Digital Policy Lab

Policy Digest #7

12 October 2022

Sponsored by

ISD Powering solutions to extremism, hate and disinformation

Federal Republic of Germany Foreign Office

Policy Digests offer an overview of recent digital policy developments in Digital Policy Lab (DPL) member countries, including regulatory and non-regulatory initiatives aiming to combat online harms such as disinformation, hate speech, extremist or terrorist content. In addition to general updates, each Policy Digest provides a snapshot of topic-specific schemes relevant to the upcoming DPL session.[1]

Section 1 Digital policy developments

EU: Digital Services Act (DSA) and Digital Markets Act (DMA)

Type Regulatory (legislative acts)
Status Approved

On 4 October 2022, the Council of the European Union gave its final approval of the Digital Services Act (DSA), which defines responsibilities and due diligence obligations for providers of intermediary services, including very large online platforms (VLOPs) and very large online search engines (VLOSEs). The DSA introduces measures to counter illegal content online, imposes limits on the presentation of advertising and on the use of sensitive personal data for targeted advertising, and prohibits misleading interfaces. VLOPs and VLOSEs will have to offer recommender systems that are not based on profiling, and assess and mitigate the systemic risks they pose (meaning various risks, for example related to the dissemination of illegal content, negative effects on fundamental rights, on electoral processes and on gender-based violence or mental health). The final signature by the European Parliament and the Council is expected on 19 October. The text will be published in the Official Journal of the EU and enter into force 20 days after its publication. It will apply 15 months after entering into force, while the application to VLOPs/VLOSEs will be 4 months after their designation as such.

On 14 September 2022, the Czech EU affairs minister, Mikuláš Bek, and the President of the European Parliament, Roberta Metsola, signed the Digital Markets Act (DMA). The DMA defines rules for large online platforms ("gatekeepers") to ensure a digital level playing field. For example, such platforms will have to ensure that unsubscribing from core platform services is just as easy as subscribing; ensure that the basic functionalities of instant messaging services are interoperable; give business users access to their marketing or advertising performance data on the platform; and inform the European Commission of their acquisitions and mergers. The Commission will also be the sole enforcer of the rules, and will cooperate closely with the competition authorities and courts in the EU Member States. In case of non-compliance, the Commission will be able to impose fines of up to 10% of the gatekeeper's total annual turnover, or 20% in cases of repeated non-compliance. The DMA will be published in the Official Journal of the EU on 13 October 2022 and will start to apply six months later.

EU: GDPR fine of 405 million euros for Instagram following EDPB intervention

Type Regulatory (enforcement)
Status Adopted

On 2 September 2022, following the European Data Protection Board's (EDPB) binding dispute resolution decision in July 2022, the Irish Data Protection Authority (DPA) adopted its final decision regarding Instagram, issuing a record General Data Protection Regulation (GDPR) fine of 405 million euros. The decision follows an inquiry into Instagram's public disclosure of email addresses and/or phone numbers of children using the Instagram business account feature and a public-by-default setting for personal accounts. This was the first binding decision of the EDPB addressing one of the fundamental pillars of the GDPR: the lawfulness of processing in accordance with Article 6. The EDPB provided clarification on the applicability of the legal bases of 'performance of contract' and 'legitimate interest'. The dispute resolution found that there were no grounds to conclude that the processing at stake was necessary for the performance of a contract or, if it were to be

[1] We welcome any feedback from DPL members regarding additional developments, as well as own submissions from DPL members who wish to be featured in the digest.

considered necessary, did not pass the balancing test required when determining legitimate interest. The final decision taken by the Irish DPA is available in the Register for Decisions taken by supervisory authorities and courts on issues handled in the consistency mechanism.

New Zealand: Christchurch Call Initiative on Algorithmic Outcomes

Type Non-regulatory (initiative)
Status Launched

On 22 September 2022, New Zealand and the US, on behalf of the wider Christchurch Call community, announced the launch of an initiative that will support the creation of new technology to understand the impacts of algorithms on user experiences. New Zealand, the US, Twitter, and Microsoft will invest in developing technologies to enable data scientists to remotely study data and algorithms distributed across multiple secure sites. Technologies such as remote execution, federated learning, differential privacy, and secure multi-party computation will enable this research in a way that conforms to the data-use policies of the platform under study. The current project plan envisages implementation approximately over a nine-month timeframe with a total cost of approximately 1.5 million USD. Milestones include building the underlying software and systems for the new infrastructure, conducting a proof-of-concept test (will be provided by Twitter's machine learning (ML) Ethics, Transparency and Accountability team) using synthetic data, while trusted partners test out the system on real datasets to demonstrate reproducibility and proof of function.

