Columbia World Projects | Thinking Doing Columbia Global

Hertie School Centre for Digital Governance

October 2025

Building Capacity for Data Access, Analysis + Accountability

Report of the Columbia-Hertie Working Group:
Mapping Strategies and Resources to Safeguard Democracy in the Digital Age

Authors

Mark Scott
Senior Resident Fellow at the Digital Forensic Research Lab, Atlantic Council; Research Fellow, Centre for Digital Governance, Hertie School

Daniela Stockmann
Professor of Digital Governance; Director, Centre for Digital Governance, Hertie School

Thomas Asher
Director of Research and Engagement, Columbia World Projects

Anna Marchese
Senior Project Officer, Columbia World Projects

Acknowledgements

We are grateful to those who participated in a series of working group meetings on data access for their continued collaboration and commitment to this effort, and to colleagues from Columbia World Projects and the Centre for Digital Governance for their contributions, advice and overall support. We also thank our student assistant, Shruti Das, for conducting valuable research that informed the report's contents. This report was produced with support from the John F. and James L. Knight Foundation.

About Columbia World Projects

Columbia World Projects (CWP) is an initiative of Columbia University that connects policymakers, practitioners and researchers to explore pressing issues, including what sustains and ails liberal democracies around the world, while offering focused solutions for revitalizing democracy.

About the Centre for Digital Governance

The Centre for Digital Governance at the Hertie School is a gateway to governance of digital transformation, based in Berlin. The Centre promotes digitalization where appropriate to improve public well-being. To this end, the Centre enables and supports synergies between rigorous academic research, world-leading education and socially relevant policy insights on the challenges and opportunities of the digital era.

Table of Contents

Executive Summary

Over the last year, Columbia World Projects and the Hertie School's Centre for Digital Governance, with support from the John S. and James L. Knight Foundation, brought together more than 120 experts from across the regulatory and researcher community from the United States and Europe with one objective:

To identify the current gaps in social media data access and outline where public and private funders can meet these opportunities to support democratic institutions and norms worldwide.

The ability to access social media data may not appear to be fundamental to addressing such questions — especially at a time of heightened geopolitical uncertainty, significant pullback from both philanthropic and public support for such work and increasingly antagonistic relations between platforms, regulators and the research community.

But in late 2025, little is still known about the internal policy choices made within some of the world's largest companies that fundamentally shape what type of content is displayed to billions of people from California to Croatia to Cambodia.

This lack of transparency has real-world consequences.

Citizens cannot make active choices about what they see on social media. Independent regulators cannot hold companies accountable for their obligations under a growing number of national and regional online safety regimes. The research community — made up of academics, civil society groups and the media — cannot highlight potential deficiencies in both platform and regulatory action.

Collectively, it represents a deficit in social media platform transparency and accountability that is a direct threat to individuals' fundamental rights, as well as to wider societal democratic norms.

Funders, regulators and researchers must act within the next 6-12 months to establish foundational infrastructure and standards related to social media data access. Without swift action, democratic institutions are vulnerable to the weaponization of social media platforms whose activities remain opaque and subject to potential manipulation by malign actors.

It is within this context the Columbia-Hertie initiative provides clear funding recommendations, as outlined in the chart below.

At its core, this work is based on upholding the highest levels of data protection and security practices so that any form of social media data access protects the privacy rights of individual social media users — no matter where they are located. That is the guiding principle for all recommendations.

The report is divided into three sections:

1 **Supporting Underlying Data Access Infrastructure**
2 **Building Best Practices for the Research Community**
3 **Fostering Researcher-Regulator Relationships**

Each of these sections provide specific recommendations on how public and private funders can meet the existing opportunities within social media data access. The recommendations include which type of funder is most appropriate; how much money is required to meet the objectives; and a time-scale for results.

Recommendations key

Funding Source	Costs		Time Frame	
Public	$	$25K–$75K	Short-term	1-2 years
Private	$$	$75K–$200K	Mid-term	2-4 years
Public / Private	$$$	$200K–$500K	Long-term	4+ years
	$$$$	$500K+		

Recommendations

1 Supporting Underlying Data Access Infrastructure

1.1 Regulatory, Legislative and Legal Certainty for Data Access

Public/Private | $$$ | Short-term
Clarify legal frameworks for independent data access (e.g., data donations and public-interest scraping) and support policymaking, advocacy and strategic litigation.

1.2 Wholesale Data Access Infrastructure

Public/Private | $$$$ | Long-term
Develop a secure, society-wide technical infrastructure as a public good to reduce costs, centralize protections and accelerate research.

1.3 Research Safe Harbors and Legal Liability Funds

Private | $$ | Short-term
Establish safe harbors for public-interest research and community-wide legal support to protect researchers from liability.

1.4 International Standards for Researcher Ethics and Data Protection

Public/Private | $$ | Mid-term
Create a multi-stakeholder, internationally recognized ethics and privacy body to establish global standards for cross-border social media data access.

2 Building Best Practices for the Research Community

2.1 Standardized Legal and Regulatory Documentation

Private | $ | Short-term
Overcome bespoke data access by creating data management plans, DPIA templates and other standardized submission documents for regulated regimes.

2.2 Standardized Data Protection and Security Protocols

Public/Private | $$ | Short-term
Integrate standardized documentation with training and capacity-building initiatives to ensure researchers meet high data protection and security standards.

2.3 Outreach and Onboarding New Researchers

Public/Private | $ | Short-term
Expand access via grants and fellowships that train new researchers in data access methods, with mentoring and peer-to-peer learning opportunities.

2.4 Data Access Community Engagement and Coordination

Public/Private | $$ | Mid-term
Coordinate disparate research efforts to maximize data access opportunities and outcomes, and support grassroots engagement and in-person workshops to foster information sharing.

3 Fostering Cooperative Researcher-Regulator Relationships

3.1 Information Sharing Between Researchers and Regulators

Public | $$ | Long-term
Establish institutional mechanisms to share regulatory priorities and actionable evidence, and support neutral convenings with equal stakeholder participation.

3.2 Funding Support for Independent Data Access Research

Public | $$$$ | Mid-term
Expand public funding for flexible, independent research on data access, online safety and platform governance through open, transparent and agile tendering processes.

3.3 Formalized Researcher Secondments and Fellowships

Public | $ | Mid-term
Create secondments and fellowships to embed researchers within regulatory agencies — and vice versa — to build mutual understanding, enhance research impact and strengthen the capacity of under-resourced institutions.

3.4 Engagement with National Data Protection Regulators

Private | $ | Mid-term
Engage national data protection regulators to clarify how independent data access (e.g., data donations and public-interest scraping) complies with privacy laws.

3.5 Demonstrations and Case Studies for Data Access

Public/Private | $ | Mid-term
Create a repository showcasing exemplary research to support learning and highlight the utility of data access.

