin[13].) But the conversation should be about error rates and biases, not focused on examples that by their very nature must be unrepresentative. The DSA's due diligence obligations are a step in that direction, but even analysis of systemic risks and mitigation must be accompanied by an awareness that individual failures will be inevitable even in the best of all possible worlds. And the DSA's due process obligations for individual users point, like the Texas and Florida laws, in the other direction. A hobgoblin is haunting content moderation; we should face it directly.

References

1. 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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).
2. Emerson and Wilde on consistency, Wikipedia, https://en.wikipedia.org/w/index.php?title=Wikipedia:Emerson_and_Wilde_on_consistency&oldid=1202333184 (last visited Mar 24, 2024).
3. Evelyn Douek, *Content Moderation as Systems Thinking*, 136 Harvard Law Review (2022), https://harvardlawreview.org/print/vol-136/content-moderation-as-systems-thinking/ (last visited May 11, 2024).
4. Kate Klonick, *Of Systems Thinking and Straw Men*, 136 Harvard Law Review Forum, https://harvardlawreview.org/forum/vol-136/of-systems-thinking-and-straw-men/ (last visited May 11, 2024).
5. Amy Howe, *Justices Take Major Florida and Texas Social Media Cases*, SCOTUSblog (Oct. 30, 2023), https://www.scotusblog.com/2023/09/justices-take-major-florida-and-texas-social-media-cases/ (last visited Mar 24, 2024).
6. Brittan Heller, *Coca-Cola Curses: Hate Speech in a Post-Colonial Context*, 29 Michigan Technology Law Review (2023), https://doi.org/10.36645/mtlr.29.2.coca-cola (last visited May 11, 2024).
7. Aaron Rupar, *Bret Stephens's "Bedbug" Meltdown, Explained*, (2019), https://www.vox.com/2019/8/27/20834957/bret-stephens-bedbug-meltdown-dave-karpf-new-york-times-explained (last visited Mar 24, 2024).
8. Oversight Board 2022 Annual Report, https://oversightboard.com/attachment/795921088637952/ (last visited May 10, 2024).
9. United States Supreme Court, Brief of Professors Barton Beebe and Jeanne Fromer as Amici Curiae Supporting Respondent, in *Iancu v. Brunetti*, No. 18-302, (2019), https://www.supremecourt.gov/DocketPDF/18/18-302/93049/20190325143914702_Brunetti%20amicus%20for%20e-filing.pdf (last visited May 10, 2024).
10. Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, And the Hidden Decisions That Shape Social Media (2018), http://dx.doi.org/10.12987/9780300235029 (last visited May 11, 2024).
11. James Grimmelmann, *The Platform Is the Message*, 2 Georgetown Law Technology Review (2018), https://georgetownlawtechreview.org/the-platform-is-the-message/GLTR-07-2018/ (last visited May 11, 2024).

12. Josh Constine, *Facebook Will Change Algorithm to Demote "Borderline Content" That Almost Violates Policies*, TechCrunch (2018), https://techcrunch.com/2018/11/15/facebook-borderline-content/ (last visited May 11, 2024).

13. Patrick McKenzie, *Falsehoods Programmers Believe About Names*, Kalzumeus (2010), https://www.kalzumeus.com/2010/06/17/falsehoods-programmers-believe-about-names/ (last visited Mar 24, 2024).

Eric Goldman, Sebastian Schwemer

How the DMCA Anticipated the DSA's Due Process Obligations

In 1998, Congress enacted the Digital Millennium Copyright Act[1] (DMCA), a major copyright reform act intended to modernize copyright policy for the next millennium. Among other provisions, the DMCA established the well-known "notice-and-takedown" scheme that reduces the copyright liability exposure of user-generated content (UGC) services. The DMCA puts the burden on rightsowners to monitor online activities and affirmatively take action to stop perceived infringement.[2]

In Europe, the DMCA and its notice-and-takedown paradigm became the blueprint for the liability exemptions in the e-Commerce Directive[3] of 2000, the prevailing legal framework in the European Union (EU) for UGC services for two decades. Unlike the DMCA, the e-Commerce Directive applies to all types of illegal content, not just copyright-infringing items. More recently, the EU reformed the liability framework for user-caused copyright infringement in the 2019 Directive on copyright in the Digital Single Market[4]. In 2022, the EU followed that reform up with an even more comprehensive UGC liability reform in the Digital Services Act[5] (DSA).[6]

Among other things, the DSA requires UGC services to provide "due process"-like protections for user-authors. This regulatory approach is an important Internet Law development, but it's not completely novel. The DMCA also contains several due process-like protections for user-authors. This post identifies some of the DMCA's due process elements, compares them to the DSA's analogous provisions, and discusses the lessons from the DMCA for the DSA. Though the DSA uses a different policy paradigm than the DMCA, it's unclear if it will achieve better outcomes.

Comparison of the DMCA and DSA

The DMCA has several design features that anticipate the DSA's due process approaches. We discuss four such features: notice-and-appeal, disclosure of editorial policies, trusted flaggers, and user recourse for wrongful takedown demands.

Notice-and-appeal

The DMCA's notice-and-takedown scheme (17 U.S.C.§512(c)) provides rightsowners with a lot of leverage to remove allegedly infringing UGC. Knowing that rightsowners would sometimes make mistakes or even intentionally abuse their privilege, the DMCA contemplated that services would notify users when their items were targeted by rightsowners and provide an opportunity to correct mistakes. However, instead of expressly requiring services to provide notice-of-action or an appellate process, the DMCA provides services with a liability safe harbor if they honor users' "putback" notices (17 U.S.C. §512(g)). In other words, if users ask to restore the targeted content and provide the service with sufficient assurances, then the service could restore the content without incurring additional liability. Aggrieved rightsowners must then pursue the matter in court or drop it.

The DMCA's indirect approach to notice and appeals doesn't provide full due process protections to users. First, services don't have to notify users about the rightsowner's complaint or the service's action in response, though services may voluntarily choose to do so. Second, services do not have to inform users about the putback mechanism, so many users may be unaware of the possibility. Third, and most importantly, services do not have to honor putback requests. There are few legal downsides if the ser-

vice chooses to ignore it. Thus, the DMCA's putback safe harbor only vaguely resembles a proper notice-and-appeal process.

In Europe, the e-Commerce Directive remained silent on notice-and-appeals. In contrast, the DSA now provides both notice-of-action (Art. 16 DSA) and an appellate process (Art. 20 DSA). The DSA requires UGC services to provide a notice-of-action along with an explanation for the removal (Art. 17 DSA). In addition, the services must provide a complaint-resolution function that includes human review (Art. 20 DSA).

Disclosure of editorial policies

To qualify for the DMCA safe harbor, services must publicly announce their rules for recidivist (alleged) infringers (17 U.S.C. §512(i)(1)(A)). This provision nominally provides transparency about the governing rules to facilitate user compliance, but the statute doesn't specify any details about the required disclosures. Not surprisingly, many services make complex disclosures that users aren't likely to understand (see, e.g. Reid 2021[7]).

Though ruleset transparency is an essential part of due process, users probably aren't the main audience for the DMCA-compliant disclosures. Instead, the disclosures are more likely intended to help rightsowners monitor if services are appropriately disciplining recidivists. Given that audience focus, the ruleset disclosures don't provide the level of notice required for due process.

The DSA requires services to make much greater policy disclosures (e.g., Arts. 14 and 27 DSA). The disclosures apply to all policies about all types of illegitimate content, not just copyright infringement. Further, the disclosures must provide details about what facts and circumstances will influence the service's decision, including what considerations affect the service's decisions.

Trusted flaggers

The DMCA's scheme of removing the safe harbor upon notice creates substantial incentives for services to honor takedown notices, and remove content on the sender's demand, regardless of the request's legitimacy. To protect targeted users from improper removal demands, only notices from rightsowners or their designees implicate the service's safe harbor, and only when the sender declares that it is "authorized to act on behalf of the owner of an exclusive right that is allegedly infringed" (17 U.S.C. §512(c)(3)). Implicitly, this provision classifies copyright owners and their designees as "trusted flaggers" by giving their notices enhanced legal significance.

However, copyright owners do not always deserve that privileged status. For example, copyright ownership is often disputed. In those circumstances, the trusted flagger status gives one putative owner extra leverage over their ownership rivals online. Furthermore, rightsowners widely use automated means to detect alleged infringement, producing a flood of "robo-notices" with dubious margins of error inconsistent with their "trusted" status (see, e.g., Karaganis & Urban 2015[8]). Instead of protecting users, the DMCA's trusted flagger paradigm exacerbates content overremoval.

The DSA requires services to prioritize handling of notices submitted by trusted flaggers (Art. 22). Regulators can designate a limited number of flaggers who are accorded enhanced legal standing for their notices due to their expertise and competence. Services must notify the regulators if trusted flaggers are submitting too many erroneous reports, which may cause the entity to lose its trusted flagger status. Next to the trusted flagger status awarded by regulators, UGC platforms can voluntarily designate other trusted

flaggers (for detailed analysis of the trusted flagger regime, see Rosati's contribution to this symposium[9]).

User recourse for wrongful takedown demands

To provide users with recourse if the notice-and-takedown process is misused, the DMCA allows users to sue the senders of abusive takedown requests (17 U.S.C. §512(f)). However, 512(f) has helped only a trivial number of users. A 2004 Ninth Circuit Court of Appeals ruling[10] permits 512(f) claims only when takedown notice senders subjectively believed their notices were wrongful. However, users almost never have evidence of the sender's subjective beliefs when initiating the lawsuit. Thus, the DMCA's main mechanism to curb overreaching takedown notices doesn't properly function.

The DSA does not enable users to sue submitters of takedown notices. Instead, users may seek compensation from services for inadequately performing their content moderation duties. The DSA also requires services to suspend users who submit manifestly unfounded takedown notices (Art. 23 DSA).

DMCA's rightsowner focus v. DSA's user focus

As this post shows, the DMCA online safe harbors contain several features that nominally provide users with some due process protections. However, those features were incidental to the DMCA's primary goal of facilitating interactions between rightsowners and services.

In contrast to the DMCA, the DSA imposes affirmative obligations that services must comply with, rather than safe harbors that services can choose to opt-into. In that vein, the DSA repeatedly dictates how services must interact with users. Will the DSA's em-

phasis on user protections, compared to the DMCA's focus on rightsowners, lead to better outcomes?

Uncertainty #1: Traditionally, "due process" governs the actions of government actors, not private actors, due to structural differences between the two types of entities. Government actors have far greater powers over, and remedies against, their citizens than private companies have towards their "customers". Furthermore, government actors are sole-source providers of constituent services, and constituents must deal with them even if they don't want to. Thus, constituents need due process from government actors and constituents because of the government actor's powers and constituents' lack of choice.

It can be tempting to analogize large services like Google and Facebook to nation-states, but no private actor has the same monopoly or remedial powers over their "constituents" as any government entity. (Plus, the DSA doesn't limit its regulatory reach only to big services, such as very large online platforms and search engines.) Within the US regulatory context, this raises questions about the appropriateness of imposing government-like due process obligations on private actors.

Uncertainty #2: Government actors fund their procedural mechanisms using mandatory taxes; but when it's a mandatory cost to for-profit businesses, they will implement it as cheaply as possible (i.e., minimal viable compliance). The DSA's enforcers surely will question the sincerity of the regulated entities' implementations. What will the enforcers do about it?

The costs of providing due process may exceed the economic value of any individual user to the for-profit business, so services have incentives to disregard those users' interests – a common outcome under the DMCA (see, e.g. Keller 2021[11]). The DSA's due pro-

cess mandates may raise costs to the point where the services no longer can afford to support users at all. To the extent that the costs cause services to exit the industry or reduce their commitment to user-generated content, some users might lose authoring rights online due to the DSA's economic impact. Furthermore, increased regulatory costs usually reward incumbents over startups and reduce competitive dynamism (the DSA's micro- and small-enterprise exemption in Art. 19 doesn't eliminate all compliance costs), which suggests further shrinkage of online expression.

Uncertainty #3: Regulatory enforcement of the DSA's due process obligations may be hard to distinguish from censorship. Regulatory investigations and enforcements will send strong signals about what the government wants services to do, and those signals may not be bias-free. The DMCA didn't pose the same risks because it focuses on copyrights, whereas the DSA cuts across all content categories, including topics of substantial political and partisan interest. The DSA's obligations for very large online platforms to do risk assessments and risk mitigations (Arts. 33 to 35) provide a heightened potential for censorial interventions.

Conclusion

The DMCA shows how policymakers have been thinking about user due process since the earliest days of Internet regulation. However, solutions like the DMCA prioritized the interests of rightsowners over those of users. The DSA flips those priorities, substantively structuring the interactions between services and users. In theory, those revamped priorities might protect users better, but they also pose risks of unwanted regulatory-caused outcomes as highlighted in this post. As a result, by significantly extending the limited due

process principles attempted in the DMCA, the DSA raises important questions about the appropriateness and implications of imposing due process obligations on private entities.

References

1. H.R.2281 - Digital Millennium Copyright Act, https://www.copyright.gov/dmca/ (last visited Mar 24, 2024).
2. Nomenclature note: the DMCA applies to "online services". The DSA uses several different terms for regulated entities (including hosting, online platform, and very large online platform) and subjects them to heterogeneous obligations. For simplicity, this post uses the general descriptor "services" except where more specificity is required.
3. 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), 178, http://data.europa.eu/eli/dir/2000/31/oj/eng (last visited Mar 24, 2024).
4. 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, https://eur-lex.europa.eu/eli/dir/2019/790/oj (last visited Mar 24, 2024).
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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).
6. Many of the DSA's provisions discussed in this post trace their roots to the non-binding Commission Recommendation of 1.3.2018 on measures to effectively tackle illegal content online. For background on the complex relationship between the DSA and the Copyright Directive, see Quintais & Schwemer 2022.
7. Amanda Reid, Readability, Accessibility, And Clarity: An Analysis of DMCA Repeat Infringer Policies, 61, https://www.americanbar.org/digital-asset-abstract.html/content/dam/aba/publications/Jurimetrics/summer2021/reid.pdf, also available here: https://papers.ssrn.com/abstract=3921231 (last visited 24 Apr 2024).
8. Joe Karaganis & Jennifer Urban, *The Rise of the Robo Notice*, 58 Communications of the ACM (2015), https://dl.acm.org/doi/10.1145/2804244 (last visited Apr 23, 2024).
9. Eleonora Rosati, *The Dsa's Trusted Flaggers: Revolution, Evolution, Or Mere Gattopardismo?*, Verfassungsblog (2024), https://verfassungsblog.de/the-dsas-trusted-flaggers/ (last visited May 10, 2024).
10. *Rossi v. Motion Picture Association of America Inc.*, 391 F.3d 1000 (9th Cir. 2004), https://casetext.com/case/rossi-v-motion-picture-assn-of-america-inct (last visited Apr 24, 2024).

11. Daphne Keller, *Empirical Evidence of over-Removal by Internet Companies under Intermediary Liability Laws: An Updated List*, The Center for Internet and Society (2021), https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws (last visited Apr 23, 2024).

Giancarlo Frosio, Christophe Geiger

Towards a Digital Constitution

How the Digital Services Act Shapes the Future of Online Governance

Digital Service Providers (DSPs) like Google, Facebook, and Twitter/X have become key players in the modern digital landscape, influencing social interactions, political discourse, education and research, and cultural norms. Their widespread impact, however, brings challenges such as intellectual property (IP) infringement, privacy issues, hate and dangerous speech, misinformation, and political manipulation, highlighting the need for effective governance.

The European Union's Digital Services Act[1] (DSA) is a significant step in addressing these challenges, redefining digital platform regulations. It focuses on content moderation, user rights, and balancing regulation with innovation. The DSA aims to clarify platform responsibilities in content moderation, ensuring transparency and accountability, while protecting user rights and fostering digital market growth.

The DSA exemplifies the EU's efforts to create a fairer, more responsible digital environment. Through the DSA, the EU appears to be advancing a process of constitutionalisation[2] of Internet governance, as an important milestone in the evolving landscape of "digital constitutionalism"[3], aiming to establish a unified framework of rights, principles, and governance norms for the digital space, while also contributing to the development of new governance structures and regulatory bodies dedicated to effectively safeguarding fundamental rights online. This shift from reliance on private, market-driven solutions to a democratic, fundamental rights-centered approach[4] represents a major change in perspective. Importantly, this trend extends beyond the EU, gaining traction globally in various jurisdictions. Legislative initiatives like the UK's Online Safety Bill[5] and Brazil's "Fake News" Bill[6] also reflect a move towards public governance in moderating online con-

tent. Such a multi-faceted approach to digital constitutionalism[7] is increasingly seen as a practical response to the legitimacy crisis in privately managed online content moderation.