UK: Online Safety Bill (OSB)

Type Regulatory (legislative act)
Status Review process

On 7 September 2022, the new UK Prime Minister Liz Truss confirmed in a statement to Parliament that her government would proceed with the OSB, albeit with "some tweaks". In March 2022, the government had formally introduced the amended OSB into Parliament. The bill would apply to "user-to-user services" and "search services", requiring such services to identify, remove and limit the spread of illegal content. Secretary of State of the Department for Digital, Culture, Media and Sport (DCMS) Michelle Donelan confirmed that changes would be made to the adult safety duties that cover so-called 'legal but harmful' content. As a result, the timeline for the bill to return to Parliament is not yet clear. Ofcom's "Roadmap to regulation" had confirmed that under the most recent version of the bill services would be able to decide whether to host content that is legal but harmful to adults, and Ofcom would not compel them to remove it. "Category 1" services with the highest risk functionalities and the highest user-to-user reach would be required assess risks associated with certain types of legal content that may be harmful to adults, have clear terms of service explaining how they treat such content, and apply those terms consistently.

US: Principles for Enhancing Competition and Tech Platform Accountability

Type Non-regulatory (principles)
Status Published

On 8 September 2022, the White House convened a listening session with experts and practitioners on the harms that tech platforms cause and the need for greater accountability. Experts and practitioners identified concerns in six key areas: competition; privacy; youth mental health; misinformation and disinformation; illegal and abusive conduct, including sexual exploitation; and algorithmic discrimination and lack of transparency. Following the session, the White House released "Principles for Enhancing Competition and Tech Platform Accountability". Embracing bipartisan Congressional efforts to pass antitrust and privacy legislation, the principles call for more competition in the technology sector and federal protections for Americans' privacy. The principles advocate for stronger privacy and online protections for children, including prioritising safety by design standards and practices for online platforms, products and services. In addition to transparency from platforms on their algorithms and content moderation policies, as well as ending discriminatory algorithmic decision-making, the principles call for reforming Section 230 Communications Decency Act that broadly shields companies from liability for the content on their platforms.

US: California social media transparency bill (AB-587)

Type Regulatory (legislative act)
Status Signed

On 14 September 2022, California Gov. Gavin Newsom announced he signed a "first-of-its-kind" bill into law designed to "protect Californians from hate and disinformation spread online." The bill (AB-587) by Assemblymember Jesse Gabriel (D-Encino) will require social media companies to publicly post their terms of services regarding hate speech, disinformation, harassment and extremism on their platforms, and report data on their enforcement. Companies are expected to submit a service report to the Californian Attorney General on a semiannual basis, in which they provide detailed descriptions of moderation practices (for example, when these systems involve human review) and information on flagged content. Companies must submit their first terms of service report to the Attorney General no later than 1 January 2024. The rules will apply to social media companies that generate more than 100 million USD in gross revenue during the preceding calendar year. Platforms that fail to abide with the new rules shall be liable for penalties of up to 15,000 USD per violation per day. The bill has raised concerns from legal experts, citing the First Amendment and free speech as their main issues.

Section 2 Topic-specific snapshot: "Operationalising data access requirements"

This section presents summaries of selected analyses and commentary published by civil society and academia on current methodological challenges of conducting social media research, including the prospects of sharing privacy-protected platform data with external researchers.

In their perspective article "Acknowledging Data Work in the Social Media Research Lifecycle", **Katharina E. Kinder-Kurlanda** and **Katrin Weller** address the ongoing discussions of social media research data's quality and validity. The article considers epistemological challenges and drivers of social media research as an ongoing, situated process of engaging with technology-enabled structures and affordances of tools and platforms. It addresses criticism of social media research, notably for being data-driven with research questions that are tailored to data availability and accessibility.

When working with social media data, refining initial research questions and refining data collection strategies is likely necessary if the "ideal" dataset turns out not to be accessible, and iterations are needed to define a question and look for suitable data. However, the authors find that a lack of documentation of potentially discarded ideas combined with greater innovation in methods and lack of standard epistemologies contribute to a notion of opportunism. Moreover, the authors find that there are currently no tools or standards for documenting the data acquisition process that allow, for example, to describe the rationale for choosing specific keywords or collection periods or to record other critical information, such as server downtimes during the data collection phase.