Introduction

Fast-evolving technological changes are pushing countries' political institutions and norms to the breaking point.

Social media platforms are now central to everyday communication. They have become a linchpin in how politicians speak to citizens, how citizens form their political opinions and how harmful actors attempt to target national democratic processes and undermine online safety through information manipulation.

Artificial intelligence — fueled by the collection and use of reams of information — has only accelerated these trends. These systems allow for the creation and dissemination of online content within seconds and play a central role, through companies' internal **recommender systems**, in what content people consume in their social media feeds.

Citizens and policymakers alike lack sufficient insight into the effects of algorithms that increasingly influence our notions of community and politics. Independent researchers — drawn from a wide community of academics, research and advocacy organizations and the media — similarly lack a quantifiable understanding of how these platforms function. This limits their ability to support the level of accountability and transparency necessary to uphold democratic principles and protect countries from potential national security threats associated with social media and other digital platforms.

Accountability refers to the ability of independent groups and regulators to hold some of the world's largest companies responsible for their stated public commitments to promote free speech and safeguard users from harm, as outlined in the firms' terms of service. It also extends to abiding by national and regional online safety regulation – most evident within the **European Union's Digital Services Act (DSA)** — where researchers can similarly hold regulators accountable for their commitment to implement such legislation to protect people's fundamental rights.

Transparency refers to the provision of data, information and other governance structures to facilitate independent regulators and the researchers to shed light on how social media platforms operate, what data drives decisions when displaying billions of posts worldwide to users each day and how such policies affect individuals' free speech rights and safety online.

Within this context, comprehensive and scaled researcher data access regimes — either via regulatory mandates or voluntary commitments — are fundamental to achieving these accountability and transparency goals. They enable the public to understand the forces and policy decisions shaping the content they encounter in their social media feeds.

Researchers' ability to access both publicly available and private social media data — while adhering to privacy and security best practices — can improve the documentation of potential online harms. With better access, researchers can also track how information flows across digital platforms in ways that affect countries' democratic institutions and norms, as well as individuals' fundamental rights.

As artificial intelligence becomes more prevalent across society, such data access provisions — based on existing digital platform governance — can similarly provide a benchmark for how artificial intelligence researchers might access such data to hold firms and regulators accountable for the development and implementation of these systems.

Current researcher data access does not deliver on these aspirations.

It remains a cottage industry where individual organizations access one-off datasets to meet separate research, policy and advocacy objectives. Academics collect data for specific research; news organizations access data related to particular events; and civil society groups focus on a narrow set of issues, often over a short-term basis and depending on the type of funding they receive.

This approach does not scale to provide support for democratic resilience in an era of rapid technological change.

To meet this challenge, **Columbia World Projects** and the **Hertie School's Centre for Digital Governance**, with support from **the John S. and James L. Knight Foundation**, gathered more than 120 experts over five meetings between November 2024 and May 2025 to identify the current gaps in researcher data access and to provide recommendations for public-private funding opportunities to address these gaps.

These convenings between national and regional online safety regulators, academics, civil society groups and data access infrastructure providers were held under the Chatham House Rule.

The recommendations below outline practical opportunities where targeted funding support for researcher data access would bolster countries' democratic resilience and improve platform accountability at a time of heightened geopolitical tensions and an **industry-wide pullback** on trust and safety commitments.

Public and private funding will both play a key role in furthering data access.

Private philanthropy, embodied in the work of independent foundations and individual donors, can support policy experimentation and advocacy, community-wide capacity building and cooperation among researchers and regulators. Public funding, such as that provided via public tenders and other grant programs administered by national and subnational agencies, can underwrite essential technological infrastructure, support project development and help scale philanthropy-backed initiatives to meet national, regional, or global data access needs.

Each recommendation is associated with a potential funding partner (Public, Private, or Public/Private), as well as the size of any financial commitment. That includes **a scale $ ($25-$75,000); $$ ($75-$200,000); $$$ ($200,000-$500,000); and $$$$ ($500,000+), and a time frame: Short-term (1-2 years); Mid-term (2-4 years); and Long-term (4+ years.)**

This report is intended for public and private funders seeking to address issues of democratic resilience, platform accountability and transparency and the rise of artificial intelligence in an increasingly digitally connected world — one in which countries pursue different approaches to online safety and digital platform regulation.

By supporting long-term and sustainable funding opportunities related to data access, funders can also advance aligned priorities that include (but are not limited to) social equity, free expression, climate change, economic resilience, LGBTQ+ rights, foreign interference and political discourse.

It offers a range of options — from long-term investments in digital infrastructure to short-term support for researcher-regulator capacity building — prioritized during the Columbia-Hertie workshops as critical to addressing current shortfalls in researcher data access.

Given the diverse objectives and support criteria of individual funders, the report does not prioritize which funding options should take precedence.

Participants acknowledged, however, that underlying data access infrastructure — encompassing technical, administrative and legal frameworks (see Recommendation 1) — is a critical foundation upon which other areas of support (outlined in Recommendations 2 and 3) can be built.

No single funder can meet the challenge of promoting researcher data access to protect countries' democratic institutions and norms. Different countries and funders will pursue distinct approaches when supporting national regulatory efforts or voluntary commitments from digital platforms.

Given these realities, and the transatlantic nature of the Columbia-Hertie initiative, this report takes a jurisdictionally agnostic approach to address how its recommendations can be tailored to specific national or regional needs.

The report's underlying message is clear: researcher data access is a foundational pillar to meet the challenge of promoting democratic resilience in the age of social media and artificial intelligence.

For now, that promise remains unfulfilled. It is time to realize how independent accountability and transparency can safeguard and enhance people's fundamental rights while protecting countries' democratic institutions and norms.

Background:
State of Researcher Data Access

Social media platforms have become gatekeepers to fundamental information about society. Yet a significant knowledge deficit persists between what happens on social media and what regulators, academics and other independent researchers understand about these platforms.

To address this gap, there is an urgent need to analyze how these companies interact with wider society — through greater transparency and accountability efforts led by the independent research community, as defined above, and by regulators at both national and regional levels.

These efforts focus on enabling mandated or voluntary access to both public and private social media datasets. The goal is to provide — through privacy- and security-preserving mechanisms, and in compliance with purpose limitations in existing privacy legislation — academics, civil society organizations and media outlets with the information necessary to verify how companies meet their internal and external obligations under their terms of service and evolving regulatory requirements. This data includes, but is not limited to, aggregate engagement figures, systemic user behavior patterns and information about companies' recommender systems.

Current data access regimes fall into two camps.

Regulatory: Article 40 of the DSA outlines how such transparency and accountability mechanisms should work within the 27-country bloc's regulation. The United Kingdom's Data (Use and Access) Act similarly empowers the country's lawmakers to enact a separate legally-mandated data access regime. Meanwhile, countries worldwide — not least Canada, Australia, Brazil and Taiwan — are exploring or implementing data access regimes.