Digital service providers and fundamental rights: a balancing act

Large DSPs have transcended their roles as mere content hosts to become active shapers of public discourse and gatekeepers of information access. This transformation has significant implications for the democratic process and the exercise of fundamental rights, particularly in the realms of free speech and privacy. A poignant example of the complex role DSPs play in moderating public discourse is Twitter's decision[8] to suspend the account of a U.S. President. This action sparked a global debate[9] on the limits of free speech and the power of private companies over public communication channels. Similarly, Facebook's approach to content moderation[10], especially during politically charged events, has raised questions about the role of DSPs in influencing electoral processes and shaping political narratives. These incidents underscore the delicate balancing act DSPs must perform between allowing open discourse and curbing misinformation and harmful content.

The legal and ethical considerations of DSPs' content moderation policies are multifaceted. On the one hand, there is a legal imperative to adhere to national laws and regulations regarding illegal content. On the other, DSPs face ethical dilemmas when their policies intersect with issues of free expression and censorship. The European Union's General Data Protection Regulation[11] (GDPR) and the DSA are legislative attempts to provide a framework

for addressing these challenges, aiming to ensure that DSPs operate transparently and are held accountable for their content moderation decisions.

The impact of DSPs' content moderation policies on democratic processes and individual rights is profound. The role of DSPs in shaping public discourse and information access is a double-edged sword. While these platforms have the potential to enhance democratic engagement by providing a space for public discourse, their algorithms and moderation policies can also lead to the silencing of voices and the suppression of certain viewpoints. This has led to concerns about the "echo chamber" effect, where users are only exposed to information that reinforces their existing beliefs, and the potential for algorithmic bias, which can inadvertently marginalize certain groups.

Balancing these competing interests – ranging from freedom of expression to freedom to conduct a business, and from the right to an effective remedy to privacy and data protection[12] – is a complex challenge that requires careful consideration of both legal and ethical dimensions. The EU's framework for online fundamental rights forms a complex but pragmatic scaffold upon which to construct a comprehensive platform liability regime. It emphasizes the need to strike a balanced approach that respects the nuanced interplay among various fundamental rights. While the regulatory fabric laid out by the EU Charter[13] and the European Convention on Human Rights[14] allows for the imposition of obligations on DSPs, these must be carefully calibrated to protect the ecosystem of online platforms – from large, commercial entities to smaller, non-profit players. Importantly, IP rights, although recognized, are not to be overprotected to the point of overshadowing other fundamental rights or societal interests.[15]

Evolving liability and regulatory frameworks: from e-commerce to digital services

The legal frameworks governing DSPs have undergone significant evolution, mirroring the rapid development and growing influence of digital platforms in our society. The 2000 EU E-Commerce Directive[16] marked the beginning of formalized legal regulation for online services. It set the foundation for the digital market within the EU, primarily focusing on creating a harmonized environment for electronic commerce and introducing the concept of limited liability for service providers. This directive laid the groundwork for the regulation of digital services, although it was crafted in a different era of the internet, where the roles and impacts of DSPs were considerably different from today.

The regulatory landscape has since diversified, with regions like the EU, U.S., and others adopting varying approaches. In the EU, recent legislative developments, notably the DSA, represent a paradigm shift that that could potentially widen a transatlantic divide. The DSA aims to modernize the digital market's regulatory framework, addressing contemporary challenges like online harm and platform influence. This approach contrasts with the U.S., where Section 230 of the Communications Decency Act[17] still provides broad immunity to online platforms from liability for user-generated content, a principle that has been pivotal in the growth of these platforms but also a subject of intense debate and calls for reform.

Creating a global standard for digital governance remains a formidable challenge, given the divergent legal and cultural contexts across regions. The global internet landscape comprises various

stakeholders with differing priorities and values, making the harmonization of digital laws an intricate task. This diversity often leads to conflicts of jurisdiction and enforcement, exemplifying the complexities of regulating a borderless digital space.

The shift towards more stringent regulations reflects a growing recognition of the substantial impact DSPs have on public discourse, individual rights, and market competition. The DSA, for instance, introduces more robust obligations for platforms, such as transparency in content moderation, due diligence, and increased accountability. While these regulations aim to create a safer and more trustworthy digital environment, they also pose challenges for DSPs and users. For platforms, the increased responsibility and compliance requirements could impact operational models and innovation strategies. For users, while these changes promise enhanced protection and rights, they may also lead to increased content moderation and potential overreach.

Content moderation: the interplay of private ordering and state influence

Content moderation on digital platforms represents a complex and multifaceted challenge, intricately weaving together technology, law, and ethics. DSPs are at the forefront of this challenge, grappling with the monumental task of monitoring and moderating the vast amounts of content uploaded daily. The core of this moderation effort increasingly relies on sophisticated algorithms designed to detect and filter harmful and illegal content. However, these automated systems are not without their shortcomings. Issues of algorithmic bias and a lack of transparency have raised significant

concerns,[18] as they can inadvertently silence certain voices or amplify harmful narratives.

There is an intricate relationship between government policies and the content moderation practices of private platforms, introducing complexity to the landscape. Governments worldwide, each operating within their unique cultural and legal frameworks, influence platforms to adhere to local laws and societal norms. This influence ranges from explicit legal requirements, like those in the EU's DSA, to more subtle forms such as political and public opinion pressures, shaping content moderation policies. The interplay raises concerns about the independence of DSPs and the potential for state censorship under regulatory compliance. The challenge lies in striking a balance between safeguarding freedom of expression – a fundamental right in democratic societies – and preventing the dissemination of harmful content. In this context, platforms grapple with the ethical and technical complexities of fostering open discourse while minimizing the impact of harmful content like hate speech and misinformation on public safety and social harmony.

From a societal public interest-perspective, users' freedom of expression (and information) is crucial, given its role as part of "the essential foundations of [a democratic] society, one of the basic conditions for its progress and for the development of every man"[19]. Optimal regulation in the field of platform governance must thus attempt first to preserve users' and citizens' rights, as more online enforcement – and potential over-enforcement – equates with less access to information and less freedom of expression, thus a shrinking space for debate essential to democracy. The centrality of users' rights – and the overall goal of the EU legal system to preserve those rights against invasive proactive algorithmic

enforcement – has been reiterated by the Grand Chamber of the CJEU in the Case C-401/19[20] of 26 April 2022, possibly acknowledging a fundamental right of users to share[21] content online that cannot be limited by algorithmic content moderation.

The Digital Services Act: towards a fair and transparent digital market

At its core, the DSA seeks to modernize the regulatory framework for digital services, addressing the challenges and opportunities presented by the evolving digital landscape.

The DSA is built on a foundation of key provisions that aim to reshape the way digital services operate. One of its primary objectives is to enhance transparency, particularly in areas such as content moderation and advertising. By requiring platforms to disclose how they target and amplify content, the DSA promotes a more open digital environment. Furthermore, the act introduces stringent measures against illegal content online, mandating platforms to swiftly address such issues while providing clear reporting mechanisms for users.

A pivotal aspect of the DSA is its emphasis on accountability. The legislation imposes a due diligence obligation on DSPs, making them more responsible for the content they host and the services they provide. This shift signifies a move away from the *laissez-faire* approach that has predominantly governed the digital sphere, marking a new era where platforms are held to higher standards of responsibility.

The potential impact of the DSA on innovation, user rights, and platform responsibilities is profound. By establishing clearer rules, the DSA offers a stable legal environment that can foster innova-

tion and growth. For users, enhanced protections and greater transparency mean more control over their digital experiences and improved safeguarding of their rights. For platforms, and Very Large Online Platforms and Search Engines in particular, the DSA introduces new responsibilities and challenges, requiring them to adapt their operations to comply with stricter regulatory standards, including risk assessment and mitigation for algorithmic processes that might affect users' fundamental rights.

The DSA has the potential to serve as a model for global digital governance. Its comprehensive approach to digital regulation addresses many of the issues that have emerged in the digital age, setting a precedent for other countries and regions. By striking a balance between protecting user rights and fostering a healthy digital economy, the DSA could influence future legislation worldwide, promoting a more harmonized approach to digital governance. However, although it introduces innovative regulatory mechanisms for platform governance, it is also an exceptionally intricate and lengthy legislative document, where preference for national oversight strategies over unified European approaches risks further complicating its harmonised implementation.[22] Given these complexities, subsequent revisions and fine-tuning, also via delegated regulation, will inevitably be required to best protect fundamental rights in a rapidly evolving digital landscape.

Conclusion: charting a path towards digital constitutionalism

Navigating the digital age highlights the need for effective digital governance, as discussed in this blog post. DSPs play a key role in public discourse and are central to evolving regulatory frameworks, especially in content moderation. The DSA marks a major shift in

digital governance, focusing on transparency, accountability, and user protection, setting standards for DSPs to foster a safe, reliable, and innovative digital environment. Yet, digital governance is an evolving journey. The rapidly changing digital landscape presents continuous challenges and opportunities, requiring adaptable governance strategies.

The DSA, while providing a solid foundation, is the beginning of ongoing refinement and development. Looking ahead, the DSA's emphasis on fundamental rights, transparency, and regulatory oversight could guide transatlantic and global digital governance. The DSA, serving as a potential model for other nations crafting their digital strategies, leads us to distill 10 key principles that are rooted in its fundamental rights-centered approach.[23] These principles not only offer a blueprint for global digital governance but also serve as a valuable reference for other jurisdictions looking to update their legal frameworks for platform liability:

(1) DSP regulation in information societies is crucial for democratic information access and expression, requiring a balance of fundamental rights to uphold democracy and rule of law. Past DSP regulations have faced challenges in balancing competing fundamental rights.
(2) The DSA aims to balance interests while upholding rights, but its complexity and national oversight preference complicate implementation. Revisions are needed, and "digital constitutionalism" offers insights.
(3) The EU E-Commerce Directive and C-DSM Directive shaped DSP liability, with the DSA maintaining fundamental rights balance in a changing landscape.
(4) Fundamental Rights balancing in the DSA should be guided by

European human rights texts and case law, with international standards as reference. Challenges include:
ithms in order to safeguard creativity and expression, media pluralism and the right to information online[24].
(5) The DSA modernizes the e-Commerce Directive, emphasizing ex-post moderation over proactive measures and maintaining a ban on general monitoring obligations. Exceptions to this rule should be rare, primarily for manifestly illegal content that doesn't require independent assessment. Relying solely on automated filters for content moderation is ill-advised due to technological limitations. Adhering to the "human-in-command" principle is essential for accurate and nuanced content moderation.
(6) The DSA distinguishes between illegal and harmful content, focusing on harmonizing rules for illegal content. From a freedom of expression perspective, controversial content should not be censored simply because it may make the audience uncomfortable. Different regulatory approaches should be applied to illegal and manifestly illegal content.
a) Manifestly illegal content includes content promoting offenses against human dignity, war crimes, crimes against humanity, human trafficking, incitement to violence, acts of terrorism, and child abuse. It may also encompass content blatantly infringing on IP rights without the need for equity-based assessment. Such content should be clearly defined to avoid ambiguities.
b) For content that is illegal but not manifestly so, requiring human review for legality assessment is necessary. Further independent scrutiny should be available upon request, with consistent standards for expeditious removal within a reasonable timeframe.
c) When content is harmful but not outright illegal, complete removal may not be the best approach from a freedom of expression

standpoint. Alternative strategies like content flagging by DSPs and users, along with counter-speech mechanisms like "like" or "dislike" buttons, should be explored. Users should have more control over the type of content they engage with.

(7) Enhanced procedural guarantees for content moderation include:

a) Increased user access to information and opt-out options.

b) Efficient notice-and-action mechanisms with procedural safeguards, enabling the swift reinstatement of unjustly removed content.

c) Keeping notified content accessible during review, exempting DSPs from liability.

d) Transparency and human oversight in decision-making.

(8) The DSA regulates algorithmic content moderation, emphasizing fundamental rights and requiring:

a) Transparency and non-discrimination in algorithms.

b) Human review of algorithmic decisions.

c) Periodic audits and oversight for compliance from independent regulators.

d) Risk assessments and mitigation protocols.

e) Yearly transparency reports on algorithmic moderation.

However, there's room for refining the DSA's approach to algorithmic transparency and accountability to counter the challenges of algorithmic opacity. Specific obligations could be introduced to address issues like algorithmic bias, provide clearer explanations for automated decision-making logic, ensure transparency around data sets used for algorithmic training, and establish robust redress mechanisms to handle potential harm arising from algorithmic decisions.

(9) The DSA proposes specialized oversight bodies for monitoring DSP compliance. This oversight should be implemented according to the following guidelines:
a) A centralized EU entity for harmonized DSA implementation and policy guidelines, with a focus on fundamental rights, should operate in partnership with the EU Agency for Fundamental Rights (and, possibly, also other existing and future regulation authorities to be created to manage creativity online[25]).
b) This entity should serve quasi-judicial functions in content moderation, acting as a final dispute resolution authority for borderline cases and setting precedents for DSP moderation practices. However, this resolution option should not replace users' ability to seek recourse through an independent judiciary.
c) An Ombudsperson could represent users in these proceedings, ensuring their rights receive adequate protection.

(10) DSP obligations should be proportional and clear, avoiding impractical or ambiguous requirements that hinder business freedom and create barriers for Small and Medium-sized Enterprises. The DSA's nuanced approach to assigning responsibilities based on size and market share should be a benchmark for future content moderation regulations.

References

1. 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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).
2. Giancarlo Frosio, *Platform Responsibility in the Digital Services Act: Constitutionalising, Regulating and Governing Private Ordering*, *in* Research Handbook on EU Internet Law (Andrej Savin & Jan Trzaskowski eds., 2023), https://papers.ssrn.com/abstract=4236510 (last visited Apr 23, 2024).
3. Lex Gill, Dennis Redeker & Urs Gasser, *Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights*, Berkman Klein Center for Internet & Society Research Publication (2015), https://dash.harvard.edu/handle/1/28552582 (last visited Mar 24, 2024).
4. Giancarlo Frosio & Christophe Geiger, *Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime*, 29 European Law Journal (2023), https://onlinelibrary.wiley.com/doi/10.1111/eulj.12475 (last visited Apr 23, 2024).
5. Online Safety Act 2023, https://bills.parliament.uk/bills/3137 (last visited Apr 24, 2024.
6. Lei das Fake News, PL 2630/2020, https://www25.senado.leg.br/web/atividade/materias/-/materia/141944 (last visited Apr 24, 2024).
7. Giovanni De Gregorio & Oreste Pollicino, *The European Constitutional Road to Address Platform Power*, Verfassungsblog (2021), https://verfassungsblog.de/power-dsa-dma-03/ (last visited Mar 24, 2024).
8. Twitter 'permanently suspends' Trump's account, https://www.bbc.com/news/world-us-canada-55597840 (last visited Mar 24, 2024).
9. European Commission Press Corner, Speech by President von der Leyen at the European Parliament Plenary on the Inauguration of the New President of the United States and the Current political Situation, (2021), https://ec.europa.eu/commission/presscorner/detail/en/speech_21_167 (last visited Mar 24, 2024).
10. The Oversight Board Upholds Former President Trump's Suspension and Finds that Facebook Failed to Impose Proper Penalty, https://oversightboard.com/news/226612455899839-oversight-board-upholds-former-president-trump-s-suspension-finds-facebook-failed-to-impose-proper-penalty/ (last visited Mar 24, 2024).

11. 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) (Text with EEA relevance), (2016), http://data.europa.eu/eli/reg/2016/679/2016-05-04/eng (last visited Mar 24, 2024).

12. Christophe Geiger, Giancarlo Frosio & Elena Izyumenko, *Intermediary Liability and Fundamental Rights*, (2019), https://papers.ssrn.com/abstract=3411633 (last visited Mar 24, 2024).

13. 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 (last visited Mar 24, 2024).

14. Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR). https://www.echr.coe.int/documents/d/echr/Convention_ENG (last visited Apr 24, 2024).

