How can we apply human rights due diligence standards to content moderation?

Focus on the EU Digital Services Act

Summary

On Thursday, July 29, representatives from government, the private sector, academia, and civil society came together to discuss the EU Digital Services Act (DSA) and its obligations around human rights due diligence in a virtual event hosted by the Global Network Initiative (GNI) and the Center for Democracy and Technology (CDT), Europe. The meeting was held under the Chatham House Rule.

Published in December 2020, the DSA proposal is framed as an update of the EU's E-Commerce Directive. Its intent is to create a safer digital environment and a level-playing field in the European single market and globally. In practice, the DSA will likely establish new rules for addressing illegal content, responsibilities for intermediaries, and protection of users' fundamental rights, among other fundamental aspects of the current digital environment. The draft DSA establishes differentiated obligations for companies, depending on their services and size in the EU. The focus of this event was on the additional set of rules related to "risk assessment" and assigned to very large online platforms (VLOPs), including around identifying and analyzing systemic risks on their platforms, adopting mitigation measures, and undergoing annual independent audits.

The discussion highlighted recent developments in the emerging field of human rights due diligence, in the technology space and beyond, and how it has been shaped by the UN Guiding Principles on Human Rights and Business. Several participants praised the overall architecture and approach set forward in the DSA, positively contrasting the focus on content moderation processes, transparency, and accountability, but there was also agreement that it still leaves many unresolved questions. Some core concerns included the convergence and divergences between the principle-based structure of the UNGPs vs. the DSA's rule-based approach, as well as how to strike the appropriate balance between establishing legal clarity and allowing the right amount of flexibility for tech companies to assess their unique risks.

Overall, the session revealed that the DSA still has significant room for improvement in terms of clarifying its due diligence obligations, and anticipating their unintended consequences. This discussion helpfully illuminated important questions that EU legislators should be addressing as they continue to revise the DSA and consult stakeholders over the next few years.

Speaker affiliations:

Civil Society: Article 19, CDT, GNI
Consulting: Business for Social Responsibility (BSR)
Government: German Federal Ministry for Economic Affairs and Energy, Danish Institute for Human Rights
Multilateral: UN B-tech Project
Academia: University of Ottawa, Canadian Internet Policy and Public Interest Clinic
Private Sector: Microsoft, Mozilla

Participant affiliations:

European Parliament: 2
EU Member States: 1
European Commission: 1
Other Government/Multilateral: 1
Academia: 2
Civil Society: 10
Private Sector: 10

Table of Contents:

Session I: Current state of play of human rights due diligence in the tech policy sector 1

Session II: Deep dive on the Digital Services Act proposal; discussion about possible amendments to the current text to mitigate human rights risks 4

Session III: Deep dive on DSA's provisions about auditing of online platforms 5

Shared Resources **7**

Session I: Current state of play of human rights due diligence in the tech policy sector

Current BHR and Tech Initiatives

The initial part of the conversation highlighted some of the existing business and human rights initiatives, including work conducted by the Danish Institute for Human Rights, GNI, and the UN B-tech project (see 'shared resources'). The discussion then turned to some of the tensions between the UN Guiding Principles on Business and Human Rights (from here on referred to as the "UNGPs"), which have served as the guiding framework for many existing initiatives, versus the rules-based approach of the DSA.

Potential divergences between the DSA and the UNGPs

The UNGPs and the DSA have some differently defined provisions, which could potentially cause confusion for companies. Most notable is their divergence on the definition of due diligence. That term has taken on a particular meaning under the UNGPs, which focuses on the identification and mitigation of human rights risks. This is the foundation on which much work has already been done through voluntary company efforts, multistakeholder initiatives like the GNI, and emerging regulatory regimes requiring mandatory human rights due diligence (including a parallel effort currently underway in the EU). Meanwhile, the draft DSA takes a much broader and more ambiguous approach, which risks creating inconsistent expectations and approaches.

In addition, in its risk assessment provisions, the draft DSA appears to privilege some rights over others, including privacy, freedom of expression, and access to information, while the UNGPs focus on the holistic range of international human rights. There is concern that the DSA approach could lead companies to ignore other harms that they should be addressing.

Prioritization was identified as another issue. Specifically, the draft DSA appears to prioritize discouraging and limiting "illegal content" over countervailing concerns about limiting freedom of expression or privacy, and otherwise lacks clarity about when and where companies should act. By contrast, the UNGPs recognize that businesses cannot address all risks at once and enables them to prioritize based on the severity of the harm and the likelihood of that harm occurring. This is particularly important for the large tech companies (i.e. "VLOPs" in the DSA), which must process a significant amount of content.

A speaker also noted possible incompatibility between the auditing requirements outlined in the DSA with the UNGPs. Establishing a clear set of standards that can be applied to all types of companies in the tech sector is a challenge, which is why companies have preferred conducting assessments vs audits. While certain principles may be applied, requirements will still need to differ according to companies' business models. The UNGPs offers this flexibility and encourages

companies to take appropriate action according to the risks they identify. The DSA can also provide further clarity about what it means by publishing strategies and impact assessments. Disclosure should be meaningful, and not just be implemented for their own sake.