Voluntary: Independent researchers access social media data via direct partnerships with the companies, company-led Application Programming Interfaces (APIs), and/or separate data collection techniques like data donations (in which social media users share their data through regulatory structures or third-party applications) and public-interest scraping (the automated collection of publicly available social media data). This voluntary approach has become the de facto norm in the United States, where repeated legislative efforts to mandate researcher data access have fallen short.

Groups reliant on regulatory and voluntary data access protocols include academics, civil society organizations, regulators and media organizations. Collectively, they represent a broad church of independent researchers and oversight bodies with varying objectives, funding sources and capacities. Data includes publicly-accessible information, whose definition still needs to be determined, and private datasets, including personal information about specific social media users. At present, who can access social media data and what constitutes public or private datasets vary between jurisdictions and research communities.

All data collection must preserve the highest data protection and security safeguards. That includes the use of aggregation, encryption and anonymization techniques to protect individual social media users' posts from potential abuse (as detailed in the 2018 Cambridge Analytica scandal). Any form of private social media data — embodied, for example, by independent research on Facebook and Instagram during the 2020 US election — requires additional safety precautions due to the sensitivity of such information.

The current landscape of independent researcher data access creates a series of capacity, resource and regulatory gaps that make existing independent transparency, accountability and privacy preservation efforts less than the sum of their parts. It has led to increasingly antagonistic relationships with platforms, many of which have reduced researchers' access to publicly-available data or threatened legal action against those conducting public-interest research deemed to violate companies' terms of service.

Digital regulation, notably provisions within the DSA and the United Kingdom's updated Online Safety Act (OSA), has yet to be fully implemented, even as lawmakers emphasize that this regulatory landscape will continue to evolve to address the rapidly changing challenges of governing digital platforms.

Spotlight 1
Defining Publicly Available Data

The **Publicly Available Platform Data Expert Working Group**, organized by the Knight-Georgetown Institute, is developing "a uniform, cross-industry framework that allows for understanding the online information ecosystem as a whole." Its work begins with defining what types of platform data should be made publicly available — and how and when this should occur — in alignment with existing regulatory requirements.

Under the DSA's Article 40.12, which addresses real-time access to public data, researchers face barriers from companies whose application processes for publicly available data are bureaucratic and, to date, **underutilized**. Under Article 40.4, access to primarily private or sensitive datasets has not yet commenced, following the European Commission's publication of **detailed guidance** in July. This process is only beginning, and large volumes of data access requests are unlikely to be approved before mid-2026 at the earliest.

An **increasingly adversarial approach** from some social media companies has reduced researchers' access to both public-facing and private social media data. In August 2024, Meta **closed** CrowdTangle, its internal data analytics tool used widely by researchers to track real-time social media trends. Companies like **X** and **Meta** sued external efforts to access their publicly-available data via external scraping mechanisms that the firms said breached their terms of service.

Most social media firms **placed limitations** on independent collection mechanisms — like direct user data donations and public-interest scraping — citing potential violations of their terms of service. This has occurred even when the research aims were to identify potential risks, including the spread of hate speech, politically motivated violence against vulnerable communities or attacks on the integrity of democratic institutions and practices.

At present, independent researchers lack the collective capacity and the technical or financial resources needed to hold both platforms and regulators accountable for their commitments to governance, transparency and oversight.

Currently, social media data access is largely limited to a small number of researchers who have direct relationships with platforms, possess the technical skills to access data while maintaining required data protection and security standards, understand existing regulations and/or have the financial resources to leverage data access opportunities. Reliance on relationships with platforms imposes significant constraints on the scope of data available, the duration of access and restrictions on publication.

Other asymmetries include a reliance on US-based data access infrastructure that favors US-based organizations over their international counterparts. These dynamics limit involvement of independent researchers from other research areas and those from the so-called Global Majority, many of whom have insufficient resources to meaningfully leverage data access opportunities to address local transparency and accountability needs.

Opportunities to support democratic institutions and norms through social media data access are abundant. Yet, as of October 2025, existing efforts have fallen short of what is needed to hold the world's largest technology companies accountable for their commitments to online safety, free speech and the protection of electoral processes.

Without meaningful data access, the research community cannot effectively complement regulators in meeting the obligations of expanding national and regional online safety legislation and in uncovering the societal impacts of platform operations.

It is within this landscape — and the quickly evolving political, technical and funding environment — that this report frames potential public and private funding opportunities to meet the needs of independent researchers, social media companies, regulators and policymakers.

Recommendation 1: Supporting Underlying Data Access Infrastructure

A theme that emerged repeatedly during the Columbia–Hertie workshops was the need to prioritize open, cost-effective and secure data access infrastructure for the broader research community. This **includes** technical infrastructure, such as data archives; administrative infrastructure, including legal and policy frameworks; and organizational infrastructure that enables researchers to collaborate across institutions.

Technical infrastructure refers to both the tools for accessing data and the repositories for storing it. It encompasses offerings provided by social media companies (e.g., APIs and so-called "Clean Rooms," online platforms where researchers can analyze data, as outlined in the **Meta Content Library**). It also includes external repository providers, such as the National Conference on Citizenship's **Junkipedia**, which focuses on real-time data from numerous social media platforms and is accessible to a wide range of researchers, and the University of Michigan's **Social Media Archive**, whose archives allow predominantly academic researchers to access both publicly available and private datasets.

Infrastructure also encompasses independent data collection tools and processes, such as direct social media user data donations and public-interest scraping, which are currently conducted primarily on a project-by-project basis. These approaches can be constrained by companies' terms of service, which create legal risks around accessing data through such mechanisms.

Each of these data access options addresses different researcher needs, depending on an organization's technical capacity, the research questions being studied and the funding resources available.

Participants emphasized that the issue is not about choosing one infrastructure option over another. Instead, there is a need to combine all options — tailored to researchers' individual requirements — to reduce the costs of data storage and access and to create collective solutions. This approach enables each researcher or institution to advance greater transparency, accountability and security through mutually reinforcing efforts.

While there was an expressed need to create multiple data access offerings to ensure "strategic redundancy" in case existing tools were taken offline (for financial or legal reasons), participants stressed that the current funding environment makes it difficult to support several competing infrastructures, each of which would require significant financial backing.

What was universal across all forms of data access infrastructure discussed during the workshops was that financial and technical constraints currently limit researchers' capacity. At present, there is insufficient funding and technical expertise to scale solutions across the research community, and available resources often remain siloed between formal organizations and informal groups.

Separate non-technical infrastructure considerations present additional barriers that limit the utility of social media data access mechanisms. These include globally accepted ethics standards, legal certainty around independent data collection methods, and, relatedly, pooled liability funds for researchers who may face lawsuits related to their public-interest work.