15. Christophe Geiger, *Reconceptualizing the Constitutional Dimension of Intellectual Property – an Update*, (2019), https://papers.ssrn.com/abstract=3496779 (last visited Apr 23, 2024).

16. 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), 178, http://data.europa.eu/eli/dir/2000/31/oj/eng (last visited Mar 24, 2024).

17. 47 U.S. Code § 230 - Protection for Private Blocking and Screening of Offensive Material, https://www.law.cornell.edu/uscode/text/47/230 (last visited Mar 24, 2024).

18. Martin Senftleben, João Pedro Quintais & Arlette Meiring, *How the EU Outsources the Task of Human Rights Protection to Platforms and Users: The Case of UGC Monetization*, 38 Berkeley Technology Law Journal (2023), https://doi.org/10.15779/Z381G0HW20 (last visited May 10, 2024).

19. ECtHR, *Handyside v. The United Kingdom*, Application no. 5493/72, (1976), https://hudoc.echr.coe.int/ukr?i=001-57499 (last visited Mar 24, 2024).

20. Judgment of 26 April 2022, C-401/19, *Poland v Parliament and Council*, ECLI:EU:C:2022:297.

21. Giancarlo Frosio, *Freedom to Share*, 1145–1148 International Review of Intellectual Property and Competition Law (2022), https://doi.org/10.1007/s40319-022-01238-y (last visited Mar 24, 2024).

22. Giancarlo Frosio & Christophe Geiger, *Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime*, 29 European Law Journal (2023), https://onlinelibrary.wiley.com/doi/10.1111/eulj.12475 (last visited Apr 23, 2024).

23. Giancarlo Frosio & Christophe Geiger, *Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime*, 29 European Law Journal (2023), https://onlinelibrary.wiley.com/doi/10.1111/eulj.12475 (last visited Apr 23, 2024).

24. Christophe Geiger & Bernd Justin Jütte, *Designing Digital Constitutionalism: Copyright Exceptions and Limitations as a Regulatory Framework for Media Freedom and the Right to Information Online*, (2023), https://papers.ssrn.com/abstract=4548510 (last visited Mar 24, 2024).

25. Christophe Geiger & Natasha Mangal, *Regulating Creativity Online: Proposal for an EU Copyright Institution*, 71 GRUR International (2022), https://academic.oup.com/grurint/article/71/10/933/6656384 (last visited Apr 23, 2024).

Martin Senftleben

Human Rights Outsourcing and Reliance on User Activism in the DSA

Art. 14(4) of the Digital Services Act[1] (DSA) places an obligation on providers of intermediary services, including online platforms hosting user-generated content (see Art. 3(g) DSA), to apply content moderation systems in "a diligent, objective and proportionate manner". The provision emphasizes that online platforms are bound to carry out content filtering with due regard to the fundamental rights of users, such as freedom of expression. Considering the central role of online platforms in the current media landscape, this regulatory attempt to safeguard the right of users to share and receive information does not come as a surprise. However, fundamental rights, including freedom of expression (Art. 11(1) EU Charter of Fundamental Rights), have been designed as rights to be invoked against, and nurtured by, the state. Against this background, the approach taken in Art. 14(4) DSA raises complex questions. Does the possibility of imposing fundamental rights obligations on intermediaries, such as online platforms, exempt the state power from the noble task of preventing inroads into fundamental rights itself? Can the legislator legitimately outsource the obligation to safeguard fundamental rights to private parties (see the contribution by Geiger/Frosio[2] for a discussion of digital constitutionalism)?

In the case of user uploads to online content-sharing platforms, Art. 17(7) of the Directive on Copyright in the Digital Single Market[3] (CDSMD) adds an important guideline to the general obligation laid down in Art. 14(4) DSA (see Art. 2(4)(b) DSA as to the complementary application of these rules): the cooperation between online platforms and the creative industry in the area of content moderation (Art. 17(4) CDSMD) must not result in the blocking of non-infringing content uploads, including situations where user-generated content falls within the scope of a copyright

limitation that supports freedom of expression, such as the exemption of quotations, parodies and pastiches (explicitly mentioned in Art. 17(7) CDSMD, see also the more detailed discussion in Senftleben 2019[4]).

Joint effort of creative industry and platform providers

Evidently, this outsourcing scheme for human rights obligations relies on a joint effort of the creative industry and the online platform industry. To set the content filtering machinery in motion, copyright holders in the creative industry must notify "relevant and necessary information" with regard to those works which they want to ban from user uploads (Art. 17(4)(b) CDSMD). Once relevant and necessary information on protected works is received, the online platform is obliged to include that information in the content moderation process and ensure the unavailability of content uploads that contain traces of the protected works.

Unlike public authorities, however, the central players in this cooperation scheme are private entities that are not intrinsically motivated to safeguard the public interest in the exercise and furtherance of fundamental rights and freedoms. Despite all invocations of diligence and proportionality in Art. 14(4) DSA, the decision-making in the context of content filtering is most probably much more down to earth: the moment the balancing of competing human rights positions is confidently left to industry cooperation, economic cost and efficiency considerations are likely to occupy centre stage (see already the contribution by Goldman/Schwemer[5]).

A closer look at the different stages of industry cooperation resulting from the described regulatory model confirms that con-

cerns about human rights deficits are not unfounded. As explained, the first step in the content moderation process is the notification of relevant and necessary information relating to "specific works and other subject matter" by copyright holders (Art. 17(4)(b) CDSMD). In the light of case law precedents, in particular *Sabam/Netlog*[6] (para. 51), use of the word "specific" can be understood to reflect the legislator's hope that copyright holders will only notify individually selected works. Otherwise, content moderation may reach proportions that violates freedom of expression and information, and other fundamental rights (see Angelopoulos & Senftleben 2021[7]). In *Sabam/Netlog*, the Court declared content filtering based on a whole repertoire of collecting society repertoire excessive and impermissible (paras 48-51).

Seeking to avoid the evolution of an overbroad, general filtering obligation, a copyright holder could limit use of the notification system to those works that constitute cornerstones of the current exploitation strategy. As a result, other elements of the work catalogue would remain available for creative remix activities of users. This, in turn, would reduce the risk of overbroad inroads into freedom of expression and information.

In practice, however, rightholders are unlikely to adopt this cautious approach. The success of the risk reduction strategy surrounding the word "specific" is doubtful. In the cooperation with online platforms, nothing seems to prevent the creative industry from sending copyright notifications that cover each and every element of impressive work catalogues. Platforms for user-generated content may thus receive long lists of all "specific" works which copyright holders have in their repertoire. Adding up all works included in these notifications, the conclusion can become inescapable that the regulatory approach underlying the described

interplay of rules in the DSA and the CDSMD culminates in a filtering obligation that is very similar to the filtering measures which the CJEU prohibited in *Sabam/Netlog*. The risk of encroachments upon human rights is evident (see also Senftleben 2024[8]).

Impact of cost and efficiency considerations

Turning to the second step in the content moderation process – the act of filtering carried out by online platforms to prevent the availability of notified works – the aforementioned proportionality and diligence obligations apply: content moderation must comply with the diligence and proportionality requirements in Art. 14(4) DSA. As to the practical outcome of content filtering in the light of diligence and proportionality requirements, however, it is to be recalled that online platforms will most probably align the concrete implementation of content moderation systems with cost and efficiency considerations. In reality, the subordination of concrete industry decisions to abstract diligence and proportionality imperatives – the acceptance of more costs and less profits to reduce the corrosive effect on freedom of expression and information – would come as a surprise. Online platforms can be expected to be rational in the sense that they seek to achieve content filtering at minimal costs. A test of proportionality is unlikely to occupy centre stage unless the least intrusive measure also constitutes the least costly measure. A test of professional diligence is unlikely to lead to the adoption of a more costly and less intrusive content moderation system unless additional revenues accruing from enhanced popularity among users offsets the extra investment of money.

Hence, there is no guarantee that industry cooperation in the field of user-generated content will lead to the adoption of the

most sophisticated filtering systems with the highest potential to avoid unjustified removals of content mash-ups and remixes (further examined in Senftleben, Quintais & Meiring 2023[9]). An assessment of liability rules also confirms that excessive filtering risks must be taken seriously. An online platform seeking to minimize the risk of liability is likely to succumb to the temptation of overblocking. Filtering more than necessary is less risky than filtering only clear-cut cases of infringement. After all, primary, direct liability for infringing user uploads follows from Art. 17(1) CDSMD and dangles above the head of providers of platforms for user-generated content like the sword of Damocles. The conclusion is thus inescapable that the outsourcing strategy underlying the EU regulation of content moderation in the DSA and the CDSMD is highly problematic. Instead of safeguarding human rights, the regulatory approach is likely to culminate in human rights violations.

Reliance on user complaints

Against this background, it is of particular importance to analyse mechanisms that could bring human rights deficits to light and remedy shortcomings. This question requires the discussion of the role of users. Art. 14(1) DSA and Art. 17(9) CDSMD both make users the primary addressees of information about content moderation systems. According to Art. 14(1) DSA, users shall receive information on upload and content sharing restrictions arising from the employment of content moderation tools. If they want to take measures against content restrictions, Art. 17(9) CDSMD – and the complementary provisions in Art. 20 DSA – ensure that complaint and redress mechanisms are available to users of OCSSP services

"in the event of disputes over the disabling of access to, or the removal of, works or other subject matter uploaded by them".

Hence, users are expected to instigate complaint and redress procedures at platform level and, ultimately, go to court. The reliance placed on this mechanism, however, is surprising. Evidence from the application of the DMCA counter-notice system in the U.S. shows[10] quite clearly that users are unlikely to file complaints in the first place. This is confirmed by data from recent transparency reports from the largest user-generated content (UGC) platforms (examined by Senftleben, Quintais & Meiring 2023[11]). If users must wait relatively long for a final result, it is foreseeable that a complaint and redress mechanism that depends on user initiatives is incapable of safeguarding freedom of expression and information. Moreover, an overly cumbersome complaint and redress mechanism may thwart user initiatives from the outset.

In the context of user-generated content, it is often crucial to react quickly to current news and film, book and music releases. If the complaint and redress mechanism finally yields the insight that a lawful content remix or mash-up has been blocked, the decisive moment for the affected quotation or parody may already have passed. From this perspective, the elastic timeframe for complaint handling in Art. 17(9) CDSMD – "shall be processed without undue delay" – gives rise to concerns. This standard differs markedly from an obligation to let blocked content reappear promptly. As Art. 17(9) CDSMD also requires human review, it may take quite a while until a decision on the infringing nature of content is taken. Considering these features, the complaint and redress option may appear unattractive to users (see Senftleben 2020[12]).

Instead of dispelling concerns about human rights deficits, the reliance on user complaints, thus, constitutes a further risk factor.

Apart from being ineffective as a remedy for human rights violations, it may allow authorities to hide behind a lack of user activism and thereby conceal human rights deficits. It may also be that users refrain from complaining because they consider the mechanism too cumbersome and/or too slow. However, when taking the number of user complaints as a yardstick for assessing human rights risks, a relatively low number of user complaints may be misinterpreted as evidence that content moderation does not lead to excessive content blocking. If users refrain from taking action, human rights deficits stay under the radar. The oversimplified equation "no user complaint = no human rights problem" offers the opportunity of presenting potentially overly restrictive content moderation systems as a success. Instead of shedding light on human rights deficits, the complaint and redress mechanism can be used strategically – by platforms and regulators alike – to conceal encroachments upon freedom of expression and information.

Conclusion

In sum, closer inspection of DSA and CDSMD content moderation rules confirms a worrying tendency of reliance on industry cooperation and user activism to safeguard human rights. Instead of putting responsibility for detecting and remedying human rights deficits in the hands of the state, the EU legislature prefers to outsource this responsibility to private entities, such as online platforms, and conceal potential violations by leaving countermeasures to users. The risk of eroding freedom of expression is further enhanced by the fact that, instead of exposing and discussing the corrosive effect of human rights outsourcing, the CJEU has already rubberstamped the described regulatory approach. In its *Poland*[13]

decision (see Quintais 2022[14] and Husovec 2023[15]), the Court has even qualified problematic features of the outsourcing and concealment strategy as valid safeguards against the erosion of freedom of expression and information (see further Senftleben 2024[16]).

To safeguard human rights, the state power itself must become much more active. Litanies of due diligence and proportionality obligations for private entities and reliance on user activism are not enough. Requirements for audit reports under Art. 37 DSA should include the obligation to provide sufficiently detailed information on the implementation of human rights safeguards to allow the European Commission to exercise effective control and prevent encroachments (see Arts. 42(4), 66(1), 70(1), 73(1), 74(1) DSA). The implementation of Art. 17 CDSMD in national legislation should only be deemed satisfactory when the Member State has devised effective legal mechanisms to ensure that content filtering measures do not erode the freedom of users to upload quotations, parodies and pastiches (Art. 17(7) CDSMD). Moreover, the research community should be encouraged to throw light on violations of freedom of expression and information when analysing platform data (Art. 40(4) and (12), 34(1)(b) DSA).

References

1. 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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).
2. Giancarlo Frosio & Christophe Geiger, *Towards a Digital Constitution: How the Digital Services Act Shapes the Future of Online Governance*, Verfassungsblog (2024), https://verfassungsblog.de/towards-a-digital-constitution/ (last visited Mar 24, 2024).
3. 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, https://eur-lex.europa.eu/eli/dir/2019/790/oj (last visited Mar 24, 2024).
4. Martin Senftleben, *Bermuda Triangle – Licensing, Filtering and Privileging User-Generated Content under the New Directive on Copyright in the Digital Single Market*, (2019), https://papers.ssrn.com/abstract=3367219 (last visited Apr 23, 2024).
5. Eric Goldman & Sebastian Schwemer, *How the DMCA Anticipated the Dsa's Due Process Obligations*, Verfassungsblog (2024), https://verfassungsblog.de/how-the-dmca-anticipated-the-dsas-due-process-obligations/ (last visited Mar 24, 2024).
6. Judgment of 16 February 2012, *SABAM*, C-360/10, ECLI:EU:C:2012:85, https://curia.europa.eu/juris/liste.jsf?num=C-360/10.
7. Christina Angelopoulos & Martin Senftleben, *An Endless Odyssey? Content Moderation Without General Content Monitoring Obligations*, (2021), https://papers.ssrn.com/abstract=3871916 (last visited Apr 23, 2024).
8. Martin Senftleben, *Guardians of the UGC Galaxy – Human Rights Obligations of Online Platforms, Copyright Holders, Member States and the European Commission under the CDSM Directive and the Digital Services Act*, 14 Journal of Intellectual Property, Information Technology and E-Commerce Law (2024), https://www.jipitec.eu/archive/issues/jipitec-14-3-2023/5847 (last visited May 10, 2024).
9. Martin Senftleben, João Pedro Quintais & Arlette Meiring, *How the EU Outsources the Task of Human Rights Protection to Platforms and Users: The Case of UGC Monetization*, 38 Berkeley Technology Law Journal (2023), https://doi.org/10.15779/Z381G0HW20 (last visited May 10, 2024).
10. Jennifer M. Urban & Laura Quilter, *Efficient Process or Chilling Effects – Takedown Notices under Section 512 of the Digital Millennium Copyright Act*, 22 Santa Clara Computer and High Technology Law Journal (2006), https://digitalcommons.law.scu.edu/chtlj/vol22/iss4/1 (last visited Apr 23, 2024).

11. Martin Senftleben, João Pedro Quintais & Arlette Meiring, *How the EU Outsources the Task of Human Rights Protection to Platforms and Users: The Case of UGC Monetization*, 38 Berkeley Technology Law Journal (2023), https://doi.org/10.15779/Z381G0HW20 (last visited May 10, 2024).

12. Martin Senftleben, *Institutionalized Algorithmic Enforcement – the Pros and Cons of the EU Approach to UGC Platform Liability*, 14 FIU Law Review (2020), https://ecollections.law.fiu.edu/lawreview/vol14/iss2/11 (last visited May 25, 2024).

13. Judgment of 26 April 2022, C-401/19, *Poland v Parliament and Council*, ECLI:EU:C:2022:297.

14. João Pedro Quintais, *Between Filters and Fundamental Rights: How the Court of Justice Saved Article 17 in C-401/19 – Poland V. Parliament and Council*, Verfassungsblog (2022), https://verfassungsblog.de/filters-poland/ (last visited Mar 24, 2024).