Preparation and analysis of data (e.g., preprocessing, cleaning, labeling, sorting, and filtering) often face limitations, e.g., tools for detecting sentiments in texts are limited in accuracy. However, the authors find that some researchers were concerned that studies that they themselves saw to be limited in scope and analytic value due to the limitations of both data and tools were perceived as much more general and powerful by the media or even other researchers. Working toward a shared epistemic understanding of data and tools, the authors suggest approaches such as synthetic data or sandboxes as safe spaces for data work with datasets specifically prepared for experimental purposes.

Finally, the authors note that social media data preservation and sharing is difficult to accomplish given the lack of documentation tools, the enhanced importance of privacy protection and the restrictions on sharing imposed by platform providers. The lack of consensus about what information is required to achieve reproducibility (if that is the goal) or reusability adds uncertainty to developing documentation tools and defining archiving standards.

In his testimony before the US Senate Committee on the Judiciary, **Nate Persily**, James B. McClatchy Professor of Law and Co-Director of the Stanford Cyber Policy Center, emphasises the foundational role that transparency can play in a larger regulatory regime. Persily notes that greater transparency will affect behaviour of the platforms themselves, as well as contribute to effective public policy. Data access would be essential for a better understanding of the role that platforms and their systems play in amplifying problematic content and exacerbating online threats such as the public health mis- and disinformation, or foreign election interference. In short, Persily argues that transparency is a necessary prerequisite to understanding most contemporary policy challenges "in the real world".

Persily summarises that transparency legislation requires (1) broad obligations for public disclosures; (2) protection for researchers analysing publicly available data; and (3) supervised access for vetted researchers to the data accessible to the platforms own data scientists. Importantly, policymakers and researchers should not depend on the generosity of platforms for analysis of data in the public interest. This was a motivation for the creation of Social Science One, run by academics for academics, and independently funded by a set of foundations. Researchers were vetted by the Social Science Research Council, while Facebook's main role was to provide datasets for analysis by these vetted academics.

According to Persily, Social Science One faced two sets of problems. The first and most significant were concerns over protecting privacy. The Cambridge Analytica scandal helped to reveal how much data is collected by platforms, and potentially accessible to third-parties, but has since served to demonstrate the need for the establishment of a legally sanctioned and regulated process that will simultaneously grant researcher access while ensuring government oversight to protect user privacy. The pressures regarding user privacy constrained Social Science One as it tried to realise broad data access. For example, Facebook added statistical noise, following the principles of differential privacy, even though the data were aggregated at the URL level (that is, no individual level data appeared in the dataset) and only URLs that were shared 100 times were included. The second challenge that Social Science One confronted concerned the reliability of the data. Facebook data scientists discovered that the URLs dataset neglected to include about a third of the US population in the available data. Persily emphasises that the life cycle of Social Science One highlights how critical it is to do as much as possible to protect and safeguard user privacy in the design of new data sharing initiatives.

In their research article "It's Time to Open the Black Box of Social Media", **Renée DiResta**, **Laura Edelson**, **Brendan Nyhan**, and **Ethan Zuckerman** outline what type of research could be enabled by data access and platform transparency. The research questions below were formulated by the authors:

- Research suggests that misinformation is often more engaging than other types of content. Why is this the case? What features of misinformation are most associated with heightened user engagement and virality? Researchers have proposed that novelty and emotionality are key factors, but we need more research to know whether this is true. A better understanding of why misinformation is so engaging will help platforms improve their algorithms and recommend misinformation less often.
- Research shows that the delivery-optimisation techniques companies use to maximize revenue, and even the advertising-delivery algorithms themselves, can be discriminatory. Are some groups of users significantly more likely than others to see potentially harmful adverts, such as consumer scams? Are others less likely to be shown useful adverts, such as job postings? How can advertising networks improve delivery and optimisation to be less discriminatory?
- Social media companies attempt to combat misinformation by labelling content of questionable provenance, hoping to push users toward more accurate information. Results from survey experiments show that the effects of labels on beliefs and behaviour are mixed. We need to learn more about whether labels are effective when individuals encounter them on platforms. Do labels reduce the spread of misinformation or attract attention to posts that users might otherwise ignore? Do people start to ignore labels as they become more familiar?
- Internal studies at Twitter show that Twitter's algorithms amplify right-leaning politicians and political news sources more than left-leaning accounts in six of seven countries studied. Do other algorithms used by other social media platforms show systemic political bias as well?
- Because of the central role they now play in public discourse, platforms have a great deal of power over who can speak. Minority groups sometimes feel their views are silenced online as a consequence of platform moderation decisions. Do decisions about what content is allowed on a platform affect some groups disproportionately? Are platforms allowing some users to silence others through the misuse of moderation tools or through systemic harassment designed to silence certain viewpoints?

