Journal of the Statistical and Social Inquiry Society of Ireland
Vol. LIV

Symposium: Big Tech and Big Data - New Frontiers in Regulation

Data Access for Research: A New Era in Online Integrity

Naoise McNally

University College Dublin

(read before the Society, 30th April 2025)

Abstract: The rapid proliferation of online harms over the past fifteen years has exposed fundamental flaws in platform self-regulation, with information asymmetries preventing independent researchers from accessing data necessary to understand and address digital risks. Article 40 of the European Union's Digital Services Act represents a paradigm shift, establishing the first global mandate for research data access on social media platforms and search engines. This paper examines Article 40's implementation framework, analysing its regulatory structure, scope, and implications for evidence-based digital governance. The analysis reveals how Article 40 addresses longstanding information asymmetries through a two-tier data access system, enabling unprecedented research opportunities while establishing safeguards for privacy and platform concerns. The framework's focus on systemic risks – including illegal content dissemination, fundamental rights violations, and threats to democratic processes – ensures regulatory alignment while constraining research scope. However, successful implementation faces significant challenges, including substantial administrative burdens on regulators, institutional capacity requirements, and platform resistance to data transparency. With Ireland's Coimisiún na Meán serving as coordinator for 15 of 25 designated platforms, the framework's global implications extend beyond European borders. Article 40's success will ultimately depend on navigating complex implementation challenges to deliver the rigorous, systematic investigation that contemporary digital governance demands.

Keywords: digital governance and regulation, digital services, research data, ethical data use
JELs: K23, K24, L86, O38

1. INTRODUCTION

The past fifteen years have witnessed an unprecedented transformation in human communication through the emergence of online platforms and search engines, which, while delivering extraordinary benefits for information sharing and social connection, have also unleashed a proliferation of risks and harms that have evolved at dizzying speed. From child sexual abuse material to electoral manipulation, from information disorder to radicalisation, the scale and complexity of these issues demand evidence-based research approaches that can match the sophistication of the phenomena they seek to understand. Despite increasing policy focus and public interest in these critical societal questions, researchers have faced systematic barriers in accessing the data necessary to investigate platform operations and their effects on users and society.

For over a decade, social media platforms and search engines have maintained proprietary control over data relating to the online behaviour and activities of their users, content moderation policies and practices, and the design and decision-making algorithmic systems. This gatekeeping has created a significant information asymmetry, where platforms possess comprehensive knowledge of the online environment, while researchers, policymakers, and civil society must operate with limited, filtered, or entirely absent data. The resulting knowledge gaps have undermined efforts to develop evidence-informed policy responses to online harms.

This landscape changed fundamentally with the introduction of Article 40 of the European Union's Digital Services Act (Regulation (EU) 2022/2065) (DSA), which represents the first global mandate for research data access on social media platforms and search engines. As a novel regulatory instrument in a highly contested space, Article 40 creates unprecedented opportunities for research into critical questions facing contemporary society,

including online hate, cyberbullying and harassment, online predation and grooming, radicalisation, amplification of harmful or extremist content, identity theft, fraud, and content related to self-harm, suicide, or eating disorders.

While this regulatory innovation is a global first, the need is recognised in other jurisdictions where similar nascent initiatives are being progressed. The UK's 13/10/2025 17:47:00 includes provisions for "providing information for purposes related to the carrying out of independent research into online safety matters," while in Canada, the House of Commons' Standing Committee on Access to Information, Privacy and Ethics (Brassard, 2024) recently recommended that Canadian researchers be provided access to platform data. These developments signal a broader global recognition that effective platform governance requires moving beyond self-regulation toward evidence-informed policy making.

The DSA's approach with Article 40 aligns closely with established principles of evidence-informed policy making (EIPM), which emphasises using "different types of information in a variety of forms and from a variety of sources, reflective of and responsive to the policy and practice context" (Bowen & Zwi, 2005). Currently, much digital policy operates on assumptions about platform effects rather than systematic evidence – a situation that researchers like Cairney and Oliver (2017) have identified as creating "frustrating experiences of scientists trying to secure a response to the problems and solutions they raise." Article 40 directly addresses this knowledge gap by providing the data foundation needed to generate rigorous, policy-relevant evidence about platform harms and mitigation strategies.