15. Martin Husovec, *Mandatory Filtering Does Not Always Violate Freedom of Expression: Important Lessons from Poland V. Council and European Parliament*, 60 Common Market Law Review (2023), https://doi.org/10.54648/cola2023007 (last visited Mar 24, 2024).

16. Martin Senftleben, *Guardians of the UGC Galaxy – Human Rights Obligations of Online Platforms, Copyright Holders, Member States and the European Commission under the CDSM Directive and the Digital Services Act*, 14 Journal of Intellectual Property, Information Technology and E-Commerce Law (2024), https://www.jipitec.eu/archive/issues/jipitec-14-3-2023/5847 (last visited May 10, 2024).

Martin Husovec, Jennifer Urban

Will the DSA have the Brussels Effect?

The[1] Digital Services Act[2] (DSA) is a comprehensive effort by the European Union (EU) to regulate digital services. Many onlookers in Europe and beyond its borders wonder about whether the DSA will influence activities outside of Europe via a "Brussels Effect"[3]. In this contribution, we argue that when it comes to extraterritorial spill-over effects of the DSA that are driven by economic incentives or de facto standardisation and private ordering, the strength of any DSA Brussels Effect will depend on several factors: the type of obligations in question; compliance costs; the extent of regulatory imitation by other countries; and finally, the existence of any countervailing legal regimes. Under this analysis, the chances of spontaneous voluntary implementation beyond the EU's borders for four key parts of the DSA – content moderation procedures, transparency and governance obligations, and risk management rules – seem modest. Some content moderation rules might reach beyond the European continent through the ensuing industry standardisation.

Four key components of the DSA

The DSA regulates how online service providers make content moderation decisions. It subjects companies to an elaborate set of prescriptive rules that organize the process of notification, evaluation, removal, and contestation (Art. 16 to 21). Affected individuals are given a right to an individual explanation of such decisions (Art. 17), the right to appeal decisions internally for free (Art. 20), and the right to appeal externally before independent out-of-court dispute settlement bodies (Art. 21).

Providers are further obliged to issue annual or bi-annual transparency reports about how they conduct content moderation

(Arts. 15 and 24). Companies of a certain size and reach must also submit all their individual content moderation decisions to a centralised database (Article 24(5)). The largest online platforms or search engines (so-called Very Large Online Providers (VLOPs) and Very Large Online Search Engines (VLOSEs)) have further disclosure obligations (Article 42). VLOPs and VLOSEs must also provide access to data to researchers to study risks and mitigation strategies on their services (Article 40).

All providers have some obligations to appoint points of contact (Arts. 11 and 12), or legal representatives (if not established in the EU [Art. 13]); however, only the largest ones have extensive governance obligations. VLOPs and VLOSEs must appoint compliance officers who must have certain standing with the senior management of companies (Art. 41). VLOPs and VLOSEs also must appropriately train their staff, including content moderators, and monitor risk management within companies (Arts. 34, 35, 42(2)(b)).

Finally, mid-sized or bigger online platforms must design their services in compliance with certain statutory risk-related imperatives, such as avoiding misleading or manipulating practices (Art. 25), or design that fails to protect the safety, security, and privacy of children as their users (Art. 28). VLOPs and VLOSEs are subject to periodic risk assessment and external review by auditors under the supervision of the European Commission and national regulators.

Predicting extraterritorial spill-overs

Simply stated, the Brussels Effect causes EU rules to "spill over" into other jurisdictions through private actions of companies (a "de facto" effect) or harmonising changes in law by outside jurisdic-

tions (a "de jure" effect[4]). We focus on de facto spill-overs, which, if they occur, arise from the choices of individual firms.

Feasibility of localised implementation

Whether spill-overs occur first and foremost depends on what companies find useful and cost-effective. However, in some cases, usefulness and cost-effectiveness can be forced or incentivised by the global nature of the product. If a new EU obligation cannot be easily siloed into a specific location – whether for technical or other reasons – companies might prefer to extend their compliance across jurisdictions. For instance, if certain design features of the systems are harder to split and localise to certain jurisdictions only, companies might extend the implementation of the rules governing such features beyond the EU. The ability to localise implementation is thus one of the key issues.

Many of the DSA's rules probably can be localised. For example, companies seem to have localised all obligations regarding the opt-out from the recommender systems (see here[5], here[6], and here[7]), thus signalling that splitting markets is not too difficult in this instance. There may be areas where the services cannot be designed differently for different markets but that will probably be an exception, at least for large players that already comply with requirements from multiple jurisdictions.

Cost

In most cases, the main reason why companies might extend the application of the DSA beyond the EU is cost: because they find it cheaper to keep one set of rules for several markets. This might be the case for some notice-handling content moderation rules. Given that companies must build new processes for European users, the cost of extending some of those rules to other jurisdictions might

be lower than keeping two or more separate complaint-handling systems. However, some of the user-protecting obligations under the DSA, such as the broad possibility to appeal visibility restrictions of any kind, are both unique and quite costly, and thus less likely to be implemented in other countries without a legal mandate.

Indirect effects

Other indirect effects on the markets are possible too. The industry's reliance on a single set of rules could help to standardise processes, improve demand for and interoperability of moderation tools, and thus increase the new market entry in the area. For content moderation, the DSA can become a shipping container moment[8], which gives an entire industry vocabulary, structure and building blocks. Section 512 of the US Digital Millennium Copyright Act[9] (DMCA) has functioned in this way[10] for the last twenty-five years – companies applied many of its features well beyond US copyright disputes because it provided guidance that could be used to structure takedown practices across types of complaint and across jurisdictions. The DSA's far more detailed requirements could serve as the updated "shipping container" for companies looking to update their content moderation practices. Such dimensions can help spur a lot of innovation and business activity around industry-wide content moderation solutions that can be customised, repurposed, or applied on a cross-platform basis. A bigger market means better solutions.

Looking at specific examples, in theory, industry-wide standardisation around content moderation can also facilitate the convergence concerning voluntary and DSA-mandated transparency reporting. This might depend on how separate the US and EU back-end systems are and how convergent the DSA's requirements

are with companies' current practices. Transparency reporting is neither easy nor cheap, especially for companies running on legacy systems. The DSA could have a nudging effect that could move some providers to transparency reports, or more detailed transparency reports, in other countries, such as the US, but this might take time. And other incentives may push companies in the opposite direction. For example, being more transparent than others also has its costs associated with extra scrutiny by private actors[11] and public authorities.

Additional considerations

Further, some DSA obligations may be unattractive for voluntary compliance for other reasons: because they are too expensive, because they require local institutions that don't exist, or because they create bad policy precedents from a local perspective. For instance, out-of-court dispute settlements are both costly and cannot work without appropriate certified out-of-court dispute settlement bodies. Periodic risk assessment and auditing also is not cheap. Moreover, companies could rationally fear that legal change abroad would become politically easier if major industry players change their private policies to match the DSA requirements.

An interesting case in this regard is researchers' access to data to study platforms. Due to a lack of vetting processes in other countries, the companies would seem to be easily able to reject extending access to non-public data to researchers from outside the EU. However, these researchers, to the extent that they are interested in studying EU-relevant risks, can benefit from the EU regime. Arguably, they can use their respective countries as control groups[12], and thus undertake more globally-relevant research as well.

Moreover, the scraping/API provision of the DSA facilitates access to publicly available data[13], again without limitation to the nationality of those who are undertaking the research. Even if foreign governments do not follow the EU's approach, they might favour that their researchers can benefit from the EU regime and might complain if their researchers cannot access similar tools. Thus, public pressure in other countries might force companies to extend some of the features of the system beyond the EU, even though there usually is little immediate positive economic benefit from such transparency for the companies.

Conflicting rules

Finally, the DSA's extraterritorial effect might be enhanced by reinforcing local rules, but also undermined, or entirely prevented, by conflicting rules in other jurisdictions.

Consider how the U.S. DMCA interacts with the DSA. The DSA is mostly stricter in what it requires companies to do compared to the US law (although, exceptions do exist, such as the fact that the DSA's DMCA-style repeat-infringer rule does not apply to mere conduits). But there are potential conflicts that could prevent companies from replacing DMCA-required or -influenced practices with DSA-style practices.

For example, the DSA includes protections for targets of removal requests that could be significantly more effective than the DMCA's analogues. The DMCA says that online service providers "shall" reinstate content upon receipt of a counter-notice[14] but in practice, this often doesn't happen. The reason is the liability asymmetry: liability risk for not taking down infringing material is far greater than the liability risk for leaving up non-infringing material. Copyright remedies in the U.S. can be very severe, and the main notice senders have deep pockets, creating strong incentives

toward removal for providers. The affected target would have to rely on section 512(f) , which allows challenges only for "knowing, material misrepresentations"[15]. This is a very high barrier for recovery, thus making 512(f) effectively a dead letter.

The DSA in the EU aims to re-balance exactly that liability asymmetry by imposing countervailing due diligence obligations for the benefit of affected individuals. The question is whether they could be of any use outside the EU. For global actions that affect both markets simultaneously, the new EU liability could change the overall risk calculation. However, if companies can split their compliance by geo-locating it, in all likelihood, the DSA will not change the local US risk calculation, at least for copyright disputes. And most other disputes fall within the safe-harbor protections of section 230 of the Communications Decency Act[16] (CDA) – which, broadly speaking – allows companies to make the moderation decisions they prefer. In other words, to the extent that companies can separate their compliance, they will act in a more balanced way in the EU than in the U.S.

Furthermore, the DMCA and the CDA both lack disclosure requirements; given the incentives described above, this could undermine the willingness of providers to provide more information to affected individuals. Because the DMCA requires only limited information from notifiers as compared to the DSA – and outside of copyright, little information is required by US law at all – providers might be stuck with the existing disclosures, whatever their intentions or other motivations.

Finally, other jurisdictions could have laws that directly conflict with the requirements of the DSA. While CDA section 230 and the U.S. First Amendment jurisprudence might conflict in spirit with some of the DSA obligations, they do not prevent companies from

voluntarily extending the DSA rules to the U.S. market and individuals. The more tangible conflicts may ultimately be those where federal or state statutes prevent companies from doing what they are required to do in the EU, such as removing or disclosing something. In such cases, the only way how companies can comply with both regimes is to geo-localise their compliance.

Conclusions

Our brief analysis suggests that many most ambitious parts of the DSA will probably have modest impact on the other countries unless these countries adopt similar laws. From all the new obligations, the most promising is the potential impact of the DSA content moderation rules on the industry standards for processes and tools, and potentially the data access regime. However, the overall outcome depends on many variables, including those that might not be even well understood within the companies at the time when they are implementing the DSA. Changes that seem costly and without benefits today, might easily prove to be useful and less costly tomorrow. Thus, we shouldn't be too quick to judge the law's overall de facto Brussel's Effect. However, DSA compliance offers a moment for another important lesson. The compliance attitude of companies – that is, how ready they are to redesign their products for specific markets and engage in geo-localised enforcement to prevent giving the DSA an extraterritorial effect – will show us the true colours of how "universal" or "global" the digital services as products are in the mid-2020s.

References

1. Views expressed in this piece by Jennifer Urban are her own and should not be attributed to the University of California, the California Privacy Protection Agency, or the California Privacy Protection Agency Board.
2. 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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).
3. Anu Bradford, The Brussels Effect: How the European Union Rules the World (2019), https://doi.org/10.1093/oso/9780190088583.001.0001 (last visited May 10, 2024).
4. Anu Bradford, The Brussels Effect: How the European Union Rules the World (2019), https://doi.org/10.1093/oso/9780190088583.001.0001 (last visited May 10, 2024).
5. TikTok, An Update on Fulfilling our Commitments under the Digital Services Act, (2019), https://newsroom.tiktok.com/en-eu/fulfilling-commitments-dsa-update (last visited Apr 23, 2024).
6. Nick Clegg, *New Features and Additional Transparency Measures as the Digital Services Act Comes into Effect*, Meta, https://about.fb.com/news/2023/08/new-features-and-additional-transparency-measures-as-the-digital-services-act-comes-into-effect/ (last visited Mar 24, 2024).
7. New Features and Transparency Measures for Snapchatters in the European Union to Comply with the Digital Services Act, https://newsroom.snap.com/en-GB/digital-services-act-snap (last visited Mar 24, 2024).
8. Christoph Beuttler, *80 Moments That Shaped the World: The Shipping Container*, British Council (2014), https://www.britishcouncil.org/voices-magazine/80-moments-shaped-world-shipping-container (last visited May 10, 2024).
9. 17 U.S. Code § 512 - Limitations on liability relating to material online, https://www.law.cornell.edu/uscode/text/17/512 (last visited Mar 24, 2024).
10. Jennifer M. Urban, Joe Karaganis & Brianna Schofield, *Notice and Takedown in Everyday Practice*, (2017), https://papers.ssrn.com/abstract=2755628 (last visited Apr 23, 2024).
11. Jennifer M. Urban, Joe Karaganis & Brianna Schofield, *Notice and Takedown in Everyday Practice*, (2017), https://papers.ssrn.com/abstract=2755628 (last visited Apr 23, 2024).
12. Martin Husovec, *How to Facilitate Data Access under the Digital Services Act*, (2023), https://papers.ssrn.com/abstract=4452940 (last visited Apr 23, 2024).
13. Martin Husovec, *How to Facilitate Data Access under the Digital Services Act*, (2023), https://papers.ssrn.com/abstract=4452940 (last visited Apr 23, 2024).

14. 17 U.S. Code § 512 - Limitations on liability relating to material online, https://www.law.cornell.edu/uscode/text/17/512 (last visited Mar 24, 2024).
15. 17 U.S. Code § 512 - Limitations on liability relating to material online, https://www.law.cornell.edu/uscode/text/17/512 (last visited Mar 24, 2024).
16. 47 U.S. Code § 230 - Protection for Private Blocking and Screening of Offensive Material, https://www.law.cornell.edu/uscode/text/47/230 (last visited Mar 24, 2024).

Eleonora Rosati

The DSA's Trusted Flaggers

Revolution, Evolution, or mere Gattopardismo?

One of the most-publicized innovations brought about by the Digital Services Act[1] (DSA or Regulation) is the "institutionalization" of a regime emerged and consolidated for a decade already through voluntary programs introduced by the major online platforms: trusted flaggers. This blogpost provides an overview of the relevant provisions, procedures, and actors. It argues that, ultimately, the DSA's much-hailed trusted flagger regime is unlikely to have groundbreaking effects on content moderation in Europe.

The DSA's trusted flaggers

The (unsurprising) rationale of the system found in Art. 22 DSA is encapsulated in recital 61: by prioritizing the handling of notices submitted by trusted flaggers, "[a]ction against illegal content can be taken more quickly and reliably". Trusted flagger status shall be awarded by the appointed Digital Service Coordinator (DSC) where the applicant is established. Once there, such status shall be recognized by all platforms targeted by the DSA.

During the negotiations leading up to the adoption of the Regulation, a key issue became the eligibility criteria for trusted flaggers. Indeed, the European Commission's original proposal[2] was that only entities (not individuals) representing "collective interests" could – among other requirements – aspire to receive such a recognition. If such a proposal had made its way into the eventual text of the DSA, this would have meant, for example, that corporate entities only representing private interests would have not been in position to access the DSA trusted flagger regime.

The final text of the DSA (thankfully) does not contain such a requirement and instead indicates "private bodies" as also poten-

tially eligible for a trusted flagger designation. Overall, Art. 22(2) provides that an entity (thus, like the Commission's proposal, also excluding individuals) aspiring to receive such a status shall: (a) have particular expertise and competence for the purposes of detecting, identifying and notifying illegal content; (b) be independent from any provider of online platforms; and (c) carry out its activities for the purposes of submitting notices diligently, accurately and objectively.

Recital 61 itself provides examples of entities that will be eligible to become trusted flaggers under the DSA. Reference is made to internet referral units of national law enforcement authorities or of Europol, organizations part of the INHOPE network of hotlines for reporting child sexual abuse material, and organizations committed to notifying illegal racist and xenophobic expressions online.

The list is merely exemplificative. Hence, with reference to, e.g., the creative industries, their trade bodies and industry associations are also obvious candidates for trusted flagger status under the DSA given that (i) one of their key tasks is the online enforcement of their members' rights through specialized and experienced teams and (ii) that is why they are already trusted flaggers through private agreements with platforms, from which they are clearly independent.