Below is an overview of potential public-private funding opportunities to address gaps in the current social media data access research infrastructure. As noted above, all recommendations are associated with likely costs, timeframes and funding partners, based on participants' feedback.

1.1 Regulatory, Legislative and Legal Certainty for Data Access

Cost ($$$); Timeframe (Short-term); Funding partner (Private/Public)

The DSA represents the world's only mandatory social media data access regime. The OSA is now being updated to include similar provisions. Both lack legal clarity regarding which independent, non–platform-led data access techniques — such as public-interest scraping and data donations — are permissible, due to **potential infringements** of existing data protection rules and possible violations of companies' terms of service.

This ambiguity must be resolved, especially as public-interest data scraping in the US continues to be litigated in court, and previous legislative efforts in Washington (notably the **Digital Services Oversight and Safety Act** and the **Platform Accountability and Transparency Act**) failed to secure a critical mass of bipartisan support.

Data access mechanisms, whether regulatory or voluntary and administered by platforms, often prohibit independent data collection techniques within companies' terms of service. This exposes researchers to potential legal liability, highlighting a fundamental tension between privacy concerns and the need for social media transparency and accountability.

If platforms can take legal action against public-interest scrapers for breaching user agreements — independently of any public prosecution under privacy or consumer protection laws — the future of this form of data collection for public-interest research remains uncertain.

Funders have two short-term opportunities to mitigate these issues.

Within existing regulation: Funders can support independent policymaking and advocacy initiatives to clarify when such independent data access research techniques are permitted in line with countries' existing online safety and data protection legislation. They can also promote changes in companies' terms of service — via researcher engagements with regulators and policymakers — to provide greater legal certainty for public-interest research.

Recent examples include the **Social Platforms Data Access Taskforce**, supported by **UK Research and Innovation (UKRI)**, an independent public funding body. The taskforce's remit includes engaging with British regulators and policymakers to address independent data access approaches within the country's existing legislative framework.

Despite similar efforts in the EU, including a **report** from the European Digital Media Observatory that attempts to clarify privacy concerns associated with data access mechanisms, this work is generally absent in the region. Ongoing uncertainty around data donations and public-interest scraping has stalled researchers' use of these techniques within the bloc's DSA.

Within voluntary commitments: Funders can support strategic litigation initiatives aimed at providing clarity for researchers reliant on data donations and public-interest scraping. This is particularly important in countries with no current legal data access mandates.

Previous lawsuits, primarily in the US, already provide a degree of legal certainty. However, a well-structured legal campaign to test the rationale for public-interest scraping — in the name of social media transparency and accountability — across strategically-important jurisdictions would allow researchers confidence to conduct such analysis without fear of being sued for breaching companies' terms of service or violating privacy laws.

It also would resolve the **grey zone between commercial vendors**, which are scraping social media platforms for either advertising or AI-training purposes, and public-interest groups, which are using similar techniques to identify potential harms associated with social media content.

1.2 Wholesale Data Access Infrastructure

Cost ($$$$); Timeframe (Long-term); Funding partner (Public/Private)

Comprehensive and ongoing access to social media data represents a fundamental building block for transparency and accountability. Yet the current patchwork of platform- and non–platform-provided technical infrastructure lacks the economies of scale needed to meet the research community's long-term collective requirements.

What is needed is an underlying technical infrastructure layer — built to world-class security and data protection standards — that functions as a society-wide public good.

This requires a public-private funding partnership to provide long-term, sustained support, likely in the tens of millions of dollars annually (based on costs associated with previous, ongoing, and proposed data access infrastructure tools and repositories), guaranteed over a three- to five-year period, to deliver scalable infrastructure for the global research community.

Private funders can provide initial start-up capital to test proofs of concept for how such community-wide infrastructure is developed, either within regulatory or voluntary data access jurisdictions or, ideally, within both types of jurisdictions to maximize the utility of such research tools. Public funders can subsequently support long-term sustainability for such projects by contributing multi-year financing through transparent and public tendering processes.

Access should be available to all researchers who meet specific vetting requirements, either as defined within existing regulation or by the individual infrastructure provider. An international standard (see below) can facilitate a common definition for data access. Such a shared structure would ensure the highest data protection and security standards are maintained, even for researchers lacking personal expertise in these areas, thereby supporting robust data management practices.

Wholesale data access infrastructure would mitigate potential security and privacy concerns by centralizing these issues, allowing often non-technical researchers to focus on individual outputs rather than on maintaining datasets in a secure environment.

Centralizing data access infrastructure does carry potential privacy and security vulnerabilities, such as exposure to systemic breaches or misuse of sensitive information. However, participants determined that these risks are lower than in a decentralized approach, where privacy and security safeguards would be outsourced to individual researchers. A middle-path approach could combine centralized data access protocols with decentralized data storage within individual research projects.

At its heart, such infrastructure would allow faster, more secure and quantifiable research into the role that digital platforms play within society.

Existing social media data access projects — including those being developed by the Leibniz Institute for the Social Sciences in Germany and Princeton University's Research Accelerator, as well as existing offerings from the National Conference on Citizenship and University of Michigan — provide a groundwork for what can be achieved.

Long-term funding to sustain, connect, and scale such infrastructure projects — most likely through a combination of public support, private donors, and company involvement — remains absent. This represents a clear opportunity for public-private partnerships to support the creation of digital public infrastructure, which is fundamental to overlapping national, regional and global accountability and transparency efforts.

This report does not make a judgment on which initiative should receive support. Instead, its recommendations prioritize wholesale technical data access infrastructure as a foundational building block for quantifiable independent research and for the additional funding opportunities outlined below.

1.3 Research Safe Harbors and Legal Liability Fund

Cost ($$); Timeframe (Short-term); Funding partner (Private)

Almost all countries lack regulatory mandates for social media data access. This gap creates widespread legal challenges for independent researchers, who may face lawsuits from companies for violating terms of service in their pursuit of platform-based data.

Even within the EU and the UK — where mandatory access is already available or will soon be implemented — researchers still face legal uncertainty related to their independent data access techniques.

To address these shortfalls, funders have two immediate options.

Research Safe Harbors: Support policymaking and advocacy projects to establish safe harbors for independent public-interest research within existing legislative norms. These lawmaking efforts — either at a subnational, national or regional perspective — focus on removing legal liability from specific research projects that involve the collection of social media data if those efforts solely focus on:

- Publicly-available datasets;
- Automated data collection, user data donations, and research accounts designed to test platforms' responses;
- Public-interest research (and not those with a commercial focus); or
- Privacy-conscious research techniques that reasonably protect individuals' rights.

Legal Liability Funds: Underwrite the creation of community-wide legal support mechanisms to provide assistance for researchers when they face lawsuits resulting from public-interest data access projects.