Ireland's national media regulator, Coimisiún na Meán, occupies a pivotal position in this new regulatory framework. As Digital Services Coordinator (DSC), An Coimisiún has supervision and enforcement functions as well as a role in coordinating and cooperating with other national and EU authorities that also regulate online content. Further, as Digital Services Coordinator of Establishment (DSCoE) for 15 of the 25 Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) designated under the DSA, Ireland effectively serves as the regulatory hub for the platforms that dominate global digital discourse. This unique position places Irish regulators at the centre of facilitating research that could contribute to mitigating systemic risks not only within the European Union but globally, making Ireland's implementation of Article 40 a critical case study for the future of platform governance worldwide.

This paper examines the implementation framework of Article 40, analysing its regulatory structure, scope, and potential implications for research into online integrity. Through analysis of the data access mechanisms and assessment criteria, we explore how this unique mandate for platform data access could create new possibilities for interdisciplinary research while establishing safeguards to protect user privacy and platform security and confidentiality concerns. The analysis identifies both the transformative potential of mandated research access and the significant implementation challenges that will determine whether this regulatory innovation can deliver on its promise of evidence-based digital governance.

2. BACKGROUND

Proliferation and Understanding of Online Harms

The digital landscape has evolved almost beyond recognition over the past decade and a half, with a proliferation of potential online harms spanning individual safety, democratic processes, and societal wellbeing. Issues such as misinformation and disinformation, radicalisation, extremism and polarisation, threats to election integrity, and challenges to information integrity and freedom of expression that were either unknown or barely recognised ten years ago have now become central concerns in public discourse and policy debates through the investigative work of independent researchers. Alongside these systemic risks, researchers have documented the emergence of individual harms including cyberbullying and gendered harassment, online predation and grooming, financial fraud and exploitation, and the proliferation of content promoting self-harm, suicide ideation, and eating disorders. The targeting of vulnerable populations – from children exposed to sexual abuse material to elderly users susceptible to scams – has revealed new forms of digital exploitation, while algorithmic systems have introduced novel concerns around bias, discrimination, and the creation of filter bubbles that shape user behaviour in previously unimagined ways. This evolution reflects both the changing affordances of digital platforms (Bucher & Helmond, 2018) and the corresponding shifts in user behaviours that exploit or are affected by these technological capabilities (Papa & Ioannides, 2023).

The transformation of political discourse provides a clear demonstration of the rapid evolution of digital harms associated with online platforms and search engines. Ten years ago, threats to democratic processes through digital platforms were largely unimagined. The 2016 Brexit referendum and US Presidential election marked a turning point, with bot-nets (Bastos & Mercea, 2019), Russian troll farms, and "fake news" (Aïmeur et al., 2023) emerging as salient threats to democratic integrity – concepts that were barely in the public consciousness before these

watershed moments. Misinformation and disinformation have evolved from isolated propaganda efforts to sophisticated information campaigns that can reshape public understanding of critical issues (Allcott et al., 2019) (Bradshaw, 2019). Radicalisation and extremism facilitated through platform recommendation systems (DeCook & Forestal, 2023; Mølmen & Ravndal, 2021; Ribeiro et al., 2021) and community formation have enabled the growth of transnational networks that can coordinate real-world violence (Veilleux-Lepage & Archambault, 2019), while political polarisation (Baumann et al., 2020; Conover et al., 2021) and conspiracy theories (Faddoul et al., 2020) amplified by algorithmic content curation contributes to the fragmentation of shared democratic discourse.

The rapid emergence of these harms is exemplified by key watershed moments: the 2016 Brexit referendum and US Presidential election brought electoral manipulation into mainstream awareness, while the trajectory from Ukraine-specific disinformation concerns in 2014 to the global information integrity issues experienced during the COVID-19 pandemic demonstrates how quickly these phenomena can scale. The speed at which these issues proliferate and adapt continues to accelerate as platform features evolve and new forms of problematic conduct and content emerge alongside changing user behaviours.