Does all this suggest, however, that the trusted flagger "floodgates" are now open to many, if not all? The answer appears to be in the negative, as otherwise the very rationale for having a fast-track notice handling procedure would be lost. Indeed, the DSA specifies that "the overall number of trusted flaggers awarded in accordance with this Regulation should be limited"

in order "[t]o avoid diminishing the added value of such mechanism".

All this means that, while trade bodies and industry associations are encouraged to submit applications to the competent DSC, the DSA shall not affect the ability of private entities and individuals to conclude agreements with online platforms outside of the DSA trusted flagger framework. To be blunt, this sounds like a "nothing new under the sun" result as such agreements have been in place for a long time already. If one thinks for example of copyright, YouTube inaugurated its trusted flagger program as early as 2012.

Nevertheless, the institutional framework that the DSA has created has the potential to be still meaningful, at least for two reasons. The first is that it will likely prompt a standardization of practices and approaches. This consideration is further reinforced by the (very welcome and much needed) harmonization of notice-and-action brought about by Art. 16 DSA. The second reason is that it will serve to complement – in a lex generalis to lex specialis fashion – the regimes contained in subject-matter specific legislation. One such example is Art. 17 of Directive 2019/790 (DSM Directive).

Trusted flaggers and Art. 17 of the DSM Directive

As Art. 17 of the DSM Directive moves from the consideration that, by storing and making available user-uploaded content, online content-sharing service providers (OCSSPs) directly perform acts of communication and making available to the public, the operators of such platforms are required to secure relevant authorizations from concerned rightholders to undertake such activities. Never-

theless, it might be the case that, despite the "best efforts" made by OCSSPs in accordance with Art. 17(4)(a), no such authorization is ultimately secured, given that rightholders are not required to grant it. In such a case, OCSSPs can still escape liability by complying with the cumulative requirements under Art. 17(4)(b)-(c).

In *Poland* (C-401/19[3]), the Grand Chamber of the Court of Justice of the European Union (CJEU) considered that the liability mechanism referred to in Art. 17(4) "is not only appropriate but also appears necessary to meet the need to protect intellectual property rights". In this regard, two notable points may be extrapolated.

The first is that the use of automated content recognition technologies appears unavoidable under Art. 17(4)(b)-(c): content moderation at a scale cannot be performed manually. Nevertheless, the CJEU has only allowed such technologies insofar as they are capable to distinguish adequately between lawful and unlawful uploads. In this regard the DSA will once again play a key role: the transparency obligations set forth therein will serve indeed to determine if the technologies employed by platforms that qualify as OCSSPs satisfy the CJEU mandate.

The second point reflects the scale of OCSSPs' content moderation obligations: obviously, someone must be sending all those notices! In this regard, it is apparent that, at least in certain sectors (think of music, for example), "trusted rightholders" will continue playing a very substantial role within the architecture of Art. 17. In turn, platforms will need to prioritize their notices in order to comply with the obligations set forth in Art. 17(4)(b)-(c).

The latter point is further confirmed if one considers the six key safeguards identified by the CJEU in *Poland*, notably the third one: OCSSPs shall be led to make content unavailable under Art. 17(4)

(b)-(c) upon condition that rightholders provide them with the relevant and necessary information. Clearly, entities that qualify as trusted flaggers in the creative industries will play a most significant role, whether it is through the DSA-sanctioned model or through existing or new private agreements with OCSSPs. In this sense, it will be intriguing to see if a competition arises between private trusted flagger programs and DSC-run ones, in the sense that the former might prove to be more attractive to rightholders (also because of fewer and/or less stringent obligations than those under Art. 22 DSA) than the latter. In any event, it appears that the notices that rightholder will submit shall comply with the requirements set forth in the DSA.

So what?

In light of everything that precedes, is the much-publicized DSA's trusted flagger regime to be regarded as a ground-breaking innovation? For the time being, that does not seem to be the case. All this might evoke – at least in the minds of the most cynical readers, perhaps even including myself – that statement from Giuseppe Tomasi di Lampedusa's Il Gattopardo, which famously reads: "Se vogliamo che tutto rimanga com'è, bisogna che tutto cambi" ("If we want things to stay as they are, things will have to change.").

Nevertheless, and at the very least, the institutional and harmonized shape conferred to trusted flaggers has the potential to smooth out divergences emerged in practice and meaningfully complement the legal regimes provided for in subject-matter specific legislation, including but obviously not limited to the field of copyright.

For this (positive) development to happen and thus avoid an insidious form of gattopardismo, however, it will be first necessary to see how appointed DSCs will handle their role, who will be awarded the trusted flagger status, and how the procedure will work in practice, including having regard to trusted flaggers' own obligations under Art. 22. In any event, it appears safe to conclude the "institutionalized" trusted flagger regime of the DSA shall not replace but, rather, complement (or even compete with!) the voluntary trusted flagger programs already in place.

References

1. 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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).
2. European 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, (2020), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0825 (last visited Mar 24, 2024).
3. Judgment of 26 April 2022, C-401/19, *Poland v Parliament and Council*, ECLI:EU:C:2022:297.

Rachel Griffin, Erik Stallman

A Systemic Approach to Implementing the DSA's Human-in-the-Loop Requirement

Policymakers and the public are increasingly concerned about a lack of transparency and accountability in content moderation. Opaque and incontestable content moderation decisions have potential impacts on freedom of expression and media freedom[1], and well-known issues of discrimination and bias[2]. In the EU, improving fairness and accountability in content moderation is one important policy objective[3] of the 2022 Digital Services Act[4] (DSA).

Our contribution focuses on a core component of this legislative framework: Art. 20 DSA, which sets out rules for online platforms' internal complaint-handling systems. Art. 20 requires platforms to allow users to challenge moderation decisions, and have their complaints reviewed "under the supervision of appropriately qualified staff". Although scholars and commentators have raised important questions about the utility of trying to regulate complex, large-scale content moderation systems[5] via "due process" for individuals[6], this approach is now entrenched in European law. Accordingly, our focus here is on how Art. 20 can and should be interpreted going forward. Specifically, does Art. 20 require a human content moderator to review every content moderation decision on request? And should it?

Drawing on the broader literature on "human in the loop" requirements in artificial intelligence (AI) governance, we argue that formalistically requiring a human to look over every complaint is both normatively problematic and practically counterproductive. We set out an alternative approach, in which human review is oriented towards improving automated moderation systems at a systemic level, rather than correcting individual decisions. We argue that this is both permitted by the DSA text, and normatively preferable as a way of achieving the DSA's ultimate policy goals[7] of preventing arbitrariness and discrimination in moderation.

What level of human review does Art. 20 require?

Art. 20 requires online platforms to establish "easy to access, user-friendly" systems which allow users to complain about any moderation decision. This includes all kinds of actions (or inaction) on flagged content – from terminating an entire account to hiding a single comment – as well as decisions not to remove content, and decisions to reduce visibility or impose other interventions short of removal. This implies a vast number of decisions[8] potentially subject to review. Art. 20(4) requires platforms to consider complaints "in a timely, non-discriminatory, diligent and non-arbitrary manner" and reverse decisions where the complaint shows that they are not justified by the law or by platforms' content policies.

The vast majority of moderation decisions potentially subject to Art. 20 complaints are fully automated[9] – the only feasible way of monitoring content across platforms with millions or billions of users. A crucial question is therefore whether Art. 20 requires complaints to be reviewed by human moderators. The answer not only implies potentially enormous investments of labour time and resources, but also has important implications for the overall effectiveness of the DSA.

Superficially, requiring human moderators to review complaints could seem like the most natural interpretation of Art. 20. However, a close reading suggests otherwise. The key provision is Art. 20(6), which requires "decisions [to be] taken under the supervision of appropriately qualified staff, and not solely on the basis of automated means" (our emphasis). This seems to leave space for humans to play a more high-level supervisory role, rather than examining every individual complaint. Further guidance is provided by Recital 58:

> *"providers of online platforms should be required to provide for internal complaint-handling systems**, which** meet certain conditions that aim to ensure that **the systems are** easily accessible and lead to swift, non-discriminatory, non-arbitrary and fair outcomes, **and are subject to human review** where automated means are used."*
> (our emphasis)

Crucially, "are subject to human review" here refers to "systems", not to "complaints". Thus, it is the complaint-handling system as a whole which must be subject to human review and supervision – not necessarily every individual moderation decision. In the following sections, we will argue that this interpretation is not just legally permissible, but strongly preferable as a way of improving the quality, reliability and fairness of content moderation.

What is the point of human review?

The optimal design of human review processes in content moderation ultimately depends on what purposes they are meant to serve. Yet the DSA provides surprisingly little guidance on this. Recital 58 states that, "Recipients of the service should be able to easily and effectively contest [moderation] decisions […] Therefore, providers of online platforms should be required to provide for internal complaint-handling systems". The ultimate purpose of allowing recipients to contest moderation decisions is left unstated.

Turning to the broader literature on human oversight in AI governance, Rebecca Crootof, Margot Kaminski and W. Nicholson Price identify[10] six possible reasons to impose "human in the loop" requirements:

> *"Humans may play (1) corrective roles to improve system performance, including error, situational, and bias correction; (2) justificatory roles to increase the system's legitimacy by providing reasoning for decisions; (3) dignitary roles to protect the dignity of the humans affected by the decision; (4) accountability roles to allocate liability or censure; (5) interface roles to link the systems to human users; and (6) "warm body" roles to preserve human jobs."*

Considering their relevance to content moderation, we first want to emphasise that (6) is here a very bad reason. Moderators' working conditions are notoriously appalling. Major platforms outsource most such labour[11] to Global South countries with lower wages and fewer worker protections, but even for workers in Global North markets[12] – often migrants with few other employment options – it is characterised by fast-paced and stressful work, poor pay, and intense managerial surveillance. While these conditions could conceivably be improved, there is nothing in the DSA (a supposedly "comprehensive"[13] regulation of online content governance) that tries to achieve this – an important point we will return to later. Reviewing harmful or offensive content is also, to some extent, an inherently repetitive, unpleasant, and psychologically taxing job.

It follows from this that reason (3) is also of questionable relevance. We do not believe it serves human dignity to allow every social media user to demand that some poorly paid and treated worker on the other side of the world quickly glances at their content. Reason (4) is also less relevant to content moderation, as the DSA's provisions on intermediary liability[14] and regulatory oversight[15] already regulate platforms' liability for moderation decisions. The most relevant goals for "human in the loop" require-

ments in relation to Art. 20 are therefore (1) improving the performance of moderation systems, including by correcting errors and bias, and (2) and (5), justifying decisions and making them comprehensible to human users.

Human review of every contested decision is neither practical nor desirable

To effectively achieve these goals, we start with two observations about the roles that humans should not play in content moderation. First, it is neither practical nor desirable to have humans review every contested automated moderation decision. Automated moderation exists largely because humans can't operate at the scale required for timely action on content hosted on large platforms. In three months, YouTube removed 9 million videos and 1.16 billion comments[16]. As Evelyn Douek notes,[17] "even the smaller fraction of content moderation decisions that are appealed would still overload anything but an impractically large workforce".

Arguably, moderation workforces already have become impractically large and overloaded. Facebook alone has 15,000 content moderators worldwide.[18] Yet moderators are also highly overworked, required to follow rigidly-defined workflows and meet demanding quotas which do not permit nuanced consideration[19] of individual decisions. Research on "humans in the loop"[20] in AI shows that it is generally difficult for humans to identify and correct errors, due to "automation bias", where people tend to trust and defer to decision-making software. Increasing moderators' workloads is less likely to improve content moderation decisions

than it is to lead to more frequent rubberstamping of automated decisions.

Furthermore, if Art. 20 is interpreted as relying on a huge workforce to review and correct an enormous volume of contested automated moderation decisions, it is remarkable that the DSA contains virtually no regulation[21] of these workers' pay, working conditions, qualifications and training (beyond some basic transparency requirements for very large online platforms, set out in Art. 42). An inflexible and ill-defined human oversight requirement[22] which effectively requires a permanent layer of low-paid, overworked, and over-stressed content moderators is not only in itself normatively problematic, but also seems like a suboptimal way to improve moderation quality.

Second, even assuming platforms could overcome workforce constraints, it is doubtful that a body of consistent reasoned decisions resolving content moderation complaints is a realistic or even desirable outcome. The scale, complexity, and diversity of content available on large online platforms means that "invoking judicial-style norms of reasoning and precedent is doomed to fail[23]." Removing a platform's discretion as to which decisions are subject to further review still leaves a lot of room to tailor the reasoning and outcome of those reviews to limit their current or future impact, as an intensive study of the Meta Oversight Board[24] has shown. And even a fully independent review body faithfully applying its own reasoned decisions to emerging cases would frequently find itself needing to depart from that precedent or a platform's own guidance. Community guidelines are perpetually revised in response to changing circumstances that those guidelines did not anticipate and for which they are a poor fit.

A better approach to human supervision

All content moderation systems are human/machine hybrids[25] regardless of the degree of automation. Moderation software is designed by human engineers, and AI systms[26] are trained on human decisions and evaluations, while hash-matching systems (like YouTube's ContentID system[27] for copyright enforcement) are designed to search for copies of rightsholder-supplied reference files. On the basis that these hybrid systems are the appropriate target for supervision, rather than individual contested decisions, we identify four key considerations to improve their accuracy and proportionality.

First, instead of requiring cursory human review of every individual decision, the best way to evaluate and improve automated moderation is through more systematic oversight: for example, requiring policy experts to review statistically representative samples of decisions. Today's advanced AI tools, which are increasingly being deployed[28] by major platforms for moderation tasks that would previously have required human intervention, rely on learning patterns from enormous datasets. However, recent technological advances are increasingly relying on smaller volumes of high-quality data,[29] carefully curated or even produced to order by highly-qualified workers. A smaller, better-trained and better-paid moderation workforce, which carefully evaluates and provides detailed feedback on a subset of decisions, can oversee and improve moderation systems more effectively than an army of low-paid clickworkers – as well as being preferable from a labour rights perspective. Similarly, where failings are identified in hash-matching tools like ContentID, which scan for and remove copies of millions of files,[30] it would be more productive to identify system-

atic flaws in the processes for (mis)identifying unlicensed and unlawful reproductions of content in their reference databases, rather than just trying to correct errors piecemeal.

Second, for this kind of systematic review to be effective, human reviewers must be able to understand what triggers automated flagging. Drawing on the extensive research literature on AI explainability, moderation systems should be designed to provide human supervisors "meaningful information about the logic involved"[31] in moderation decisions. Conversely, their feedback should improve the automated system's decision-making in future. For example, if the machine failed to distinguish news reporting about terrorist activity from terrorist recruitment propaganda, the human reviewer could identify characteristics that help reinforce the distinction. This "bilateral explainability" should also factor into Art. 20's requirement for supervision by "appropriately qualified" staff. Reviewers should have the qualifications and ability to facilitate machine-readable policy refinements that can minimise future errors.

Third, human supervision should be proportionate to different types of moderation decisions. Given the potential economic, reputational, and emotional consequences[32] when users' entire accounts are removed, such decisions should receive more thorough review than, for example, demonetising content or hiding a comment. Meaningful review of deplatforming decisions should not be reserved for sitting presidents[33]: we would suggest that in general, if someone will completely lose access to a platform, they should be able to appeal to a human customer service representative (potentially with some narrow exceptions, such as spam and duplicate accounts). In these serious cases, human review should not just involve a quick glance at a decision, but should enable

meaningful communication with moderators[34]. Furthermore, where machine learning led to an erroneous deplatforming decision, the human supervisor should ensure the machine learns from its mistake. That could mean reviewing and reannotating the relevant pieces of content used to train the machine learning classifiers that contributed to the erroneous decision.

Finally, human supervisors can appreciate what types of content pose particular concerns in a specific social, cultural, or political context: for example, political misinformation in the lead-up to a close election, or vaccine misinformation during a pandemic. Expert staff can dynamically allocate limited human and computing resources to address current and emerging threats. And given that the DSA itself may increase the risk of "coordinated flagging"[35], including misuse or manipulation of the Art. 20 complaint system, platforms should dedicate some of their data science and cybersecurity resources to monitoring and addressing these risks – as they have historically done for threats like coordinated disinformation campaigns[36].