Aligning the DSA and the UNGPs

Going forward, regulators should think about how the DSA can complement the UNGPs. For instance, the DSA could strengthen stakeholder engagement, focus on the salient risks versus imposing a one-size fits all approach, and adopt a complimentary definition of due diligence. On this last point, it is worth noting that the DSA's provisions on due diligence already overlap with the UNGPs approach in several key manners, including assessing actual and potential human rights impacts, acting on and integrating the assessments within company operations, tracking process, and communicating results publicly.

Avoid a "Check the Box" Exercise

The speakers agreed that it's important for the DSA to not result in "check the box" exercises. Companies need to foster meaningful relationships with relevant stakeholders, as challenging as it can be to identify the specific rights-holders, and devise collaborative solutions. Moreover, company executives need to think critically about the oversight they want to create and ensure there is a company culture that respects human rights across the value chain. It remains to be seen how this dynamic will play out once stakeholder engagement becomes an obligation under EU regulation.

Extraterritoriality

A brief note was made about the DSA's jurisdictional approach. Specifically, content that is considered illegal in the EU and is removed from a platform, will remain up in other jurisdictions. Policymakers should not avoid this issue, but rather anticipate and proactively address concerns around extraterritorial application, conflicts of laws, and comity.

Session II: Deep dive on the Digital Services Act proposal; discussion about possible amendments to the current text to mitigate human rights risks

Lack of Definitional Clarity and Legal Certainty

The speakers discussed some of the issues with Articles 26 and 27 of the DSA, including their implications for legal certainty and the rule of law.[1] It is not clear, for instance, what the proposal means by "systemic risk" and when it is reached, despite being a central component of the legislation. Additionally, there are different types of risks related to preventing the dissemination of illegal content. Other ambiguous terms include the "manipulation of a service" and "inauthentic use." "Civic discourse" is another potentially broad term that could encompass all of human discussion and debate, leading one civil society participant to urge its removal.

Rather than clarifying the above concerns, many of the proposed amendments to the DSA also present new uncertainties. For one, making immunity dependent on compliance with the due diligence requirements creates an uncertain legal environment for companies, and could have problematic consequences for privacy and freedom of expression. Additionally, another amendment that proposes adding risk assessments to a company's Terms of Service, as seen from one speaker's perspective, effectively demands risk assessments for "legal, but harmful" content, thereby creating more uncertainties, given problematic definitions around "fake news," "disinformation," and other types of speech that fall under this category.[2] In this light, the DSA can do more to ensure that intermediary's obligations are proportionate to actual risk.

Value Tensions

Problems also arise around the definition of the risk mitigation measures, and how they deal with tensions between different rights. A civil society representative noted that protecting the rights of children might conflict with encryption and privacy. That said, the DSA does not provide guidance on how to resolve these conflicting interests, or how a company can comply with these obligations. Several speakers spoke about the importance for EU legislation to balance these "value tensions."

---

[1] **Article 26** requires VLOPs to identify, analyze, and assess any significant systemic risks, including the (a) dissemination of illegal content, (b) any negative effects for the exercise of fundamental rights (privacy, child rights, and freedom of expression), and (c) intentional manipulation of their service, including by means of inauthentic use or automated exploitation of the service, with an actual foreseeable negative effect on the protection of public health, minors, civic discourse, or actual or foreseeable effects related to electoral processes and public security. **Article 27** requires VLOPs to put in place "reasonable, proportionate, and effective mitigation measures, tailored to the specific systemic risks identified pursuant to Article 26."

[2] Along these lines, the process by which a company handles content requests through its Terms of Service vs. national legislation is also a point of contention for some European governments. It was also mentioned that platforms currently have discretion to make decisions that significantly impact society, which they should not, in fact, have the power to do.

Enumerated or Broader Mention of Rights

In the DSA's risk assessment provisions, there is an enumeration of particular rights. The speakers wrestled with the appropriate balance between clearly defined protections versus flexibility for companies, given their unique business models and societal impacts. This sparked further discussion around the UNGPs vs. the DSA. Traditional human rights standards generally have states as their primary focus and precisely outline allowable restrictions and their associated legal basis. Meanwhile, the UNGPswere developed to guide businesses but also reinforce the necessity that any state regulation follows a human rights-based approach.

Broad range of companies

Risk assessments will need to be flexible enough to account for different business models and services, which the DSA does not currently reflect. For instance, while Art.26(a) concerning illegal content applies to all types of platforms, Art 26(b) and 26(c) only work for platforms that host content.

There is also a difference between platforms with "interaction" vs. "transaction" functionalities, and their impacts on fundamental rights. A company that blocks illegal goods in a marketplace presents different risks, versus a platform that blocks the speech of politicians.

Remedy

It is important for the DSA to carefully incorporate remedy mechanisms, and for these to be consistent with and complementary to the auditing process,. Civil society should have ways to be involved in the process, as regulators do not necessarily have as deep an understanding of the human rights issues.