Understanding Online Harm – the Problem of Information Asymmetry

Crucially, many of these emerging issues were not identified by the platforms themselves, but rather discovered and documented by independent researchers and journalists working with limited access to platform data. These external investigators have played a vital role in bringing public attention to issues ranging from algorithmic bias and filter bubbles to coordinated inauthentic behaviour and targeted harassment campaigns. However, their ability to conduct comprehensive, systematic research has been severely constrained by fundamental information asymmetries.

Over the past decade researchers studying online platforms have faced significant and increasing barriers to accessing the data they need. Social media companies and other digital intermediaries have treated their datasets as proprietary assets, maintaining strict gatekeeping policies. It is well-documented that platforms unilaterally determine what types of data to share and with whom, sometimes cutting off access entirely (Walker et al., 2019). Preferential access is offered to certain researchers and institutions (Urman et al., 2023), raising concerns about potential academic capture(Zingales, 2013). Many impose artificial constraints on research methodologies (Young et al., 2022) and limit what findings can be published, undermining the independence and rigor of academic inquiry (Timberg, 2021). This selective approach has created significant imbalances in research coverage, with some topics receiving disproportionate attention while others remain understudied (Urman et al., 2023).

These access restrictions have created two distinct but interconnected forms of information asymmetry that fundamentally disadvantage independent research. First, there is a profound asymmetry in analytical capability**.** Platforms possess comprehensive datasets that allow them to understand user behaviour, content effects, and systemic patterns with granular precision. Meanwhile, researchers and the public must rely on limited, filtered, or aggregate data that platforms choose to share – if they share anything at all. This creates a knowledge gap where platforms can investigate and understand phenomena that remain opaque to external researchers, leading to vastly different understandings of the same issues based on differential access to evidence. Second, there is an asymmetry in system knowledge – the "unknown unknowns" problem where researchers operate largely in the dark about what information exists and could potentially be studied. Researchers have only a limited understanding of the vast amounts of data collected by platforms (Shapiro et al., 2021).

While platforms have intimate knowledge of their own data architecture, collection methods, and available datasets, researchers operate largely in the dark about what information exists. This makes it difficult a priori to set out specific data needs across a wide variety of subjects and fields. Researchers cannot effectively request data they don't know exists, while platforms hold comprehensive knowledge of their data landscape but have little incentive to reveal its full scope.

This information asymmetry became particularly evident in debates about pressing issues like social media's impact on teenage mental health. As the American Psychological Association noted in a recent review, "the data required to make cause-and-effect conclusions are challenging to collect and/or may be available within technology companies, but have not been made accessible to independent scientists." (Health Advisory on Social Media Use in Adolescence, 2023). The self-regulatory model, which relied on platforms to voluntarily share data and police their own systems, proved inadequate for addressing these systematic barriers to independent oversight.

Addressing Online Harm through Digital Regulation

The growing public disquiet with issues relating to the online space has prompted legislative action across numerous jurisdictions worldwide to regulate online platforms. A comprehensive review by NYU researchers

identified online safety regulations in 19 countries (Olaizola Rosenblat et al., 2025). These initiatives encompass diverse regulatory approaches including the UK's Online Safety Act, Australia's Online Safety Act 2021, New Zealand's Harmful Digital Communications Act, and Singapore's Online Safety (Miscellaneous Amendments) Act 2022. The approaches vary significantly, ranging from content-based regulation that establishes prohibited content categories, to design-based regulation targeting platform architecture and features, transparency mandates requiring disclosure of operational information, and procedural safeguards ensuring fairness and accountability.

The European Union has developed a comprehensive legislative framework focused on digital safety, comprising the DSA, the Digital Markets Act (DMA), the Terrorist Content Online Regulation (TCOR) and the AI Act. The DSA aims to create a transparent and accountable online environment, and represents a fundamental shift in platform governance, moving beyond self-regulation to systematic oversight. The DSA's legislative architecture is built around a progressive obligation structure, imposing increasingly stringent regulatory requirements on online intermediaries based on their size and scope of service. All online intermediaries offering their services in the EU must comply with DSA provisions, regardless of whether they are established within the EU or operate from external jurisdictions. The most onerous obligations are placed on services designated as VLOPs or VLOSEs - platforms with more than 45 million average monthly active recipients of their services in the Union. These services face the highest level of obligations under the DSA, including enhanced transparency obligations, risk assessment requirements and content moderation obligations.