Conclusions

In the context of content moderation, we have argued against formalistic interpretations of human oversight requirements that simply require a person to confirm algorithmic decisions – whether based on the premise that the "human touch" somehow makes decisions more respectful of people's dignity, or on the optimistic assumption that having humans look at a decision is sufficient to correct algorithmic errors and bias. Instead, human review under Art. 20 DSA should be geared towards improving the reliability and explainability of algorithmic moderation systems as a whole, as

well as providing meaningful communication and support to users in the most consequential decisions (deplatforming).

These basic principles have wider relevance for tech regulation. For example, "human in the loop" requirements are also established in the EU's GDPR[37] and proposed AI Act[38], as well as under various US legal frameworks[39]. Ultimately, the optimal design of hybrid decision-making systems needs to be adapted to specific contexts. However, the approach we have set out here – interpreting "human in the loop" requirements purposively, and considering how review processes can be designed to serve the legislation's underlying normative and policy goals, rather than just checking a box – could also provide a helpful starting point for interpreting such requirements across different areas of AI regulation.

References

1. European Parliament Press Room, Media Freedom Act: MEPs Tighten Rules to Protect Journalists and Media Outlets, https://www.europarl.europa.eu/news/en/press-room/20230929IPR06111/media-freedom-act-meps-tighten-rules-to-protect-journalists-and-media-outlets (last visited Mar 24, 2024).
2. Rachel Griffin, *The Sanitised Platform*, 13 Journal of Intellectual Property, Information Technology and E-Commerce Law (2022), https://www.jipitec.eu/issues/jipitec-13-1-2022/5514 (last visited May 10, 2024).
3. European Commission, The Impact of the Digital Services Act on Digital Platforms, https://digital-strategy.ec.europa.eu/en/policies/dsa-impact-platforms (last visited Mar 24, 2024).
4. 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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).
5. Evelyn Douek, *The Siren Call of Content Moderation Formalism*, (Lee C. Bollinger & Geoffrey R. Stone eds., 2022), https://papers.ssrn.com/abstract=4005314 (last visited Apr 23, 2024).
6. Rachel Griffin, *Public and Private Power in Social Media Governance: Multistakeholderism, The Rule of Law and Democratic Accountability*, 14 Transnational Legal Theory (2023), https://www.tandfonline.com/doi/full/10.1080/20414005.2023.2203538 (last visited Apr 23, 2024).
7. European Commission, The Impact of the Digital Services Act on Digital Platforms, https://digital-strategy.ec.europa.eu/en/policies/dsa-impact-platforms (last visited Mar 24, 2024).
8. Daphne Keller, *The Eu's New Digital Services Act and the Rest of the World*, Verfassungsblog (2022), https://verfassungsblog.de/dsa-rest-of-world/ (last visited Mar 24, 2024).
9. Google Transparency Report, https://transparencyreport.google.com/youtube-policy/removals?hl=en (last visited Mar 24, 2024).
10. Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 Vanderbilt Law Review (2023), https://scholarship.law.vanderbilt.edu/vlr/vol76/iss2/ (last visited Mar 24, 2024).
11. Sana Ahmad & Martin Krzywdzinski, *Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains*, *in* Digital Work in the Planetary Market (Mark Graham & Fabian Ferrari eds., 2022), https://doi.org/10.7551/mitpress/13835.003.0008 (last visited Mar 24, 2024).

12. Sana Ahmad & Maximilian Greb, *Automating Social Media Content Moderation: Implications for Governance and Labour Discretion*, 2 Work in the Global Economy (2022), https://doi.org/10.1332/273241721X16647876031174 (last visited Apr 23, 2024).

13. European Commission Press Corner, Digital Services Act: EU's Landmark Rules for Online Platforms Enter Into Force, https://ec.europa.eu/commission/presscorner/detail/en/IP_22_6906 (last visited Apr 24, 2024).

14. Folkert Wilman, *Between Preservation and Clarification: The Evolution of the Dsa's Liability Rules in Light of the Cjeu's Case Law*, (2022), https://verfassungsblog.de/dsa-preservation-clarification/ (last visited Mar 24, 2024).

15. Julian Jaursch, *Platform Oversight: Here Is What a Strong Digital Services Coordinator Should Look Like*, *in* Putting the DSA into Practice: Enforcement, Access to Justice, and Global Implications (Joris van Hoboken et al. eds., 2022), https://verfassungsblog.de/dsa-dsc/ (last visited Mar 24, 2024).

16. *NetChoice v. Paxton*, No. 21-51178 (5th Cir. 2022), Brief of Appellees, https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=3657&context=historical (last visited Apr 24, 2024).

17. Evelyn Douek, *The Siren Call of Content Moderation Formalism*, (Lee C. Bollinger & Geoffrey R. Stone eds., 2022), https://papers.ssrn.com/abstract=4005314 (last visited Apr 23, 2024).

18. Casey Newton, *The Secret Lives of Facebook Moderators in America*, The Verge, Mar. 26, 2019, https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona (last visited Mar 24, 2024).

19. Sana Ahmad & Maximilian Greb, *Automating Social Media Content Moderation: Implications for Governance and Labour Discretion*, 2 Work in the Global Economy (2022), https://doi.org/10.1332/273241721X16647876031174 (last visited Apr 23, 2024).

20. Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, Computer Law & Security Review (2022), https://doi.org/10.1016/j.clsr.2022.105681 (last visited May 10, 2024).

21. Beatriz Botero Arcila & Rachel Griffin, *Social Media Platforms and Challenges for Democracy, Rule of Law and Fundamental Rights*, (2023), https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2023)743400 (last visited May 10, 2024).

22. Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, Computer Law & Security Review (2022), https://doi.org/10.1016/j.clsr.2022.105681 (last visited May 10, 2024).

23. Evelyn Douek, *The Siren Call of Content Moderation Formalism*, (Lee C. Bollinger & Geoffrey R. Stone eds., 2022), https://papers.ssrn.com/abstract=4005314 (last visited Apr 23, 2024).

24. Evelyn Douek, *The Meta Oversight Board and the Empty Promise of Legitimacy*, (2023), https://papers.ssrn.com/abstract=4565180 (last visited Apr 23, 2024).

25. Meg L. Jones, *The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles*, 18 Vanderbilt Journal of Entertainment and Technology Law, https://scholarship.law.vanderbilt.edu/jetlaw/vol18/iss1/3 (last visited May 10, 2024).

26. Robert Gorwa, Reuben Binns & Christian Katzenbach, *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance*, 7 Big Data & Society (2020), http://journals.sagepub.com/doi/10.1177/2053951719897945 (last visited Apr 23, 2024).

27. João Pedro Quintais, Giovanni De Gregorio & João C. Magalhães, *How Platforms Govern Users' Copyright-Protected Content: Exploring the Power of Private Ordering and Its Implications*, Computer Law & Security Review (2023), https://linkinghub.elsevier.com/retrieve/pii/S0267364923000031 (last visited Apr 23, 2024).

28. Rachel Griffin, *Algorithmic Content Moderation Brings New Opportunities and Risks*, Centre for International Governance Innovation (2023), https://www.cigionline.org/articles/algorithmic-content-moderation-brings-new-opportunities-and-risks/ (last visited May 10, 2024).

29. Josh Dzieza, *Inside the AI Factory*, The Verge (2023), https://www.theverge.com/features/23764584/ai-artificial-intelligence-data-notation-labor-scale-surge-remotasks-openai-chatbots (last visited Mar 24, 2024).

30. João Pedro Quintais, Giovanni De Gregorio & João C. Magalhães, *How Platforms Govern Users' Copyright-Protected Content: Exploring the Power of Private Ordering and Its Implications*, Computer Law & Security Review (2023), https://linkinghub.elsevier.com/retrieve/pii/S0267364923000031 (last visited Apr 23, 2024).

31. Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, (2017), https://papers.ssrn.com/abstract=3039125 (last visited Apr 23, 2024).

32. Carolina Are & Pam Briggs, *The Emotional and Financial Impact of De-Platforming on Creators at the Margins*, 9 Social Media + Society (2023), https://journals.sagepub.com/doi/10.1177/20563051231155103 (last visited Apr 23, 2024).

33. Nick Clegg, *Oversight Board Upholds Facebook's Decision to Suspend Donald Trump's Accounts*, Meta, https://about.fb.com/news/2021/05/facebook-oversight-board-decision-trump/ (last visited Mar 24, 2024).

34. Carolina Are & Ysabel Gerrard, *Violence and the Feminist Potential of Content Moderation*, *in* The Routledge Companion to Gender, Media and Violence (Karen Boyle & Susan Berridge eds., 2023), https://www.taylorfrancis.com/chapters/edit/10.4324/9781003200871-51/violence-feminist-potential-content-moderation-carolina-ysabel-gerrard (last visited May 10, 2024).

35. Cynthia Khoo, *Coordinated Flagging*, *in* Glossary of Platform Law and Policy Terms (online) (Luca Belli, Nicolo Zingales, & Yasmin Curzi eds.), https://platformglossary.info/coordinated-flagging/ (last visited Apr 23, 2024).

36. Meta, Coordinated Inauthentic Behavior Archives, (2018), https://about.fb.com/news/tag/coordinated-inauthentic-behavior/ (last visited Mar 24, 2024).

37. Art. 22 GDPR – Automated individual decision-making, including profiling, https://gdpr-info.eu/art-22-gdpr/ (last visited Mar 24, 2024).

38. Jorge Constantino, *Exploring Article 14 of the EU AI Proposal: Human in the Loop Challenges When Overseeing High-Risk AI Systems in Public Service Organisations*, 14 Amsterdam Law Forum (2022), https://amsterdamlawforum.org/articles/464 (last visited May 25, 2024).

39. Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 Vanderbilt Law Review (2023), https://scholarship.law.vanderbilt.edu/vlr/vol76/iss2/ (last visited Mar 24, 2024).

Niva Elkin-Koren

A2D for Researchers in Digital Platforms

Over the past decade, access to data (A2D) in digital platforms has emerged as a significant challenge within the research community. Researchers seeking to explore data hosted on these platforms encounter growing obstacles. Public policy concerning such access must navigate through conflicting interests involving various stakeholders, including platforms, its users, competitors, the scientific community, and the public at large. While legal policies in the US have generally focused on establishing safeguards for researchers against the restrictions on access imposed by private ordering, the recent EU Digital Service Act[1] (DSA) introduces a legal framework, which enables researchers to compel platforms to provide data access. These complementary legal strategies may prove instrumental in facilitating A2D for research purposes.

A2D in digital platforms

Data constitutes the fundamental business asset of digital platforms. These platforms collect data on users' online behaviour and generate income by utilizing these profiles for targeted advertising, as well as for creating additional data-driven products and services. Platforms have worries about the potential disclosure of sensitive data, which could breach users' privacy.[2] Data leaks may also trigger legal liability and could also damage platform's public reputation.

At the same time, however, strong public interests advocate for ensuing A2D for scientific purposes. Platforms often provide a unique access point to data, which can be indispensable for basic research.[3] For example, it may be essential for detecting early indicators of imminent natural disasters, identifying markers for infectious disease outbreaks, or developing new research

methodologies[4] employing Artificial Intelligence. A2D in digital platforms also plays a critical role in exploring the digital transformation. As societal, economic, and political activities migrate to digital spaces, A2D becomes imperative for mapping and analyzing the social implications of this shift. This includes investigating issues such as discrimination in labor markets[5], bias in short term rentals[6], or the impact of political advertising on elections[7].

Furthermore, as platforms continue to grow in dominance and significance, infiltrating the social, economic and political arenas, there is a stronger imperative to bolster their accountability by advancing transparency[8] and oversight. Enabling independent scientific research into these issues by granting scientists access to platform data can provide unbiased evidence to guide public oversight and complement investigative efforts undertaken by public authorities.

Occasionally, digital platforms have chosen to voluntarily share data with academic researchers. For example, recent papers published in Science[9] and in Nature[10] saw 17 researchers collaborating with Meta, concluding that there was no evidence of social media platforms, like Facebook and Instagram, polarizing voters during the 2020 US Elections. However, concerns have been raised by some scholars[11] that these findings may have been influenced by Meta's involvement in the research collaboration and could align with its business interests.

Ensuring A2D for independent researchers who are not affiliated with these platforms, has the potential to diversify the research agenda. It can foster studies driven by pure scientific curiosity and intellectual freedom, rather than profits-driven motives. Moreover, it can empower researchers to challenge conclusions drawn in other studies based through independent data

analysis. Overall, safeguarding A2D for research is of utmost importance in preserving the social and political role of academic research as an unbiased and independent source of reliable knowledge.

Private ordering and its limits

Despite its significant public implications, decisions regarding whether to permit A2D have so far rested solely with digital platforms. As users' content, personal data and activities predominantly occur on their facilities, platforms possess the capability to technically block data access. Platforms have exercised their physical control[12] over users' data, to prevent researchers from conducting studies. Instances such as the Cambridge Analytica scandal, where personal data of millions of Facebook users was misused, led platforms like Facebook and Instagram to block Application Programming Interface (API) access. More recently, X (formally Twitter) announced[13] its decision to restrict free API access for research purposes. Additionally, platforms have prevented[14] on multiple occasions, the scrapping of publicly available data, and obstructed other efforts to explore their operation from the outside. One notable example is the NYU Ad Observatory[15], which was established to analyze political advertisements on social media. Through a browser extension ("Ad Observer"), users were able to donate ad data scraped from Facebook to the Observatory, helping to verify and supplement some missing data in Facebook's own Ad Library. However, in August 2021 Facebook suspended the accounts[16] of researchers involved in this initiative.

Platforms also employ contractual claims as a means to restrict undesired research activities. For instance, the X Corp. has recently

filed a lawsuit against[17] the Center for Countering Digital Hate (CCDH), a non-profit organization that conducted research on the dissemination of hateful content on social media. X alleged that CCDH had intentionally and unlawfully scraped data from Twitter, thereby violating its terms of service (ToS). TikTok has taken a more stringent approach by imposing additional contractual requirements in its Research API ToS[18], requiring academics to provide advance notice of their forthcoming research, subject their work to pre-publication review, and delete certain data once it has been used.

U.S. and EU legal strategies compared - the shield and the sword

While self-help measures aimed at restricting A2D often serve the legitimate interests of platforms, policymakers must also ensure proper access to platform data for independent scientific purposes. Striking this balance presents a significant challenge.

The U.S. and Europe have adopted distinct legal approaches to address this challenge. In the U.S., the emphasis has been on defensive strategies designed to protect researchers from liability stemming from breach of contract and potential criminal liability related to the unauthorized scraping of platform data. In contrast, Europe has recently established a proactive framework, grating researchers a legal right to acquire data that is essential for their research endeavours. These strategies are further discussed below.

Research shield: The U.S. approach

Platforms ToS typically impose restrictions on unauthorized data collection, including for research purposes. This exposes researchers to the risk of civil liability for breaching contractual agreements. Moreover, under U.S. law, unauthorized access to platforms' computational services, allegedly may trigger criminal liability under the U.S. Criminal Fraud and Abuse Act[19] (CFAA). These risks can significantly deter independent research conducted on platforms.

However, recent court decisions have adopted a narrow interpretation of the CFAA, thereby reducing the risks faced by researchers who are studying platforms without prior authorization. For example, in the case of *Sandvig v Barr*[20], the DC District Court examined whether researchers investigating race and gender discrimination in employment websites violate the CFAA. The researchers planned to create multiple fake accounts, contravening the websites' ToS, which prohibited misrepresentation. The court held that CFAA does not criminalize mere violations of ToS on consumer websites. In another case, *hiQ Labs v. LinkedIn*[21], which was a commercial legal dispute, the Ninth Circuit determined that the CFAA does not apply to the scraping of publicly available data. Accessing such data, the court held, cannot be considered "unauthorized" under the CFAA.

