Vying for the Scales

Content Moderation Made in Europe after One Year of DSA

Lorenzo Gradoni & Pietro Ortolani

12.03.2025 doi: 10.59704/281ae3f68d546081

Earlier this year – in an instantly famous speech – Meta CEO Mark Zuckerberg disclosed his intention to "work with President Trump to push back on governments around the world that are going after American companies and pushing to censor more". The speech singled out Europe as leading this "global trend" with its "ever-increasing number of laws institutionalizing censorship". On that occasion, Mr. Zuckerberg also announced plans to scale back Meta's efforts to moderate online content.

The vitriolic remark may also stem from the fact that the EU's Digital Services Act (DSA) has introduced a new way of challenging content moderation decisions – a way that could prove quite costly for platforms. Under Article 21 DSA any out-of-court dispute settlement (ODS) body accredited by a national Digital Services Coordinator (DSC) – in Germany this is the "Bundesnetzagentur" – may arbitrate such disputes. Although certification is granted by a national authority, it is valid throughout the EU.

To be certified, an entity must be "independent, including financially independent, of providers of online platforms" (Article 21(3)(a) DSA). Consequently, an ODS body cannot be set up as a component of a platform's complaint handling system. Crucially, the platform involved in a dispute must nonetheless bear the cost of the procedure, whereas aggrieved users have the right to litigate free of charge or at a nominal fee, which is refunded if they win. While ODS bodies' decisions are not binding, platforms are obligated to engage in the procedure in good faith. Given that they are required to pay, cooperation should be in their best interest.

Since the DSA's entry into force, six bodies have been certified by the DSCs of Malta, Germany, Hungary, Ireland, Austria and Italy. These bodies are generally limited in their scope as they only use their national language, sometimes alongside English (the Maltese Adroit, with eight working languages, constitutes an exception). While all ODS bodies are equipped to deal with content moderation disputes, only two specialise in them: User Rights, the body certified by the Bundesnetzagentur, and Appeals Centre Europe (ACE), a company incorporated in Ireland that is noteworthy in several respects.

ACE operates in six European languages but deals with content "in any language". In an interview, ACE's CEO stated that in just about two months of operation the company received "many hundreds of complaints' coming 'from every single EU country'" – a flying start that is possibly due to a level of media coverage unmatched by its competitors. ACE currently handles complaints against Facebook, TikTok and YouTube, although it is certified for disputes concerning all the so-called Very Large Online Platforms (VLOPs). Most remarkably, the company got a generous start-up grant of $15 million from the Oversight Board Trust, a Delaware-based Trust set up and funded by Meta.

Despite this, the Coimisiún na Meán, acting as the Irish DSC, certified it as "platform independent". In a previous post, we argued that the ACE-Meta nexus goes beyond the financial aspect, as individuals and corporate entities linked to Meta planned the establishment of ACE and carried it out. We also announced that further evidence about this case could emerge thanks to a Freedom of Information (FOI) request that one of us filed with the Coimisiún na Meán.

The request was partially granted, in the sense that the released file is extensively redacted. In the following, quoting from an internal document recommending the certification of ACE (on file with us), we first explain why the new evidence compounds our

concerns. We then examine whether the EU could nonetheless accept the Irish regulator's liberal reading of the independence requirement – if content moderation remains aligned with EU law and above all, the Charter of Fundamental Rights.

One year after the DSA's entry into force, the time seems ripe for an existential question: should ODS under the DSA develop as an independent, informal, decentralised system of justice, or is it acceptable for it to function as a de facto extension of platforms' complaint-handling systems? Or could it be both?

Two or three more things we know about ACE

The disclosed file shows that ACE's dispute settlement process closely resembles that of the contractors to whom Meta outsources the bulk of its content moderation. At ACE, "[d]isputes will be handled by decision-makers ("case reviewers") trained in applying terms and conditions of the platforms, using content review software similar to the tools being used by the platforms. Interestingly, the Irish regulator itself found that "[t]he procedure appears, at least partially, to replicate the content moderation processes and operations of the social media platform companies, which is well suited to reviewing high volumes of disputes in a swift manner". As in the case of outsourced moderation, decision-making at ACE is tightly hierarchical: "Market and Policy Specialists [...] will review cases of higher complexity passed to them by case reviewers for "Escalated Review" where a particular expertise is required", whereas "[t]he highest complexity cases will be escalated to Content Review Directors to carry out a 'Leadership Review'". Hierarchy promotes consistency in decision making but can also be a vehicle for influence.