Central to this framework is Article 40 of the DSA, which mandates data access for researchers studying systemic risks. This provision directly addresses the information asymmetries that have long hindered independent research by requiring VLOPs and VLOSEs to provide access to their data for vetted researchers. Article 40 is designed to support the risk assessment and mitigation obligations outlined in Articles 34 and 35, creating an integrated framework where independent research underpins regulatory oversight. the data access provisions are explicitly designed to support the Act's systemic risk framework, as outlined in Articles 34 and 35. Article 34 requires platforms to identify and assess systemic risks including the dissemination of illegal content, negative effects on fundamental rights, and intentional manipulation of services. Article 35 mandates the implementation of reasonable, proportionate and effective mitigation measures. Article 40's research access provisions create a feedback loop, enabling independent researchers to contribute evidence that can inform both the supervision and enforcement of these obligations and broader policy-making processes. It is clear however that the effectiveness of the framework depends on practical implementation, whereby meaningful data access is provided to researchers who can investigate and evaluate risks and harms to deliver policy and regulatory insights.

Activating Data Access through Article 40

Article 40 represents the practical mechanism through which the DSA addresses the information asymmetries that have constrained research into online harms. By mandating data access for qualified researchers, it transforms the theoretical possibility of evidence-based platform governance into an operational reality. Article 40 establishes a two-tier system for research data access, carefully balancing the need for academic investigation with privacy protection and platform security and confidentiality concerns. This dual pathway approach recognises the different sensitivity levels of various data types while maintaining the overarching research purpose of contributing to systemic risk detection, identification, and understanding.

Public Data Pathway Article 40(12)

The first tier provides researchers with direct access to publicly available data through platform-mediated processes. Under this pathway, researchers apply directly to platforms, which then assess applications and approve or reject access requests. The scope includes access to content libraries, APIs of public posts, and other publicly accessible information that platforms make available to users and the general public. This direct researcher-platform interaction mirrors existing academic data access arrangements but formalises the process within a regulatory framework. Examples of accessible data include public posts, comments, sharing patterns, and engagement metrics that do not reveal personal information about individual users. The purpose limitation ensures that such access is "solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1)."

Non-Public Data Pathway (Article 40 (4-11))

Article 40, sub-sections 4 to 11 address access to non-public data, which may include personal data and special category data requiring enhanced protection measures. This access route requires researchers to apply to a DSC, either the DSC of the Member state of the research organisation to which they are the researcher affiliated, known as the DSC of Origin (DSCoO) or the DSC of establishment for the relevant platform (DSCoE). The application can be assessed by both DSCs where applicable, while the DSCoE has responsibility for approving or rejecting requests based on strict criteria as described by Article 40(8).

3. THE PARAMETERS OF DATA ACCESS UNDER ARTICLE 40

Scope of Data

The scope of data available under Article 40 extends far beyond the tabular user data and content metrics that researchers typically envision – such as posts, likes, and shares. Recital 13 of the Delegated Act on Data Access provides a comprehensive, though non-exhaustive, list revealing the breadth of data types potentially accessible for use in a wide variety of research domains. For computational and data science researchers, this includes platform testing data such as A/B test results and data relating to experimental features prior to deployment or the data used to personalise content for individual users. Content moderation policies, algorithmic and human moderation system documentation, changelogs, and repositories of moderated content and account actions could be utilised in legal and governance research to understand differences between jurisdictions User relationship networks, individual-level content exposure histories, detailed interaction data including comments and reactions, and comprehensive advertising and targeting information including profiling data and pricing measures could be available for research relating to polarisation, radicalisation and electoral integrity. This expansive data landscape represents a fundamental shift from the limited, aggregated datasets traditionally available to researchers, opening entirely new methodological possibilities across disciplines.