When A2D is carried out in violation of the ToS, it may also lead to civil liability for breaching a contract, along with the associated legal remedies. However, it is worth noting that restrictive provisions on A2D may not be enforceable if they are preempted under the preemption doctrine set forth in section 301(a) of the U.S. 1976 Copyright Act[22]. The preemption doctrine is designed to

uphold the Copyright Act's exclusivity in governing copyright matters. It invalidates any rules that offer copyright-like protection (e.g., restrictions on reproduction) to non-copyrightable subject matters, such as unoriginal data. Back in the mid-90s, in the case of *ProCD v. Zeindeberg*[23], the Plaintiff attempted to protect uncopyrightable digitized telephone listings using a shrink-wrap license. The Court of Appeals for the 7th Cir. held that such contracts only impact the parties involved and cannot establish rights in rem equivalent to copyright. Consequently, contractual restrictions could never be preempted. Note, however, that restrictions on A2D in platforms' ToS lack privity. They are boilerplate contracts[24] that apply to anyone accessing the platform, Therefore, if these restrictions are deemed enforceable, they effectively create de facto rights against the world[25].

Arguably, restrictions on A2D for research purposes run counter to the objectives of copyright law. These restrictions aim to prohibit the reproduction of data, a subject matter that was intentionally excluded from copyright protection to guarantee its availability for everyone to use as building blocks of additional creative works. Moreover, these limitations also appear to undermine the right to research[26], a right safeguarded under fair use provisions, which serves the overarching goals of copyright law – namely, fostering learning, generating new knowledge and upholding the principles of freedom of expression.

Despite extensive criticism from legal scholars[27] regarding the *ProCD* narrow interpretation of the preemption doctrine, most courts[28] have adopted this approach in the past decades and have rejected the preemption of contractual restrictions. However, in a recent decision the 2nd Cir. reaffirmed a pre-emption claim in the scraping lawsuit of *Genius v. Google*[29]. The decision to deny appeal

to the Supreme Court[30] may indicate that pre-emption claims in boilerplate contracts and platform ToS might gain more traction in the future.

EU: From shield to sword

Responding to mounting pressure from researchers and civil society organizations advocating for greater oversight of digital platforms through independent studies, the EU has adopted a proactive approach. This approach delegates decisions regarding access to platform data to a regulatory agency, which exercises its discretion within a set of explicit objective standards. The DSA[31] establishes an institutional framework, aiming to streamline A2D for research in the public interest while also addressing the legitimate interests of platforms and their users.

The DSA introduces a novel regulatory body, the Digital Services Coordinators (DSC, see Arts. 49 to 51), tasked, inter alia, with the management of data access authorizations. This transfer of authority shifts the decision-making power regarding A2D from profit-driven platforms to an administrative agency entrusted with upholding the public interest.

Furthermore, the DSA establishes a structured procedure for obtaining A2D for research purposes, including a filing procedure and eligibility criteria for researchers and their proposed research projects.

Most notably, the DSA obliges very large online platforms and search engines (VLOPs and VLOSEs) to provide data to "vetted researchers" (see Art. 40(4) and (8)) for the sole purpose of conducting research that contributes to the detection, identification and understanding of systemic risks in the Union, as set out pursuant

to Art. 34(1), and to the assessment of the adequacy, efficiency and impacts of the risk mitigation measures pursuant to Art. 35." (see Art. 40(4)). Through this obligation, the DSA effectively establishes a (limited) right to conduct academic research on systemic risk involving digital platforms in the EU. This right encompasses the ability to request data collection, using APIs, or other means of automatic extraction. It is critical for conducting research in the digital era and could have proven invaluable as exemplified in the case of the NYU research team, which was cut out from Facebook API.

Recently, the EU Commission has launched a call for evidence on the DSA related to data access for research purposes, intended to inform the implementation of Art. 40 DSA. Respondents to this call[32] have stressed the need to provide standard procedures and criteria for eligibility to vetted researchers, to establish an independent advisory body with professional expertise and to address liability for potential data breach. They also stressed the need to facilitate exploratory research and enable automated API based exploration. Based on the contributions received, the Commission is scheduled to prepare a delegated act on Art. 40 to be adopted in 2024.

A way forward

Science is a global collaborative endeavor that relies on cooperative efforts, peer review, and the free exchange of information and knowledge across national boundaries and disciplines. Digital platforms where A2D is essential, also operate on a global scale. However, there exists a fundamental disparity in the legal approaches to A2D for researchers between the U.S., and the EU. This

divergence has the potential to disrupt collaborative scientific initiatives and could shape where and how scientific research is conducted.

While the DSA may still have some imperfections,[33] it marks a significant stride towards establishing a legal right for researchers to request A2D and put in place an institutional framework to facilitate the exercise of this right. The U.S. currently lacks a comparable framework, although there are several bills, such as the Platform Accountability and Transparency Act[34] and the Digital Consumer Protection Commission Act[35] that propose mandating digital platforms to provide certain types of data for research purposes. However, as of now, these bills have not been enacted into law.

Meanwhile, in the EU, data protection laws[36] and more robust intellectual property protections[37] for data may create significant barriers to unauthorized data scrapping for research purposes.

Bridging the divide between the approaches of the U.S. and EU presents a formidable challenge, raising a multitude of complex issues, including the legitimate rights of digital platforms, freedom of contract, freedom of expression, privacy and data protection.

A potentially more effective strategy for fostering ongoing scientific collaboration could involve coordinating research initiatives that leverage the legal safeguards available for unauthorized research in the U.S. and the right to request A2D guaranteed by the EU's new digital strategy.

References

1. 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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).
2. Mike Clark, *Research Cannot Be the Justification for Compromising People's Privacy*, Meta (2021), https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/ (last visited Mar 24, 2024).
3. Nicholas Proferes et al., *Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, And Ethics*, 7 Social Media + Society (2021), http://journals.sagepub.com/doi/10.1177/20563051211019004 (last visited Apr 23, 2024).
4. Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 UC Davis Law Review (2019), https://lawreview.law.ucdavis.edu/archives/53/2/copyright-and-progress-science-why-text-and-data-mining-lawful (last visited Apr 23, 2024).
5. Arianne Renan Barzilay & Anat Ben-David, *Platform Inequality: Gender in the Gig-Economy*, (2017), https://papers.ssrn.com/abstract=2995906 (last visited Mar 24, 2024).
6. Benjamin Edelman, Michael Luca & Dan Svirsky, *Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment*, 9 American Economic Journal: Applied Economics (2017), https://www.aeaweb.org/articles?id=10.1257/app.20160213 (last visited Mar 24, 2024).
7. Alexander Coppock, Donald P. Green & Ethan Porter, *Does Digital Advertising Affect Vote Choice? Evidence from a Randomized Field Experiment*, 9 Research & Politics (2022), http://journals.sagepub.com/doi/10.1177/20531680221076901 (last visited Apr 23, 2024).
8. Robert Gorwa & Timothy Garton Ash, *Democratic Transparency in the Platform Society*, *in* Social Media and Democracy: The State of the Field, Prospects for Reform (Nathaniel Persily & Joshua A. Tucker eds.), https://www.cambridge.org/core/product/identifier/9781108890960/type/book (last visited Apr 23, 2024).
9. Sandra González-Bailón et al., *Asymmetric Ideological Segregation in Exposure to Political News on Facebook*, 381 Science (2023), https://www.science.org/doi/10.1126/science.ade7138 (last visited Apr 23, 2024).
10. David Garcia, *Influence of Facebook Agorithms on Political Polarization Tested*, Nature (2023), https://www.nature.com/articles/d41586-023-02325-x (last visited Mar 24, 2024).

11. Kai Kupferschmidt, *Does Social Media Polarize Voters? Unprecedented Experiments on Facebook Users Reveal Surprises*, (2023), https://www.science.org/content/article/does-social-media-polarize-voters-unprecedented-experiments-facebook-users-reveal (last visited Apr 23, 2024).

12. Deen Freelon, *Computational Research in the Post-API Age*, 35 Political Communication (2018), https://www.tandfonline.com/doi/full/10.1080/10584609.2018.1477506 (last visited Apr 23, 2024).

13. Kai Kupferschmidt, *Twitter's Plan to Cut Off Free Data Access Evokes 'Fair Amount of Panic' among Scientists*, Science (2023), https://www.science.org/content/article/twitters-plan-cut-free-data-access-evokes-fair-amount-panic-among-scientists (last visited Apr 23, 2024).

14. Jeremy B. Merrill & Ariana Tobin, *Facebook Moves to Block Ad Transparency Tools – Including Ours*, (2019), https://www.propublica.org/article/facebook-blocks-ad-transparency-tools (last visited Mar 24, 2024).

15. NYU Ad Observatory, Transition to Static, https://adobservatory.org/transition-to-static (last visited Mar 24, 2024).

16. Lois Anne DeLong, *Facebook Disables Ad Observatory; Academicians and Journalists Fire Back*, NYU Center for Cyber Security (2021), https://cyber.nyu.edu/2021/08/21/facebook-disables-ad-observatory-academicians-and-journalists-fire-back/ (last visited Mar 24, 2024).

17. Emma Roth, *Elon Musk's X Sues Anti-Hate Researchers for Allegedly Scraping Data from Twitter*, The Verge (2023), https://www.theverge.com/2023/8/1/23815515/twitter-ccdh-anti-hate-research-group-lawsuit (last visited May 25, 2024).

18. TikTok Research API Terms of Service, https://www.tiktok.com/legal/page/global/terms-of-service-research-api/en (last visited Mar 24, 2024).

19. Wikipedia, Computer Fraud and Abuse Act, (2024), https://en.wikipedia.org/w/index.php?title=Computer_Fraud_and_Abuse_Act&oldid=1213092789 (last visited Mar 24, 2024).

20. *Sandvig v. Barr*, 451 F. Supp. 3d 73 (D.D.C. 2020), https://casetext.com/case/sandvig-v-barr (last visited Apr 24, 2024).

21. *HiQ Labs, Inc. v. LinkedIn Corporation*, No. 17-16783 (9th Cir. 2022), https://cdn.ca9.uscourts.gov/datastore/opinions/2022/04/18/17-16783.pdf (last visited Apr 24, 2024).

22. 17 U.S. Code § 301 - Preemption with respect to other laws, https://www.law.cornell.edu/uscode/text/17/301.

23. *ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447 (7th Cir. 1996), https://law.justia.com/cases/federal/appellate-courts/F3/86/1447/538242/ (last visited Apr 24, 2024).

24. Amit Elazari Bar On, *Unconscionability 2.0 and the IP Boilerplate: A Revised Doctrine of Unconscionability for the Information Age*, 34 Berkeley Technology Law Journal (2020), https://lawcat.berkeley.edu/record/1136668 (last visited Apr 23, 2024).

25. Niva Elkin-Koren, *Copyright Policy and the Limits of Freedom of Contract*, 12 Berkeley Technology Law Journal (1997), https://lawcat.berkeley.edu/record/1115954 (last visited Apr 23, 2024).

26. Sean Flynn et al., *Research Exceptions in Comparative Copyright*, PIJIP/TLS Research Paper Series (2022), https://www.ila-americanbranch.org/wp-content/uploads/2023/09/Research-Exceptions-in-Comparative-Copyright.pdf (last visited May 10, 2024).

27. Jessica D. Litman & Pamela Samuelson, *The Copyright Principles Project: Directions for Reform*, 25 Berkeley Technology Law Journal (2010), https://repository.law.umich.edu/articles/1381 (last visited May 10, 2024).

28. Guy A. Rub, *Copyright Survives: Rethinking the Copyright-Contract Conflict*, (2017), https://papers.ssrn.com/abstract=2926253 (last visited Mar 24, 2024).

29. *ML Genius Holdings LLC v. Google LLC*, No. 20-3113 (2d Cir. 2022), https://storage.courtlistener.com/recap/ (last visited Apr 24, 2024).

30. Blake Brittain, *US Supreme Court Lets Google Win Stand against Genius Suit over Song Lyrics*, Reuters (2023), https://www.reuters.com/legal/us-supreme-court-lets-google-win-stand-against-genius-suit-over-song-lyrics-2023-06-26/ (last visited May 10, 2024).

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), http://data.europa.eu/eli/reg/2022/2065/oj/eng (last visited Mar 24, 2024).

32. European Commission, Digital Services Act: Summary Report on the Call for Evidence on the Delegated Regulation on Data Access, (2023), https://digital-strategy.ec.europa.eu/en/library/digital-services-act-summary-report-call-evidence-delegated-regulation-data-access (last visited Mar 24, 2024).

33. Aline Iramina, Maayan Perel (Filmar) & Niva Elkin-Koren, *Paving the Way for the Right to Research Platform Data*, (2023), https://papers.ssrn.com/abstract=4484052 (last visited Apr 23, 2024).

34. John Perrino, *Platform Accountability and Transparency Act Reintroduced in Senate*, Cyber Policy Center (2023), https://cyber.fsi.stanford.edu/news/platform-accountability-and-transparency-act-reintroduced-senate (last visited Mar 24, 2024).

35. Digital Consumer Protection Commission Act 2023, https://www.warren.senate.gov/imo/media/doc/Tech%20Bill%20One%20pager.pdf, last visited Apr 24, 2024).

36. Mathias Vermeulen, *The Keys to the Kingdom*, Knight First Amendment Institute (2021), https://perma.cc/M2WP-84YK (last visited Apr 23, 2024).

37. Björn Lundqvist, *An Access and Transfer Right to Data – from a Competition Law Perspective*, 11 Journal of Antitrust Enforcement (2023), https://academic.oup.com/antitrust/article/11/Supplement_1/i57/6696741 (last visited Apr 23, 2024).

Natali Helberger, Pamela Samuelson

The Digital Services Act as a Global Transparency Regime

On both sides of the Atlantic, policymakers are struggling to reign in the power of large online platforms and technology companies. Transparency obligations have emerged as a key policy tool that may support or enable achieving this goal. The core argument of this blog is that the Digital Services Act (DSA) creates, at least in part, a global transparency regime. This has implications for transatlantic dialogues and cooperation on matters concerning platform governance. Regulators, researchers, and civil society organizations may be able to use the DSA transparency rules to improve responsiveness of large platforms and other technology companies to the public values of the larger societies that they serve.

In the United States (U.S.), several members of Congress have proposed bills, including the Platform Accountability and Transparency Act[1], the Social Media Data Act[2], the Digital Services Oversight or the Safety Act[3], and Kids Online Safety Act[4] (KOSA), that would increase transparency obligations about platform content moderation practices, online advertising, and safeguards to protect personal data and children. None of these bills has been enacted, although KOSA is under active consideration.

The main regulatory agency in the U.S. that has engaged in online platform regulation has been the Federal Trade Commission (FTC), which has investigatory powers to demand transparency from platforms or other large companies when they may have engaged in unfair or deceptive practices. The President also has authority to issue Executive Orders, which sometimes includes rules that require technology developers to be more transparent.

Yet, now that the DSA has come into force, the European Union (EU) has taken a very large step ahead of the U.S. in making data usage and content moderation practices of platforms more transparent. Among the host of DSA mandatory transparency require-

ments are those that require preparation of transparency reports, the promulgation of a DSA Transparency Database[5] to report on content moderation practices, new rules about data access requirements for regulators and researchers, preparation of audit reports, a digital terms and conditions (T&Cs) database[6], and the Ad Library. The DSA is a very ambitious policy initiative aimed at cracking open not just one, but many, black boxes.

Although the geographical focus of the DSA is EU member states, some of its transparency provisions may contribute to a global transparency and observability[7] of platforms. The goal of this blog is to examine to what extent the DSA's transparency provisions can potentially benefit researchers and regulators outside the European Union.

Categories of DSA transparency obligations

The transparency obligations in the DSA can usefully be sorted into four[8] categories: 1) consumer-facing transparency obligations; 2) mandatory reporting and information access obligations to national regulators and the European Commission; 3) rights of access to data; and 4) obligations to contribute to public-facing databases of information.

We first discuss the DSA's consumer-facing transparency obligations that require platforms to provide certain types of information to their users. Some of these obligations target all users. For example, Art. 26 of the DSA obliges online platforms to identify advertising as such and to explain their main targeting criteria and how consumers can change these criteria. In addition, Art. 27 obliges platforms to set out in their T&Cs the main parameters used in their recommender systems.

Other DSA transparency rights accrue to individual consumers in particular circumstances. For example, Art. 32 requires online platforms to inform individual consumers if a product or service they acquired through a platform was illegal. Additionally, Art. 16(5) requires platforms to inform users that their content has been taken down.

In principle, these DSA rules are intended to benefit consumers of services established or located in the EU, and they certainly apply to non-European consumers located in the EU.

Although these rules are not directly applicable or enforceable outside the EU, they may potentially benefit non-European consumers through the so-called "Brussels effect" insofar as online platforms decide not to limit these extra transparency rights just to EU consumers. There is no language in most of these provisions that would exclude the applicability of these provisions to consumers located outside the EU.