Such a fund, based on existing programs for **investigative journalism** and for **social media researchers** targeted by governments or social media companies, would pool resources from multiple research projects as a form of insurance policy. Implementation would likely occur within specific jurisdictions due to the complexity of these legal support mechanisms. It could also be applied in countries with either voluntary or regulated data access mechanisms. Additional examples of such support include the **Climate Science Legal Defense Fund** and the work of the **Society For Civil Rights (GFF)**.

Funders could establish stand-alone legal liability funds for which researchers could apply, or require existing grantees to allocate a portion of their overall funding to such a structure. These funds could then be pooled either within a single funding institution or across a broader group of aligned funders.

1.4 International Standards for Researcher Ethics and Data Protection

Cost ($$); Timeframe (Mid-term); Funding partner (Private/Public)

Social media research, based on quantifiable data access, is inherently transnational. To understand the impact of Instagram's algorithm on Spanish teenagers, for instance, researchers must inevitably include posts from leading creators worldwide. Consequently, both regulatory and voluntary data access initiatives involve data subjects beyond the jurisdiction of any single country.

This represents ethical and data protection challenges. Individuals **have different perspectives and expectations of privacy** depending on their location and cultural norms. Additionally, countries have different data protection laws and approaches to research ethics, creating challenges for those conducting social media research.

To ensure complete datasets for research while respecting individual countries' sovereignty, the independent research community needs **collective standards** for who can access social media data, how ethical protocols are upheld across borders and which privacy mechanisms are maintained to comply with domestic data protection laws.

This requires the creation of a multi-stakeholder, internationally recognized ethics and privacy body that works in tandem with the wholesale data access infrastructure described above. Such an organization could address these concerns in a way that maximizes research opportunities while accounting for the privacy and ethical differences between jurisdictions, thereby upholding strong standards for trustworthy data access and use.

Drawn from the research, policymaking and industry communities, this body would be empowered to set global privacy and ethical standards for social media data access.

These standards would provide the foundation for **potential multilateral agreements**, akin to what already exists for **medical clinical trials research**, which over decades have developed cross-border harmonization to allow international teams to work together on public-interest research.

In social media research, such collective standards would allow separate wholesale data access infrastructure providers to become interoperable, relying on near-universal approaches to data protection, security and ethical issues.

They would aim to mitigate the inevitable privacy and ethical concerns arising from the transnational nature of research by providing an independent mechanism to align domestic research needs with international data protection and ethics standards. This approach would also reduce compliance costs for companies, which would need to adhere to a single standard rather than navigating varying data access requirements across individual jurisdictions.

Recommendation 2: Building Best Practices for the Research Community

Columbia-Hertie working group participants highlighted a wide range of technical, capacity and research methodological approaches used across the research community for accessing social media data. This diversity is expected, given the varying techniques currently in use, including platform-led APIs, virtual "clean rooms," social media user data donations and public-interest scraping.

Discussions revealed that many independent researchers lack the capacity and resources to meet the data protection and security protocols necessary for safely accessing social media companies' data. This includes insufficient awareness of legal requirements such as **Data Protection Impact Assessments**, limited technical knowledge of how to collect, store and manage data through privacy- and security-enhancing mechanisms and inadequate funding to use best-in-class tools to ensure compliance.

For regulators overseeing existing or future data access regimes, these rules are viewed as much a data protection exercise as they are an effort to facilitate social media data access.

Significant asymmetries have emerged — both within sectoral and geographic research communities — between those who already have access to social media data (through regulated or voluntary mechanisms) and those who do not. This gap is particularly acute for researchers from disciplines that do not regularly engage with, or have not historically accessed, potentially sensitive data, as well as for researchers from the Global Majority.

Structured community engagement — aimed at overcoming existing researcher asymmetries and fostering ongoing capacity-building to upskill researchers — represents an area where targeted public-private funding could yield significant benefits.

Below is an overview of potential funding opportunities to address these gaps and promote best practices within the research community.

2.1 Standardized Legal and Regulatory Documentation

Cost ($); Timeframe (Short-term); Funding partner (Private)

The current social media research landscape does not scale to meet the public good associated with broad social media accountability and transparency efforts.

To address this gap, funders have two short-term options.

Both focus on overcoming the current bespoke nature of social media data access so that the collective research community can draw upon common legal and regulatory documentation without undermining individual researchers' independence or objectives.

Standardized legal documentation: Funders can support the creation of community-wide data management and data protection impact assessments (DPIAs) required to access data via regulated and voluntary data access regimes (through mechanisms overseen by platforms).

These legal documents detail the protocols organizations must have in place to secure both public and private social media data. Currently, many groups do not have these documents due to resource constraints and the absence of sufficiently standardized practices. Developing and providing standardized templates, draft MOUs and legal support to tailor such documents to individual organization's needs would galvanize greater use of both regulated and voluntary data access opportunities.

Standardized regulatory documentation: Within the DSA data access regime, researchers face **ongoing difficulties** in successfully applying to platform-led data archives because of bureaucratic application procedures and confusion about how to submit such applications. Applying for social media datasets across multiple companies requires submitting separate forms that request different information from researchers.

Spotlight 2
A "Mass Request" for Data Access

AlgorithmWatch, the Mozilla Foundation and the DSA40 Data Access Collaboratory are working on a "**mass request**" that aims to grant multiple organizations access to daily lists of viral posts in each EU member state, as part of their collective push for widespread use of the DSA's data access provisions.

To fast-track access within existing regulated regimes, funders should prioritize initiatives that prompt the collective development of standardized submission documents, ensuring that researchers do not have to reinvent the wheel when submitting social media data access requests.

This work has already begun via the **DSA 40 Collaboratory**. However, ongoing support is needed to expand this approach, reduce existing barriers within the EU's regulated data access regime and promote the adoption of universal application documents across the wider research community.

2.2 Standardized Data Protection and Security Protocols

Cost ($$); Timeframe (Short-term); Funding partner (Private/Public)

Standardized legal and regulatory documentation provides a foundation for fostering best practices within the research community, but this approach must be paired with ongoing capacity building and training to ensure that all researchers accessing social media data — whether through regulated or voluntary regimes — uphold the highest standards of data protection and security.

Currently, existing capacity building does not meet collective needs. Funders should link the development of standardized documentation with on-boarding requirements, so that grantees demonstrate sufficient understanding of data protection and security protocols and can indicate how they comply with these ongoing requirements.

This has become an urgent priority amid social media companies' criticism that researcher data access may undermine firms' security, privacy, and intellectual property standards. Platforms **highlight** potential deficiencies in researchers' privacy and security practices — alongside risks to firms' intellectual property — as a key consideration for limiting independent access to internal datasets.

To address these concerns, funders can support community-wide capacity-building programs alongside the use of standardized legal and regulatory documentation.

Such initiatives already exist within academia, such as those associated with the European Research Council where data protection and ethics standards are often baked into research methodologies (though these differ between jurisdictions, institutions and research disciplines). Few, if any, similar cross-organization capacity-building efforts exist within independent advocacy organizations, where understanding and resources to meet data protection and security standards for quantifiable research vary widely.