Geographic Scope

The geographic scope of Article 40 operates on a risk-based rather than territorial model, focusing primarily on systemic risks within the EU while recognising that digital risks transcend borders. Data relating to users and platform activities outside the EU can be included in research datasets, but only where such data is demonstrated to aid understanding of systemic risks within the EU. The framework also embraces international research collaboration by allowing researchers based outside the EU to apply for data access, provided they meet the eligibility criteria outlined in Article 40(8) of the DSA. Data access operates within strict jurisdictional constraints designed to ensure GDPR compliance with relevant Data Protection Legislation and protect European data sovereignty. All data access requests should ensure appropriate safeguards are in place to fulfil the requirements of data security, data confidentiality and protection of personal data. In cases where providing access to data involves a transfer of personal data to third countries or international organisations (within the meaning of Chapter V of Regulation (EU) 2016/679), appropriate transfer mechanisms and necessary actions to comply with GDPR are required. Researchers, regardless of their institutional base, must establish secure data handling infrastructure that meets European data protection standards, often requiring partnerships with EU-based institutions or cloud providers. These safeguards ensure that while research collaboration can be international in scope, data governance remains firmly anchored in European privacy and protection principles.

Usage and Modality

A crucial distinction in the Article 40 framework is that researchers receive access to data for a specified period, after which access is revoked, this stands in contrast to the traditional conception of a permanent handover of data. This approach maintains platform control over sensitive information while enabling meaningful research within defined temporal boundaries.

Systemic Risk

The concept of systemic risk is fundamental to all data access requests under the Article 40 framework. Data must be requested in order to conduct "research that contributes to the detection, identification and understanding of systemic risks in the Union, as set out pursuant to Article 34(1), and to the assessment of the adequacy, efficiency and impacts of the risk mitigation measures pursuant to Article 35". This creates important practical constraints on research eligibility. Article 34(1) defines systemic risks to include the dissemination of illegal content, negative effects on the exercise of fundamental rights, negative effects on civic discourse and electoral processes, and threats to the protection of public health and minors. Consequently, general platform research driven by academic curiosity or commercial interests without clear connections to these specific risk categories would not qualify for data access. Instead, researchers must demonstrate and articulate how their proposed investigations contribute to understanding risks that have the potential to affect significant numbers of EU users, ensuring that Article 40 serves the DSA's regulatory objectives rather than functioning as a general academic data-sharing mechanism.

4. IMPLEMENTATION FRAMEWORK: CRITERIA AND SAFEGUARDS

The practical implementation of Article 40 requires rigorous assessment criteria and safeguards to balance research access with privacy protection and platform operational concerns. The framework establishes strict criteria for researcher qualification while ensuring that approved research contributes meaningfully to understanding and mitigating systemic risks.

Researcher Specific Qualification Criteria

Applications for data access must demonstrate that researchers meet specific qualification requirements designed to ensure both research integrity and data protection. Researchers must be affiliated with a research organisation as defined in Article 2, point (1), of Directive (EU) 2019/790, ensuring institutional oversight and accountability.

Independence from commercial interests represents a crucial requirement, preventing platform competitors or other commercial entities from using research access for business intelligence purposes. Applications must disclose research funding sources, enabling assessment of potential conflicts of interest that could compromise research integrity.

Technical and Organisational Safeguards

Researchers must demonstrate capability to fulfil specific data security and confidentiality requirements corresponding to each data access request. This includes describing appropriate technical, organisational and legal measures for protecting personal data, ensuring compliance with the GDPR and other applicable privacy laws. The assessment process resembles the rigour of European Research Council (ERC) funding applications, requiring comprehensive project descriptions, methodology explanations, and detailed data protection plans. Institutional support for researchers becomes crucial, as organisations must provide adequate data protection and IT security infrastructure to handle sensitive platform data.

Proportionality and Necessity Assessment

Applications must demonstrate that data access and requested time frames are necessary for and proportionate to research purposes. This requirement ensures that researchers request only the minimum data necessary, for the minimum time period required to address their research questions, preventing excessive data collection that could compromise user privacy without corresponding research benefits. The expected results must contribute to the purposes laid down in Article 40, specifically supporting systemic risk detection, identification, and understanding. This substantive requirement ensures that approved research addresses the DSA's core objectives rather than peripheral academic interests.