Indeed, influence over ODS bodies may be exactly what platform providers are currently striving for. Since providers must bear the cost of ODS procedures, facilitating the establishment of ODS bodies that align with industry priorities – chiefly keeping costs down– is clearly an excellent idea. Notably, ACE charges platforms significantly less than current competitors, likely due to the substantial one-time grant from the Oversight Board Trust. Meta has unique experience in outsourcing content moderation. ACE could simply be an iteration of this practice sophisticated enough to seemingly fit the constraints imposed by the DSA. But does it?

Three Trustees of the Oversight Board Trust – the entity set up and funded by Meta – acted as "[t]he founding Directors of ACE". As the Irish regulator pointed out, these individuals are "Directors of ACE as individuals, not in their role as [...] Trustees". Be that as it may, these same individuals appointed the current CEO of ACE, who, at the time of the application for certification, worked for a company owned by the above-mentioned Meta-funded Trust. In short, ACE is led by individuals either from a Meta-backed Trust or who have Meta's trust.

In its analysis under the independence requirement, the Irish regulator dismissed these concerns, by observing that the founding Directors "will serve at ACE as Non-Executive Directors and thus will not be involved with individual case decision-making". However, their "legal responsibilities", which "include setting the company strategy, risk management, assessing and monitoring performance and culture", do not seem insignificant for the purposes of assessing independence. The Irish regulator must have thought the same as it required ACE – as a prerequisite for certification – to appoint within six months four additional board members with no direct or indirect links to Meta or other platforms (in this the record confirms our assumptions).

The Irish regulator calls these additional members "non-relationship Directors" and set their number at four so that they have a majority over the relationship Directors (as one may call them). However, the Coimisiún na Meán accepted that ACE's statutes set "the quorum necessary for a general meeting" of ACE's Board at "a Super Majority", defined as 66% of the Board, rounded up to the nearest whole number of Directors. Based on the information we currently have, this means five out of seven members, which effectively gives the relationship Directors a sort of veto, as the quorum cannot be reached unless at least one of them participates. Furthermore, the Irish authorities did not object to the three founding Directors acting

alone in appointing the CEO, who inter alia has the power to dismiss case reviewers on such grounds as "ongoing poor performance".

Independence - what sort?

In the light of the above, it would be difficult to argue that ACE "presents an appearance of independence" within the meaning of the ECJ's case law on the independence of courts (see, e.g., C-585/18, A.K., para. 127). However, as Wilman, Kalėda and Loewenthal rightly point out, "the level of independence required from certified out-of-court dispute resolution bodies may not be the same as in the case of courts". What does it take for an ODS body to be seen as "independent" under Article 21(3)(a) DSA?

European institutions, if called upon to assess ACE's certification, may perceive more breaks than junctions in the connections from California and Delaware to Dublin. They may also consider it a good thing that wealthy platform operators, on top of bearing procedural costs, invest in a trust that helps building an efficient ODS infrastructure, if that means improving the quality of content moderation decisions. And this is something that ACE will likely bring to the table. In this policy context, the appropriate legal yardstick might well be, by analogy, the independence requirements of the Directive on alternative dispute resolution (ADR) for consumers disputes, rather than judicial independence. Article 6(4) of the Directive allows that "persons in charge of ADR are employed or remunerated exclusively by a professional organisation or a business association of which the trader is a member", provided that "they have a separate and dedicated budget at their disposal which is sufficient to fulfil their task."

However, this analogy risks breaking down as ACE is linked (through the Oversight Board Trust) to a single platform provider – one "trader" rather than an "association" or "organisation". And since ACE's jurisdiction extends beyond Meta's platforms – as already noted, it currently also covers YouTube and TikTok and "intends to expand further" – one must ask whether the independence requirement under the DSA protects only users or also platform providers from undue influence by competitors. Providers other than Meta may appreciate the low-cost service offered by ACE, but not if it comes at the expense of their autonomy, potentially affecting competitive dynamics. Other providers may accept Meta's indirect hegemony in this field, if it operates in their interest. However, should it fail to do so, any of them could challenge ACE's accreditation or set up its "own" ODS facility. Can one rule out the risk that, following the Irish precedent, a national regulator might approve an ODS facility designed by Mr. Musk's legal advisors?