A second category of transparency obligations includes mandatory reporting and information access obligations to national regulators and the European Commission. Obvious examples are the powers of national Digital Service Coordinators (DSCs) under Art. 51 of the DSA to require covered platforms to provide information and explanations upon request. Arts. 5 and 67 of the DSA gives the Commission investigatory powers as to Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs). These information and investigation powers are reserved to national European regulatory authorities and the Commission.

The DSA requires covered online services to prepare reports annually about their compliance with the DSA and to maintain data pertinent to the reports. However, they are not required to submit these materials annually to the Commission or to a DSCs. The on-

line services must, however, provide their compliance reports to EU regulators when so requested to enable regulators to analyze the extent to which the services have complied with DSA obligations. These online services bear the burden and expense of preparing annual reports and maintaining data that may never be reviewed by any EU regulator. The services can never know when (if ever) regulators will make such requests. But they must be ready to comply.

Art. 37 requires the online services to hire at their own expense independent auditors to assess their compliance with DSA obligations. It further requires services to provide auditors with access to all data needed to conduct an audit and identifies the kinds of data that should be part of an audit. We worry about the lack of well-established auditing standards akin to those long established for financial auditing. The DSA does not contemplate that these audits would be available to the Commission or to DSCs, but one can imagine EU regulators demanding access to them if the regulators were dissatisfied with an online services' annual report once they analyzed a requested copy.

VLOPs and VLOSEs must, in accordance with Art. 42 of the DSA, also prepare reports on their mandatory systemic risk assessments and mitigation measures, audits and audit implementation reports and consultations, as well as reports on the number of monthly users. The Commission and national DSCs of the countries where the platforms are established may require covered platforms and search engines to supply these reports to European authorities.

Regulators from other countries might, however, be interested in gaining access to the annual reports that the DSA requires covered online services to prepare. Art. 40 says EU regulators can only access the reports to assess compliance with the DSA. But

would the Commission object if the FTC, for example, demanded access to online services' annual reports for firms operating in the U.S.? We presume that the FTC could issue a civil investigative demand directly to the services asking for copies of reports prepared for compliance with the DSA.

If the Commission wants to achieve a "Brussels effect" by setting a regulatory standard for other nations to follow, perhaps it would welcome easing the burdens of non-EU regulators in this way.

Systemic risk assessment and monitoring are among the core transparency obligations for VLOPs under the DSA. These requirements respond to growing concerns about the impact of these platforms on the broader information ecosystem and on fundamental rights. This information about systemic risks may potentially be of great interest to regulators outside the EU.

Under Arts. 42 (4) and 42 (5) of the DSA, risk assessment information is to become accessible outside the EU three months after platform reports have been submitted to EU authorities, albeit in possibly redacted form. Under the DSA, providers of VLOPs and VLOSEs can, before the reports become public, remove certain parts that might disclose confidential information, pose security risks, or otherwise harm the firms whose reports became public.

The utility of these reports for non-EU regulators will, of course, depend on how extensively platforms excise information from these reports before making them public. Covered platforms and search engines should not, however, edit the reports to prevent non-EU authorities from being able to access information the reports contain unless one of the legitimate rationales for excision applies.

A third category of DSA transparency rules are those that create a right of access to data that is necessary to monitor and assess compliance. Art. 40's access to data provision allows EU policymakers to obtain a deeper level of observability which would address the growing information asymmetries between platforms and society at large. Professors Rieder and Hofman[9] have observed that "[t]he expanding data sets on vast numbers of people and transactions bear the potential for privileged insights into societies' texture, even if platforms tend to use them only for operational purposes". These authors suggest that an essential pre-condition for public accountability is the "institutionalisation of reliable information interfaces between digital platforms and society – with a broad mandate to focus on the public interest".

We believe that the access to data provisions in Art. 40 of the DSA should be understood to create such an interface. In addition to DSCs and the Commission, "vetted researchers" can request access to data held by VLOPs and VLOSEs to gauge compliance with DSA obligations.

Art. 40 of the DSA contemplates that researchers would submit proposals to DSCs identifying the online service providers whose data they want to access, along with a research plan. Coordinators would then "vet" researchers under the criteria set forth in Art. 40(8). Upon being vetted, the coordinators would notify the online services that the vetted researcher should be given access to data for compliance assessment purposes.

The vetting criteria include supplying information about the research organization with which the researcher is affiliated, their independence from commercial interests, sources of funding for their research, the ability to comply with data security and confidentiality rules, and an intent to carry out research for purposes set

forth in Art. 40(4). To be vetted, researchers must also agree to publish the results of their study without charge within a reasonable time after finishing their research project. This means that the research outputs about DSA compliance will become publicly available to all who may be interested in finding out about how well (or not) platforms did.

Vetted researchers are, however, restricted in the purpose for which they can request access to platform data, for the DSA says vetted researchers can access data only for "the sole purpose of conducting research that contributes to the detection, identification and understanding" of a pre-defined list of systemic risks under Art. 34 of the DSA or the assessment of the "adequacy, efficiency and impacts of the risk mitigation measures" that the DSA requires. In other words, research access is only possible to the extent that it contributes to the enforcement of the DSA.

By authorizing DSCs to require online services to grant independent researchers access to data concerning risk assessment and risk mitigation strategies and to publish results of their research, the DSA offloads some burdens that EU regulators might otherwise have to bear to those researchers whom the coordinators vet.

Practically speaking, this strategy raises important questions about the proper role of researchers in enforcement actions, the need to protect academic independence and autonomy, and how to combine the demands of the DSA with the way academic research is conducted, assessed, and funded.

So far as we can tell, the researcher data access rights set forth in Art. 40 may be available to researchers outside of the EU. There will almost certainly be U.S. researchers who would want to request access to data under this regime because there are no equivalent data access mandates under U.S. law.

Although the DSA does not define which researchers are eligible for data access rights, it refers to the definition of this term in Art. 2(1) of Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market.[10] That provision requires researchers to be affiliated with a "research organization", such as a university, a research institute, or another entity whose primary goal was to conduct scientific research on a not-for-profit basis or pursuant to a public interest mission recognised by a European member state. There is no explicit requirement that this must be a European university or research entity. Nor does Art. 40 (8) say that the DSC can deny an application for data access to non-Europeans (in this sense also Dergacheva, Katzenbach, Schwemer & Quintais 2023[11] and Husovec 2023[12]).

Arguably, it is in the interest of EU policymakers to open up Art. 40 of the DSA to non-European researchers. A significant share of research that has been conducted on platform auditing originates from the U.S. Using the extensive expertise and experience of non-EU researchers for the purposes of assessing compliance with the DSA would be very much in the interest of Europe. (More information on how non-EU researchers might exercise the access right can be found here[13] and here[14].)

A fourth category of transparency rules of the DSA is the obligation of platforms to make certain information publicly available in data bases and ad archives.[15] Examples are the Ad Archives that mandated by Art. 39 of the DSA. Providers of VLOPs and VLOSEs are obliged to make available in a specific portion of their online interface a searchable repository containing information about the content on their online commercial and political advertisements. Also required is disclosure about on whose behalf the advertisement was presented, who paid for the advertisement, groups tar-

geted and targeting parameters, and the total number of recipients. (For an insightful discussion of the design requirements of ad archives, see van Drunen & Noroozian 2024[16]).

Moreover, all platforms covered by the DSA, not just the VLOPs and VLOSEs, must publish statements about reasons for their content moderation actions. Platforms must send those statements to the DSA Transparency Database[17], which is operated by the Commission under Arts. 17 and 24(5) of the DSA. These statements must include information about the type of content moderation restrictions they have adopted, as well as the grounds and the surrounding facts and circumstances that influenced the decision.

Yet another platform transparency resource established by the Commission is the T&Cs Database[18]. Platforms use their T&Cs for registered users as an important source of private governance. The goal of the database is to give the public more information into this element of the legal landscape. Currently, the database includes 790 T&Cs from more than 290 service providers, including Terms of Service, Privacy Policies, but also developer terms.

All of these information resources created and maintained in the EU will be available to anyone in the world who wants to access them.

Does the DSA have a global reach?

The DSA is an ambitious step towards a global transparency regime. A significant share of the transparency obligations in the DSA are not limited to European regulators, consumers and researchers. This includes transparency about platforms' statements of reasons for taking actions, information about political and commercial ads, T&Cs, audits and systemic risk assessments as well as

access to the deeper layers of the algorithmic infrastructure through access to data rights available to stakeholders outside the European Union.

The benefits of transparency for EU and non-EU regulators provided by the DSA may be mutual. By extending the scope of potential observers, the EU too can benefit from the expertise and insights from actors outside the Union.

This more inclusive approach to global transparency resonates with a push for more international coordination and participation in (EU-led) platform governance. In the emerging digital regulatory framework, there are various ways[19] in which non-EU stakeholders, including civil society and potentially non-EU regulators can become involved in and influence EU platform governance.

Under Art. 51 (3) of the DSA, for example, DSCs can invite "interested parties" and "any other third parity demonstrating a legitimate interest" to submit written observations on planned enforcement actions and participate in the proceedings. There is nothing in the text that would exclude non-European regulators, such as the FTC, or non-European competitors from taking an active part in the enforcement deliberations of national DSCs.

The Digital Markets Act[20] (DMA) likewise entitles "[a]ny third party" to inform the national competent authority of the Member State or the Commission about "any practice or behaviour by gatekeepers that falls within the scope of this Regulation" in the context of an infringement procedure under Art. 27 of the DSA. The European Media Freedom Act (EMFA) foresees explicitly the possibility that the Board could coordinate with non-EU regulators under Art. 16 EMFA, and introduces the instrument of so-called "structured dialogues" that are also open to non-EU civil society actors under Art. 18.

In a similar way, the draft AI Act foresees explicitly cooperation and coordination with non-European authorities and international organisations under Art. 58e of the AI Act. The planned Advisory Forum and Scientific Panel are also open to non-EU stakeholders under Arts. 58 a and b, giving those an influential role in the further implementation and operationalisation of the European approach to AI governance.

Another aspect of the AI Act, which is open to non-EU stakeholders, concerns international standardisation in the field of AI. According to Art. 40 (1) (c) of the AI Act, the actors involved in the standardisation process must "contribute to strengthening global cooperation on standardisation and taking into account existing international standards in the field of AI" but also as part of EU-U.S. cooperations such as the EU-U.S. Trade and Technology Council (TTC) [21].

Has the EU, through the DSA and related initiatives, gone a long ways toward achieving a "Next Level Brussels Effect?" From EU regulators' optimistic view, not only would global platforms adhere to, and export European standards into their operations outside of the Union, but there would be a new push to an EU-led approach in the creation of global observability and governance frameworks through transparency, cooperation, codes of conduct and coordination on standardisation.

While we recognize the ambition and optimism that underlies promulgation of the DSA and related initiatives, these new regulations are still in early stages and the regulatory cultures of the EU, U.S., and other nations are distinctly different. Some clashes over the burdens and costs that these new rules impose and the impacts of the rules on competition and innovation in information techno-

logy industries seem quite likely. We look forward to seeing how they play out in coming years.

References

1. S.5339 - Platform Accountability and Transparency Act, https://www.congress.gov/bill/117th-congress/senate-bill/5339 (last visited Apr 24, 2024).
2. H.R.3451 - Social Media DATA Act, https://www.congress.gov/bill/117th-congress/house-bill/3451 (last visited Apr 24, 2024).
3. H.R.6796 - Digital Services Oversight and Safety Act of 2022, https://www.congress.gov/bill/117th-congress/house-bill/6796 (last visited Apr 24, 2024).
4. S.1409 - Kids Online Safety Act, https://www.congress.gov/bill/118th-congress/senate-bill/1409/text (last visited Apr 24, 2024).
5. European Commission, DSA Transparency Database, https://transparency.dsa.ec.europa.eu/ (last visited Mar 24, 2024).
6. Online Platforms Terms and Conditions Database, https://platform-contracts.digital-strategy.ec.europa.eu/ (last visited Mar 24, 2024).
7. Bernhard Rieder & Jeanette Hofmann, *Towards Platform Observability*, 9 Internet Policy Review (2020), https://policyreview.info/articles/analysis/towards-platform-observability (last visited Mar 24, 2024).
8. Paddy Leerssen, *The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems*, (2020), https://papers.ssrn.com/abstract=3544009 (last visited Mar 24, 2024).
9. Bernhard Rieder & Jeanette Hofmann, *Towards Platform Observability*, 9 Internet Policy Review (2020), https://policyreview.info/articles/analysis/towards-platform-observability (last visited Mar 24, 2024).
10. 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, https://eur-lex.europa.eu/eli/dir/2019/790/oj (last visited Mar 24, 2024).
11. Daria Dergacheva et al., *Improving Data Access for Researchers in the Digital Services Act*, (2023), https://papers.ssrn.com/abstract=4465846 (last visited Mar 24, 2024).
12. Martin Husovec, *How to Facilitate Data Access under the Digital Services Act*, (2023), https://papers.ssrn.com/abstract=4452940 (last visited Apr 23, 2024).
13. European Commission, FAQs: DSA Data Access for Researchers, https://algorithmic-transparency.ec.europa.eu/news/faqs-dsa-data-access-researchers-2023-12-13_en (last visited Mar 24, 2024).
14. John Albert, *A Guide to the Eu's New Rules for Researcher Access to Platform Data*, AlgorithmWatch (2022), https://algorithmwatch.org/en/dsa-data-access-explained/ (last visited Mar 24, 2024).

15. Paddy Leerssen et al., *Platform Ad Archives: Promises and Pitfalls*, (2019), https://papers.ssrn.com/abstract=3380409 (last visited Apr 23, 2024).

16. M. Z. van Drunen & A. Noroozian, *How to Design Data Access for Researchers: A Legal and Software Development Perspective*, 52 Computer Law & Security Review (2024), https://www.sciencedirect.com/science/article/pii/S026736492400013X (last visited Mar 24, 2024).

17. European Commission, DSA Transparency Database, https://transparency.dsa.ec.europa.eu/ (last visited Mar 24, 2024).

18. Online Platforms Terms and Conditions Database, https://platform-contracts.digital-strategy.ec.europa.eu/ (last visited Mar 24, 2024).

19. Laureline Lemoine & Mathias Vermeulen, *The Extraterritorial Implications of the Digital Services Act - DSA Observatory*, DSA Observatory (2023), https://dsa-observatory.eu/2023/11/01/the-extraterritorial-implications-of-the-digital-services-act/ (last visited Mar 24, 2024).

20. European Commission, Digital Markets Act (DMA) Legislation, https://digital-markets-act.ec.europa.eu/legislation_en.

21. Meredith Broadbent, *Implications of the Digital Markets Act for Transatlantic Cooperation*, Center for Strategic and International Studies (2021), https://www.csis.org/analysis/implications-digital-markets-act-transatlantic-cooperation (last visited Mar 24, 2024).

Read More

Verfassungsblog
Verfassungsblog is a not-for-profit academic and journalistic open access forum of debate on topical events and developments in constitutional law and politics in Germany, the emerging European constitutional space and beyond. It sees itself as an interface between the academic expert discourse on the one hand and the political public sphere on the other. Check out Verfassungsblog.de to discover all our articles, debates and other resources.

Our Books
We've got more open access books on other topics available for you at Verfassungsblog.de/Books.

Our Journal
With Verfassungsblatt, we collate a month's worth of texts that have been published on the blog into one publication. This format enables our readers to better keep an eye on which topics were important in a given month and to more easily find what interests them. Take a look at Verfassungsblog.de/Blatt.

Support Us
As a not-for-profit organisation, Verfassungsblog relies on its readers' support. You can help us keep up our work by making a donation here.

On 17 February 2024, the Digital Services Act (DSA) became fully applicable in Europe. The DSA takes a novel regulatory approach to intermediaries by imposing not only liability rules for the (user) content they host and moderate, but also separate due diligence obligations for the provider's own role and conduct in the design and functioning of their services. This new approach fundamentally reshapes the regulation and liability of platforms in Europe, and promises to have a significant impact in other jurisdictions, like the U.S., where there are persistent calls for legislative interventions to reign in the power of Big Tech. This book brings together a group of renowned European and American scholars to conduct an academic transatlantic dialogue on the potential benefits and risks of the EU's new approach.