The authors highlight that researchers require access to data on the structures of social media, such as platform features and algorithms, so researchers can analyse how systems shape the spread of information and affect user behaviour.

In his commentary "Researcher Access to Platform Data: European Developments", **Mathias Vermeulen**, Public Policy Director at AWO, examines the scope of the proposed data-sharing regime under the Digital Services Act (DSA). Vermeulen's commentary also evaluates the guidance and commitments contained in the EU's strengthened Code of Practice on Disinformation (CoPD) and the European Digital Media Observatory's (EDMO) proposed Code of Conduct, including how these instruments relate to one another and operate within the broader regime.

First, the DSA requires Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSE), covering companies that have more than 45 million monthly active users in the EU, regardless of their size or turnover to provide data to researchers on request from a regulator. In the DSA context, Vermeulen argues, researchers can be considered quasi-auditors: they can assess emerging risks that may not have been covered by a platform's internal risk assessment report, and assess whether self-regulatory mitigation measures have been effective in practice. Details on the procedures for vetting researchers and providing access to data will be specified in a 'delegated act'—a secondary piece of EU legislation that will be adopted in the next 18 months. The DSA lists a number of conditions that researchers need to fulfil to be vetted, and outlines a five-step process for requesting platform data.

Second, a number of tech companies (Google, YouTube, Twitter, Microsoft Bing, LinkedIn, Meta, Instagram, and TikTok) commit to make data available to enable research on disinformation under the EU's Code of Practice on Disinformation (CoPD). In practice, the Commission will consider adherence to these commitments when assessing compliance with the DSA's obligations, specifically on risk mitigation measures. By proactively giving researchers access to data, a company can signal to a regulator that it approaches its due diligence obligations under DSA Article 34 seriously. Still, the DSA makes clear that participating in and implementing this code "should not in itself presume compliance". Importantly, under the Code, the ultimate purpose of the access to data regime is different from that of the DSA, as access can be granted for any research purpose on "disinformation," and is not limited to assessing platforms' roles as they address risks or take risk mitigation measures. Moreover, relevant signatories commit to "developing, funding, and cooperating with an independent, third-party body that can vet researchers and research proposals". Vermeulen asserts that this body could be seen as the same "independent advisory mechanism" mentioned in the DSA, and elaborated in the EDMO report.

The third initiative, led by EDMO, developed a draft Code of Conduct under the EU's General Data Protection Regulation (GDPR), which specifies how platform-to-researcher data access can be achieved in compliance with Europe's most stringent privacy regime. Vermeulen notes that, from a GDPR perspective, it does not matter who performs the research using platform data. The main consideration would be that a researcher/organisation is equipped to properly protect the data it receives and processes. The EDMO Working Group unanimously agreed that an independent intermediary body should be created to certify that research proposals and proposed data safeguards comply with the EDMO Code of Conduct. Streamlining these review and certification processes and housing them in an independent intermediary body would reduce the burdens placed on smaller, under-resourced universities and research institutions, thereby offering data access to a much more diverse pool of researchers. Such a body could simultaneously review and certify that the platforms' datasets, codebooks, and technical systems adhere to the EDMO Code requirements.

About the Digital Policy Lab
The Digital Policy Lab (DPL) is an inter-governmental working group focused on charting the regulatory and policy path forward to prevent and counter disinformation, hate speech, extremism and terrorism online. It is comprised of a core group of senior representatives of relevant ministries and regulators from key liberal democratic countries. The DPL aims to foster inter-governmental exchange, provide policymakers with access to sector-leading expertise and research, and build an international community of policy practice around key regulatory challenges in the digital policy space. We thank the German Federal Foreign Office for their support for this project.