Funders can therefore play a galvanizing role in promoting best-in-class standards by linking funding applications and opportunities to the completion of funded data protection and security training programs.

This could include incorporating such programming into existing grants or creating a separate funding source — potentially shared among like-minded funders — to support community-wide capacity building through workshops organized by trusted and reputable third-party organizations.

2.3 Outreach and Onboarding New Researchers

Cost ($); Timeframe (Short-term); Funding partner (Private/Public)

Current data access opportunities are limited to a small number of researchers who possess the necessary technical skills, platform relationships or regulatory understanding to make use of existing regulated and voluntary mechanisms.

A wider range of research organizations — including those working in sectors such as climate change or LGBTQ+ rights, or those based in different geographies, particularly within the Global Majority — have yet to play a significant role in ongoing accountability and transparency efforts.

Many of these groups lack sufficient resources, technical capacity or awareness of available data access mechanisms. More seasoned researchers can also lack the requisite skills, due to the rapid change in technical advances related to data access, leaving them less well positioned to advocate for or implement cutting-edge data access techniques.

Funders can meet this need via short-term grants and fellowships to train researchers — potentially drawn from existing grantees — in the latest data access techniques and policymaking discussions relevant to their respective research priorities. They can also facilitate mentoring schemes with other grantees with existing knowledge and expertise of social media research best practices.

This can be combined with the separate data protection and security training, described above. It can also include direct collaborations with existing research organizations (potentially via in-person or remote secondments) that have more advanced awareness of the opportunities and challenges posed by accessing social media data.

Such peer-to-peer learning (either through funders' existing portfolio of grantees or organizational relationships with aligned funders) would generate mutually reinforcing benefits. Recent examples include a cross-organizational "**hackathon**," organized by AlgorithmWatch, Mozilla Foundation and the DSA 40 Collaboratory, to allow researchers to share best practice and expertise on using data access mechanisms within the DSA.

Sharing best practices with such initiatives would help reduce existing asymmetries in organizations' ability to access and utilize regulated and voluntary data access regimes. It would amplify funders' existing investments by leveraging those grantees with existing data access awareness for other groups in funders' portfolios without such expertise and capacity.

2.4 Data Access Community Engagement and Coordination

Cost ($$); Timeframe (Mid-term); Funding partner (Public/Private)

Across the regulated and voluntary data access landscape, different parts of the independent research community are pursuing often overlapping efforts to boost transparency and accountability. There is a clear need to better coordinate these separate initiatives to maximize the opportunities provided by wholesale data access.

Current challenges include the need to overcome difficulties in uniting research agendas to make the most of, and share input on, social media access mechanisms. Engagement efforts should include reaching out to regulators, policymakers and new researchers, as well as managing shared risks and coordinating on emerging opportunities.

Funders could support grassroots community engagement to facilitate information sharing, while maintaining individual organizations' separate research, policy and advocacy aims. The goal of this work is to break down current cross-organizational barriers, as described above, without creating a centralized structure for data access research that could introduce unnecessary bottlenecks.

Support for in-person workshops — in some cases hosted on a rotating basis among aligned organizations — would provide opportunities to share best practices, raise awareness of the data access landscape and foster informal networks of researchers willing to collaborate on respective data access projects.

For instance, as part of the Columbia-Hertie initiative, a meeting aimed at overcoming data access fragmentation was held in May 2025, during which representatives from academia, civil society, and data access infrastructure providers outlined common areas for coordination and utilization of existing and emerging data access regimes. This gathering served as a companion discussion to the Columbia-Hertie initiative's main convenings.

A subsequent workshop will take place during the **Mozilla Foundation conference** in November 2025. Ongoing funding support to a range of institutions to bring researchers together for future events (often at the sidelines of existing gatherings) would build on such community knowledge sharing.

Recommendation 3: Fostering Cooperative Researcher-Regulator Relationships

Participants underscored that relationships between regulators and researchers were entering unchartered territory around social media data access. Both sides are being drawn closer together as regulators within regulated regimes enable researchers to analyze datasets. This creates potential opportunities for greater cooperation in the name of accountability and transparency, but it may also lead to skewed power dynamics in how researcher–regulator relationships evolve.

These complex researcher-regulator relationships will only intensify in an environment where platforms and certain governments are **pushing back** against online safety legislation, and where regulators are relying on civil society to publicly "defend" such regulation.

Regulators highlighted that obtaining feedback from researchers on how data access mechanisms and tools support the research community remains a challenge. Officials acknowledged that not all research using data access methods to track social media platforms' roles in society aligns with their regulatory priorities. They also noted a "translation" issue, whereby researchers are often unaware of regulatory priorities and/or do not provide evidence in ways that can inform agencies' enforcement work.

Another bottleneck was the ability of researchers and regulators to access the most up-to-date and useful information provided by the research community based on social media data access. Most regulators have a mandate to engage with external stakeholders, including the research community, for this purpose. Yet, many of these relationships are ad hoc, informal, opaque and not scaled to meet ongoing needs.

There are also significant differences in the informal information-sharing networks that better-resourced agencies are able to maintain with the research community compared with less-resourced regulators, which often lack the knowledge and connections to access those conducting the most cutting-edge research.

For researchers focused on regulated and voluntary data access mechanisms, questions around long-term funding support for work associated with online safety efforts at the national and regional levels remain unanswered. Likewise, questions persist about how researchers' outputs will feed into regulatory priorities and enforcement actions.

To ensure transparent, equitable and mutually supportive relationships between regulators and the research community, there is a need to improve understanding and cooperation between both sides in ways that maintain their independence and distinct objectives. Simply put, participants voiced concerns that funding provided by governments or platforms could compromise the work of researchers, civil society organizations and other entities responsible for holding platforms and regulators to account.

Insulating researchers from the risk of capture requires prioritizing funding sources that do not compromise researcher independence while assisting groups to build out fundraising capacities over time. That includes structuring potential funding using independent bodies that adhere to best practices such as peer review and conflict-of-interest disclosures.

Ensuring that regulators themselves are not unduly influenced by platforms also requires deeper scrutiny of existing ties and communication channels between these parties, as well as the exploration of multilateral initiatives that facilitate the exchange of best regulatory practices in an accountable and transparent manner.

Below is an overview of potential funding opportunities to address these gaps and foster mutually beneficial researcher-regulator relationships.

3.1 Information Sharing Between Researchers and Regulators

Cost ($$); Timeframe (Long-term); Funding partner (Public)

Researchers and regulators need more formalized information-sharing mechanisms to maximize the opportunities presented by social media data access. These relationships must preserve each side's independence at a time of increasing political scrutiny over how online safety rules are implemented. This applies to both regulated and voluntary data access mechanisms.