Session III: Deep dive on DSA's provisions about auditing of online platforms

Mandatory auditing regime - positive next step

The speakers in this session expressed approval of the EU's overall decision to mandate audits and assessments through the DSA, as it will provide a legal basis for the work that companies, CSOs, and third-party auditors have already been doing worldwide, as well as add a regulatory architecture to the third-party auditing industry.

The rest of the discussion focused on the ambiguities around the DSA's auditing provisions, and the questions that need to be further addressed. For instance, while Article 28 sets out a very basic auditing requirement, it does not specify how audits should be conducted or by whom, nor how the expertise and independence of the evaluators will be assessed. In this light, the speakers highlighted aspects that need further thinking/clarification, including: the necessary level of granularity of audits; the nature of quality assurance, standards, and oversight mechanisms; and the challenges auditors will face.

Establishing greater granularity

There was a consensus among the speakers that auditing regimes need to be more granular if they are going to be effectively implemented and complied with.

That said, some of the auditing requirements are easier to conduct than others. For instance, transparency reporting has relatively clear standards, compared to assessing a complaints handling mechanism. There are similar uncertainties around auditing an algorithm versus its outputs, and these require regulators to specify what fundamental interests they are trying to achieve. The auditor is also required to provide an opinion, but it is not clear on what basis they will provide that opinion nor how detailed that information must be.

Developing Standards

If enacted as drafted, the DSA will require certain standards and certification processes to be developed, which Article 34 explicitly calls for. There is already precedent in this area of work by organizations like GNI and the Global Reporting Initiative, which should be acknowledged and built upon. Ideally, countries would strive to align their global accountability and transparency standards and allow for mutual recognition. These standards will take time to develop and the DSA will need to allow corresponding flexibility with respect to implementation.

At the same time, one speaker cautioned against creating a structure that is too rigid, thereby preventing the industry from emerging at all. If the standards are too burdensome, there is a

concern that this may only encourage a few, dominant third-party auditors, versus a plurality of companies.

The primary goal is to ensure that the platform auditing industry develops in a way that is trustworthy and in line with standards. The DSA will not get everything right in the first instance. In fact, GNI offers a compelling model, in that their "assessment" process does not define compliance as perfection, but rather, good faith implementation with improvement over time. This kind of structure could encourage conversations around the outputs and give stakeholders an opportunity to align regulation and improve compliance.

Assessing effectiveness and oversight

There needs to be robust oversight of the auditing itself. In other sectors, there are mechanisms to ensure that auditors are not incentivized to simply give the company a stamp of approval. Sometimes, this is accomplished through regulatory sanctions, while in other cases, shareholders can sue an auditor for producing a flawed report.

On that note, the DSA should clarify how effectiveness will be measured, if at all. It is currently unclear what is necessary to establish effectiveness. WIthout further clarity, some companies will simply "check the box." One speaker recommended that it should be mandatory for auditors to assess the quality of platforms risk assessments and mitigation measures.

Challenges facing auditors

When the issue of how to ensure the auditors' independence was raised, the speakers did not express a high level of concern, citing precedence in other sectors and industries. In addition to the scrutiny that comes with the auditing process, independence may also be assured given that cooking the audit would reflect poorly on a company. Rather, more significant challenges include the auditors' potentially large workload and need for relevant expertise, as well as ensuring a joint approach between the European Commission, the DSCs, third party researchers, and auditors.

The auditors will have to operate in a rapidly evolving and changing regulatory space. There are many EU and Member State regulations, particularly regarding platforms and digital services, that will come into play. Here, coordination will be key (though harmonization across the EU will be a challenge), as well as greater leniency in the auditing process for companies to get the "harder" questions wrong without significant penalties.

Resources Mentioned/Shared:

Human rights impact assessment of digital activities (Danish Institute for Human Rights, November 2020)

EU Mandatory Human Rights Due Diligence Directive: Recommendations for the European Commission (United Nations Human Rights, July 2021)

A Human Rights-Based Approach to Content Governance (BSR, March 2021)

Cross-Cutting Stakeholder Engagement: Guidance on HIRA of Digital Activities (Danish Institute of Human Rights, 2020)

Mandatory Due Diligence (Business & Human Rights Resource Centre, Updated Sept 2021)

Designing and implementing effective company-based grievance mechanisms (B-Tech, Jan 2021)

Bridging governance gaps in the age of technology – key characteristics of the state duty to protect (B-Tech, May 2021)

Civil Society Initiative Urging LIBE Committee to Uphold Key Principles in the DSA (AccessNow, Article 19, CDT, EFF, June 2021)

Content Regulation and Human Rights Brief (Global Network Initiative, 2020)

Federal Court of Justice on claims against the provider of a social network who deleted posts and blocked accounts on charges of "hate speech" (Federal Court of Justice, July 2021)

GNI's Assessment Process (Global Network Initiative)

Mozilla's position paper on the EU Digital Services Act (Mozilla, May 2021)