Public Benefit Requirements

Researchers must commit to making research results publicly available free of charge within a reasonable period after completion, subject to the rights and interests of platform users. Through this open access requirement the EU underscores the aim that research conducted utilising data accessed via Article 40 contributes to broader understanding of online safety and integrity issues, as well as to the regulatory framework by directly informing risk assessments and mitigation, ultimately improving platform safety and governance.

5. OPPORTUNITIES, CHALLENGES AND FUTURE IMPLICATIONS

The implementation of Article 40 presents significant challenges while creating transformative opportunities for research into online integrity. Managing these challenges effectively will determine whether the regulatory framework achieves its ambitious goals of supporting evidence-based digital governance.

Opportunity for Novel Research Methods

Access to comprehensive platform data through Article 40 enables entirely new methodological approaches that were previously impossible due to data limitations. The scale and granularity of platform datasets create opportunities to utilise research methodologies, which were previously impossible to pursue such as implementing sophisticated causal inference methods, machine learning models, and longitudinal analyses. Researchers can now investigate questions that were previously unanswerable, such as how specific algorithmic changes affect user behaviour over time, or how network effects influence the spread of particular types of content across different demographic groups. These large-scale datasets particularly enable the application of methods that require substantial statistical power to detect effects. Similarly, the temporal depth of platform data allows for dynamic modelling approaches that can capture how online phenomena evolve and interact over time.

Interdisciplinary Collaboration Requirements

The systemic risks noted within the DSA are inherently interdisciplinary challenges that cannot be adequately investigated by any single discipline. Examining the dynamics at play between teenage mental health and social media, for example, requires not only data science capabilities to analyse behavioural patterns, but also psychological expertise to interpret mental health outcomes, developmental psychology to understand adolescent vulnerabilities, and public health perspectives to assess population-level impacts. Similarly, studying electoral integrity demands collaboration between political scientists, data scientists and legal experts understanding democratic institutions.

The complexity of questions coupled with scale and scope of data creates natural requirements for disciplines that don't traditionally collaborate to work together to develop new modes of interdisciplinary collaboration. Thus, the research possibilities enabled by Article 40 data access extend far beyond conventional qualitative-quantitative mixed methods approaches. Instead, researchers can now combine computational analysis of behavioural data at massive scale with targeted qualitative investigation of specific user experiences, policy document analysis, and

experimental interventions – all within the same research framework. This represents a fundamentally new form of mixed methods research that combines big data analytics with targeted deep investigation in ways that neither traditional quantitative nor qualitative approaches could achieve alone.

Managing Demand and Expectations

The position of DSCoE for 15 of 25 VLOPs and VLOSEs creates a substantial administrative burden for Coimisiún na Meán. The assessment of complex research applications requires considerable specialised expertise in digital platform operations, research methodology, data protection, legal and information security. Building adequate regulatory capacity to handle expected application volumes while maintaining assessment quality represents a significant implementation challenge and will require significant resourcing.

The contentious nature of platform data access adds political and legal complexity to implementation. Platforms may resist extensive data sharing requirements, while researchers may have unrealistic expectations about data availability and access timelines. Managing these competing interests requires careful balance between regulatory authority and practical feasibility.

Institutional Capacity Building

Supporting researchers with adequate data protection expertise and IT infrastructure will require significant institutional investment. Universities and research organisations will require additional capabilities for handling large-scale, sensitive platform data while maintaining GDPR compliance and research ethics standards. This transformation necessitates research offices adopting governance frameworks and protocols similar to those deployed in medical research, where handling sensitive human data requires rigorous institutional oversight, specialised ethical approvals, and comprehensive data management and data protection systems and supervision. The scale and complexity of platform datasets also create compelling arguments for shared research infrastructure that can serve multiple institutions and research communities. Rather than each university developing independent capabilities, collaborative models could include shared secure computing environments, common data analysis platforms, and pooled technical expertise – similar to how research in particle physics depends on collaborative laboratory infrastructure. Such shared infrastructure could reduce individual institutional costs while enabling smaller universities and researchers to participate in platform data research that would otherwise be prohibitively expensive. In France, the viability of such collaborative research infrastructure approaches has been demonstrated (Fabre et al., 2021).