Regulators already engage with the research community through stakeholder initiatives, including formal structures such as **Ofcom** — the UK's online safety regulator — the **advisory council** of Germany's Digital Services Coordinator and the **European Commission**. This engagement provides regulatory feedback to independent researchers, as well as quantifiable external evidence to support regulatory objectives and potential enforcement actions. In the US, engagement with regulatory agencies typically **takes the form** of responses to Requests for Information, appointments of researchers to official federal advisory committees or participation in informal workshops.

Yet these existing researcher–regulator relationships need to be institutionalized at both the national and regional levels. Such interactions should clearly outline regulatory priorities for the research community and facilitate the sharing of specific, actionable information of "evidentiary value," without undermining the independence of either regulators or researchers, through external review processes and mandatory disclosure requirements.

The Columbia-Hertie initiative — based on transparent, regular, and actionable interaction between regulators, researchers and infrastructure providers — offers a test case of how such future engagement can be addressed. This includes establishing both informal and formal channels of interaction between researchers and regulators to foster mutual understanding of each side's priorities within a neutral setting where all participants have equal standing. Another example is the Institute for Data, Democracy and Politics (IDDP) at The George Washington University, which convened a **Council for Pervasive Data Ethics** to advance ethical practices in social media research.

Funders should continue to provide long-term support to facilitate such neutral convenings. All parties within the data access ecosystem must have an equal voice in ongoing discussions, and any funding must remain separate from potential regulatory enforcement actions.

This peer-to-peer approach supplements **formalized consultation processes** through which independent researchers can contribute to the regulatory and enforcement priorities of specific agencies.

It can be directed by domestic agencies seeking greater input from respective countries' research communities or implemented across jurisdictions — potentially through coalitions like the **Global Online Safety Regulators Network** — to foster long-term relations between researchers and regulators to address the "translation" dilemma described above.

3.2 Funding Support for Independent Data Access Research

Cost ($$$$); Timeframe (Mid-term); Funding partner (Public)

The funding environment for independent social media research **has become acute**. Public support from the United States government has declined, and many philanthropic donors have shifted their priorities to other areas of digital policy. European public programs are either under-resourced, require significant allocations for compliance and reporting or are too small to meet researchers' needs.

These shifting funding dynamics come at a time when societal demand — for greater accountability and transparency in social media, both in countries with online safety laws and those without — **has grown exponentially**.

What is required are more public funding opportunities to support flexible data access research projects linked to specific regulatory priorities (based on improved and ongoing regulator-research interactions), as well as open-ended tenders for wider research associated with online safety, platform governance and social media accountability and transparency. Researchers are not only creators of knowledge and educators, but also function as **policy advisors, watchdogs, and social innovators** who **can shape and clarify** the construction and implementation of specific policy and regulatory priorities.

Such additional funding support must be administered through neutral public intermediaries and tendering processes to avoid perceived or actual political or regulatory capture.

Similar programs already exist within the likes of **Horizon Europe** and the **Economic and Social Research Council**, but these opportunities primarily provide public support to academics rather than civil society groups. They also impose levels of bureaucracy on grantees that make it difficult for smaller organizations to apply. Existing public funding opportunities lack the flexibility needed for independent researchers to adapt to rapidly evolving regulatory and policymaking priorities.

What is required are long-term, flexible public support structures — delivered through open, transparent and agile tendering processes — that enable the independent research community to meet ongoing societal, regulatory and policymaking needs related to platform data access. These mechanisms must preserve researcher independence while promoting equitable regulator–researcher relationships.

Funding opportunities should be divided into two buckets. The first includes support for long-term infrastructure (as described in Recommendation 1) to enable the maintenance of community-wide public goods. The second includes grants for individual projects, covering both funding for specific research topics and infrastructure-related costs (e.g., usage fees associated with infrastructure, consulting fees, commercial scraping fees, etc.).

Policymakers **should prioritize** such funding support in the upcoming EU budget negotiations for the **2028-2034 Multiannual Financial Framework**. Public funding for independent data access research aligns with the 27-country bloc's stated dual priorities of upholding democratic values and jumpstarting the EU's industrial competitiveness.

Other countries, particularly the UK, whose separate online safety regime now includes prospects for regulated data access, should similarly expand existing funding structures through public bodies such as UKRI to better align national policymaking objectives with the current funding shortfall for independent social media research.

3.3 Formalized Researcher Secondments and Fellowships

Cost ($); Timeframe (Mid-term); Funding partner (Public)

To overcome existing gaps between researchers and regulators, participants highlighted the opportunity to expand and create formalized secondments and fellowships that embed researchers within regulatory or government agencies. These initiatives would provide individuals with a greater understanding of how public bodies operate, enabling more effective research outputs by academic and civil society groups.

These secondment and fellowship opportunities are currently more common in Anglo-Saxon countries than in the EU. Current examples include the **Intergovernmental Personnel Act** and **TechCongress** in the US, as well as the **Royal Society Pairing Scheme** and **British Academy Fellowships** in the UK. Some participants noted that such formalized structures are not currently feasible within the legal frameworks of certain EU countries.

Nevertheless, short-term secondments and fellowships — whether publicly funded or supported by researchers' institutions — could provide long-term benefits by helping to overcome the regulator-researcher "translation" dilemma described above.

They would allow researchers to gain real-world regulatory and policymaking experience that could directly inform outputs once they return to academia or civil society. For less-resourced government agencies and regulators, these opportunities would also offer the advantage of accessing world-class expertise to meet immediate capacity needs without requiring these public bodies to hire full-time employees.

3.4 Engagement with National Data Protection Regulators

Cost ($); Timeframe (Short-term); Funding partner (Private)

Within regulated data access regimes, the majority of researcher–regulator engagement has so far focused on interactions with online safety regulators, such as the European Commission, EU member countries' Digital Services Coordinators and Ofcom (although the UK's data access regime has yet to be established despite **legislation** directing such a scheme). These agencies have led the implementation of mandated data access regimes or undertaken work to establish them, and they have therefore been prioritized by the research community for engagement.

Spotlight 3
Data Access in Action

"The Case for Transparency: How Social Media Platform Data Access Leads to Real-World Change" from the Institute for Data, Democracy & Politics at The George Washington University argues for greater transparency by showcasing notable social media research projects and the changes they have prompted in platform operations, regulatory efforts and academic initiatives.

To scale the benefits of independent researchers' social media data access for transparency and accountability, private funders need to support similar engagements with national data protection regulators. These regulators oversee countries' separate privacy legislation, which will increasingly overlap with online safety efforts such as the DSA and the OSA. This includes engagement by the **Social Platforms Data Access Taskforce** with the UK's Information Commissioner's Office to clarify how data donations and public-interest scraping would be permissible under the country's current data protection legislation.

As outlined in Recommendations 1 and 2, many of the current data access gaps relate to open questions around data protection measures associated with social media data access. Addressing these gaps requires input from national privacy regulators — in conjunction with their national online safety counterparts — to ensure that any infrastructure, legal documentation and community-wide training align with these agencies' respective regulatory oversight.