Building on users informed choices

Be that as it may, ODS is destined to remain a multi-operator space for the time being, with partially overlapping and thus competing jurisdictions. One aspect that users should consider when choosing an ODS is the extent to which competing facilities ensure the protection of fundamental rights. Such protection is indeed the capstone of the "safe, predictable and trusted online environment" that the DSA aims to create (Article 1(1) DSA). Particularly at a time when Meta expresses impatience with rights whose protection could conflict with freedom of expression, it is concerning that ACE's current provision on "applicable rules" seems to give human rights a limited or uncertain role.

According to this provision, a "case reviewer" decides solely based on "the platform's content policies". A so-called "Normative Framework", which incorporates "fundamental rights standards", applies only if the first reviewer refers the dispute to an "escalation reviewer". But – notice – this framework applies only insofar as such standards inform 'the platform's stated values, principles and policy exceptions. What if the platform's internal values, principles and policy exceptions are not (or no longer) so "informed"? And what if the case is not escalated at all? In such instances, ACE's decision should not prevent a user from bringing their human rights complaint before another ODS body. Article 21(2) DSA, which allows a platform provider to 'refuse to engage' in ODS twice over the same dispute, would arguably not be an obsta-

cle, since the complaint would rest on different legal grounds.

User Rights, the ODS body certified by the Bundesnetzagentur, takes a different approach to defining applicable standards. Its decisions are said to be "based on European and, where relevant, German law, considering freedom of expression and other fundamental rights". User Rights also claims to provide, "[l]ike no other out-of-court dispute settlement body, [...] well-reasoned decisions accounting for fundamental rights." Having seen a couple of those decisions (anonymised), we can confirm that they indeed resemble short arbitral awards or judgments. This got us thinking about the modus operandi of Meta's Oversight Board, which mimics that of a human rights body and prioritises internationally recognised human rights over Meta's community standards.

Interestingly, one of the co-founders of User Rights used to work at the Oversight Board. In the realm of ODS, most roads seem to run from Meta. However, while User Rights (whose independence is not in question) embraces the Oversight Board's "judicial approach", and strives to make it work at scale, ACE arguably replicates, at least for non-escalated cases, the bureaucratic-corporate template suspicious of human rights (and of law tout court) that has prevailed since the early days of the mind-boggling enterprise that is content moderation. The emerging ODS landscape may be spacious enough to accommodate both approaches. What matters is that users are put in a position to choose from competing facilities in full awareness of the different options.

In the past, ACE had taken commendable steps in this regard. The original version on its website contained clear hints of the company's financial and corporate ties. It even displayed the Facebook logo (presumably as a synecdoche for social media), whereas the current version has been stripped of any reference to Meta. The average user is therefore unlikely to be aware of such relationships, potentially preventing informed user choice from being a driving force in shaping the ODS landscape.

ABOUT THE AUTHORS

Lorenzo Gradoni
Lorenzo Gradoni holds a PhD in European Union Law (Bologna). He is currently a member of the Luxembourg Centre for European Law (LCEL).Lorenzo Gradoni holds a PhD in European Union Law (Bologna). Formerly Associate Professor of International Law at the University of Bologna and Senior Research Fellow at the Max Planck Institute Luxembourg for Procedural Law, where he also served as the Chief Scientific Advisor to the Max Planck Encyclopedia of International Procedural Law. He currently works at the University of Luxembourg.

Pietro Ortolani
Pietro Ortolani is a full professor of digital conflict resolution at Radboud University in Nijmegen, the Netherlands.

SUGGESTED CITATION

Gradoni, Lorenzo; Ortolani, Pietro: Vying for the Scales: Content Moderation Made in Europe after One Year of DSA , Verfassungsblog, 12.03.2025, doi: 10.59704/281ae3f68d546081.

Licensed under CC BY-SA 4.0.

Langzeitarchiviert bei IntRechtDok unter: doi.org/10.17176/20250313-001130-0.