The technical complexity of platform data analysis demands new methodological expertise within the research community. Training programs, methodological workshops, and collaborative research networks will be essential for maximising the research potential of Article 40 access.

6. CONCLUSION

The rapid evolution of online harms over the past decade has exposed fundamental flaws in the self-regulatory model that long governed digital platforms. The information asymmetries that enabled platforms to gatekeep research access while maintaining comprehensive internal knowledge of their own systems proved incompatible with the public interest in understanding and mitigating these emerging risks.

Article 40 of the Digital Services Act represents a paradigm shift moving from voluntary data sharing to mandated research access that directly addresses the dual information asymmetries that have long constrained independent investigation. By requiring VLOPs and VLOSEs to provide vetted researchers with access to both public and non-public data, the framework creates unprecedented opportunities for evidence-based platform governance. The integration of research access with the DSA's broader risk assessment and mitigation obligations – particularly Articles 34 and 35 – establishes a feedback loop in which independent research can inform both regulatory oversight and policy development.

The transformative potential of Article 40 extends beyond simply providing more data to researchers. The framework enables entirely new methodological approaches that combine large-scale computational analysis with targeted qualitative investigation, fostering interdisciplinary collaboration between fields that rarely work together. Questions that were previously unanswerable due to data limitations – from understanding algorithmic influence on vulnerable populations to mapping the dynamics of coordinated inauthentic behaviour – become tractable when researchers can access comprehensive platform datasets with appropriate safeguards.

However, realising this potential requires careful navigation of significant implementation challenges. The administrative burden on Digital Services Coordinators, particularly Coimisiún na Meán as coordinator of establishment for the majority of designated platforms, demands substantial regulatory capacity building.

Universities and research institutions must develop new technical and organisational capabilities to handle sensitive platform data while maintaining GDPR compliance. The research community itself must evolve to embrace interdisciplinary collaboration and develop methodological expertise suited to the scale and complexity of platform data analysis.

The global implications of Article 40 extend well beyond European borders. As other jurisdictions develop their own platform governance frameworks – from the UK's Data (Use and Access) Act to emerging legislative proposals in Canada – the European model offers both inspiration and practical lessons for evidence-based digital regulation. The risk-based geographic scope that allows inclusion of non-EU data relevant to European systemic risks, combined with provisions for international researcher collaboration, positions Article 40 as a potential foundation for global research infrastructure addressing transnational digital issues.

Ultimately, Article 40's success will be measured not merely by the volume of research it enables, but by its contribution to more effective, evidence-based responses to online harms. By reducing the information asymmetries that have long disadvantaged independent research relative to platforms, the framework creates conditions for the rigorous, systematic investigation that complex digital governance challenges demand. As online harms continue to emerge and evolve, the ability to generate timely reliable evidence about platform operations and their societal impacts becomes not just an academic aspiration, but a democratic necessity. Article 40 provides the regulatory architecture to meet this critical need, but its success hinges on navigating significant and complex implementation hurdles, not least the resistance of platforms to meaningful data transparency.

References

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining*, *13*(1). https://doi.org/10.1007/s13278-023-01028-5

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, *6*(2), 2053168019848554.

Bastos, M. T., & Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*, *37*(1), 38–54. https://doi.org/10.1177/0894439317734157

Baumann, F., Lorenz-Spreen, P., Sokolov, I. M., & Starnini, M. (2020). Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, *124*(4), 048301. https://doi.org/10.1103/PhysRevLett.124.048301

Bowen, S., & Zwi, A. B. (2005). Pathways to "evidence-informed" policy and practice: A framework for action. *PLoS Medicine*, *2*(7), e166.

Bradshaw, S. (2019). Disinformation optimised: Gaming search engine algorithms to amplify junk news. *Internet Policy Review*, *8*(4). https://doi.org/10.14763/2019.4.1442

Brassard, J. (2024). *Oversight of Social Media Platforms: Ensuring Privacy and Safety Online*. House of Commons Standing Committee on Access to Information, Privacy and Ethics. https://www.ourcommons.ca/Content/Committee/441/ETHI/Reports/RP13390047/ethirp16/ethirp16-e.pdf

Bucher, T., & Helmond, A. (2018). The Affordances of Social Media Platforms. In J. Burgess, A. Marwick, & T. Poell, *The SAGE Handbook of Social Media* (pp. 233–253). SAGE Publications Ltd. https://doi.org/10.4135/9781473984066.n14

Cairney, P., & Oliver, K. (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy?. *Health Research Policy and Systems*, *15*(1), 35.

Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., & Flammini, A. (2021). Political Polarization on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1), 89–96. https://doi.org/10.1609/icwsm.v5i1.14126

DeCook, J. R., & Forestal, J. (2023). Of Humans, Machines, and Extremism: The Role of Platforms in Facilitating Undemocratic Cognition. *American Behavioral Scientist*, *67*(5), 629–648. https://doi.org/10.1177/00027642221103186

Fabre, R., Egret, D., Schöpfel, J., & Azeroual, O. (2021). Evaluating the scientific impact of research infrastructures: The role of current research information systems. *Quantitative Science Studies*, *2*(1), 42–64. https://doi.org/10.1162/qss_a_00111

Faddoul, M., Chaslot, G., & Farid, H. (2020). *A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos* (arXiv:2003.03318). arXiv. http://arxiv.org/abs/2003.03318

Health Advisory on Social Media Use in Adolescence. (2023). American Psychological Association. https://www.apa.org/topics/social-media-internet/health-advisory-adolescent-social-media-use

Mølmen, G. N., & Ravndal, J. A. (2021). Mechanisms of online radicalisation: How the internet affects the radicalisation of extreme-right lone actor terrorists. *Behavioral Sciences of Terrorism and Political Aggression*. https://doi.org/10.1080/19434472.2021.1993302

Olaizola Rosenblat, M., Agrawal, A., & Yap, I. (2025). *Online safety regulations around the world: The state of play and the way forward – A resource guide.* Center for Business and Human Rights, NYU Stern School of Business.

Papa, V., & Ioannides, N. (2023). Reviewing the impact of Facebook on civic participation: The mediating role of algorithmic curation and platform affordances. *The Communication Review*, 1–23. https://doi.org/10.1080/10714421.2023.2186680

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2021). *Auditing Radicalization Pathways on YouTube* (arXiv:1908.08313). arXiv. http://arxiv.org/abs/1908.08313

Shapiro, E. H., Sugarman, M., Bermejo, F., & Zuckerman, E. (2021). *New approaches to platform data research.* https://www.netgainpartnership.org/resources/2021/2/25/new-approaches-to-platform-data-research

Timberg, C. (2021, September 10). Facebook made a big mistake in data it provided to researchers, undermining academic work. *Washington Post*. https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/

Urman, A., Smirnov, I., & Lasser, J. (2023). *The right to audit and power asymmetries in algorithm auditing*. arXiv; arXiv. https://doi.org/10.48550/arXiv.2302.08301

Veilleux-Lepage, Y., & Archambault, E. (2019). *Mapping Transnational Extremist Networks: An Exploratory Study of the Soldiers of Odin's Facebook Network, Using Integrated Social Network Analysis*. *13*(2).

Walker, S., Mercea, D., & Bastos, M. (2019). The disinformation landscape and the lockdown of social platforms. *Information, Communication & Society*, *22*(11), 1531–1543. https://doi.org/10.1080/1369118X.2019.1648536

Young, M., Katell, M., & Krafft, P. M. (2022). Confronting Power and Corporate Capture at the FAccT Conference. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1375–1386. https://doi.org/10.1145/3531146.3533194

Zingales, L. (2013). Preventing Economists' Capture. In D. Carpenter & D. Moss (Eds.), *Preventing Regulatory Capture: Special Interest Influence and How to Limit it*. Cambridge University Press.