As part of this short-term funding opportunity, particular emphasis in these policymaking and advocacy engagements should be placed on clarifying legal issues related to independent data access mechanisms, such as social media user data donations and public-interest scraping.

In a **recent data access report** from Ofcom, for instance, the agency suggested that clarifying existing legislation around these two separate data access mechanisms could be a priority for the UK government's efforts to establish a mandatory data access regime. Funders can facilitate such discussions between researchers and regulators, including within the EU, where national data protection agencies have so far not engaged with the privacy-related implications of researcher data access.

3.5 Demonstrations and Case Studies For Data Access

Cost ($); Timeframe (Mid-term); Funding partner (Private/Public)

To facilitate greater awareness among researchers and regulators of the role social media data access plays in transparency and accountability efforts, a public repository should be created to showcase best-in-class research. This would allow researchers to learn from the existing body of work and provide regulators with proven examples of the utility of such data access for their respective oversight roles.

This funding opportunity brings together private donors, whose existing grantees already conduct much of this public-interest transparency work, with public financing associated with the more flexible funding mechanisms outlined in Recommendation 3.

Taken together — and aligned with other tools like the **DSA Transparency Database** and the **DSA Data Access Portal** — such public-private support can provide the public, the research community, regulators and policymakers with demonstrable and quantifiable evidence for why data access remains a crucial foundation for accountability and transparency in the social media age.

Conclusion

The Columbia-Hertie initiative focused on identifying current gaps in researcher data access, as well as public-private funding opportunities to address those needs. The recommendations were derived from input from more than 120 stakeholders drawn from regulatory agencies, academia, independent research and advocacy groups and data access infrastructure providers.

The report represents the most comprehensive review of current global researcher data access efforts, based on input from participants actively engaged in various associated projects worldwide. It outlines the necessary steps to improve what are currently disparate and under-resourced efforts to support digital platform governance and transparency through both regulatory and voluntary data access mechanisms.

Over the course of this work, it became clear that no single solution could achieve this goal of holding both companies and regulators accountable for upholding online safety principles and fundamental rights during a period of unprecedented technological change.

If such a solution existed, it would have already been implemented.

Participants also emphasized that different aspects of researcher data access all require support. These range from digitalinfrastructure as a public good and legal documentation to be shared across the community, to formalized information-sharing structures with regulators, and more flexible public funding mechanisms for independent researchers.

As of late 2025, this remains a challenging goal.

The current philanthropic funding landscape has shifted priorities away from platform governance toward other digital policy areas. Some countries, such as the US, have reduced long-term support for independent research, while others, including those within the EU, have prioritized research linked to economic competitiveness rather than digital regulation.

Given the widespread and varied funding needs of the research community outlined in this report, it is unlikely that all of its recommendations can be supported solely through private philanthropy or public institutions. Yet the case for reinvigorated support for researcher data access — as a linchpin of democratic resilience and the promotion of fundamental rights — is now more urgent than ever.

Social media platforms play a central role in society. They connect us daily to friends and family. They are increasingly ground zero in political communication campaigns. They are routinely targeted by malign actors to undermine democratic norms at a time when domestic polarization — across almost all democratic countries — **has reached record levels** in the post-World War II era.

Given these political headwinds, the independent research community **has less quantifiable insight** into platforms' role in society than ever before.

Almost all of the companies **have scaled back their trust and safety efforts**. Countries like the US, and increasingly other Western democracies, openly criticize efforts to boost digital platform governance standards. Many within the research community face attacks for their public-interest work addressing online harm, illegal content and other platform-related issues that fundamentally affect democratic resilience and long-standing societal norms.

The question is not whether public-private funding should support researcher data access to tackle these existential challenges in a society buffeted by the dual dilemmas of polarization and unprecedented technological change.

The real question is whether private philanthropy and public institutions can afford to miss the opportunity to invest in public goods that enhance transparency and accountability for digital platforms — central to 21st-century life but still largely opaque to wider society.

This report provides practical pathways to meet that challenge.

We invite public and private funding organizations to engage with Columbia World Projects, the Hertie School's Centre for Digital Governance, and working group participants to turn those opportunities into a reality.

Appendices

Appendix 1: Working Group Participants

Working Group Chair: Mark Scott, Atlantic Council

Working Group Members:

David Alders Stiftung Mercator
Chris Beall Project CONNIE
Niamh Clarke Coimisiún na Meán (CNaM)
Claes de Vreese University of Amsterdam
Mark Dempsey Article 19
Kate Dommett University of Sheffield
Maria Donde Coimisiún na Meán (CNaM)
Mauritius Dorn Institute for Strategic Dialogue
Lishu Gang Reset Tech
Brandi Geurkink Coalition for Independent Technology Research
Sam Gill Doris Duke Foundation
Niamh Hanafin Coimisiún na Meán (CNaM)
Cameron Hickey National Conference on Citizenship
Julian Jaursch Federal Network Agency (BNetzA) German Digital Services Coordinator
Marley Kalt University of Michigan | ICPSR
David Karpf The George Washington University School of Media & Public Affairs
David Klotsonis Centre for Democracy & Technology Europe
Ramya Krishnan Knight First Amendment Institute at Columbia University
Felix Kröner Reset Tech
Paddy Leerssen University of Amsterdam | Faculty of Law
Anna Lenhart The George Washington University Institute for Data Democracy and Politics (IDDP)
Benoît Loutrel Autorité publique française de régulation de la communication audiovisuelle et numérique (ARCOM)
Josephine Lukito University of Texas at Austin Moody College of Communication
Oliver Marsh AlgorithmWatch
Kirsty Park Coimisiún na Meán (CNaM)
Courtney Radsch Open Markets Institute | Center for Journalism and Liberty
Buse Raziye Çetin AI Forensics
Hilary Ross Global Network Initiative
Andrea Sanders-Winter Federal Network Agency (BNetzA) | Internet, Digitalization and Market Analysis
Anya Schiffrin Columbia University | School of International and Public Affairs
LK Seiling Weizenbaum Institute | DSA40 Data Access Collaboratory
Brandon Silverman The George Washington University Institute for Data Democracy and Politics (IDDP)
Amanda Starling Gould Partnership for Public ServiceAlison Sweet, University of Michigan | ICPSR
Rebekah Tromble The George Washington University Institute for Data Democracy and Politics (IDDP)
Katrin Weller GESIS – Leibniz-Institute for the Social Sciences

The report reflects a diversity of perspectives, and the views expressed should not be attributed to any individual participant or organization. Affiliations are current as of May 2025.

Appendix 2: Matrices for the Implementation, Monitoring and Evaluation of Recommendations

The following appendix presents matrices for the implementation, monitoring and evaluation of these funding recommendations in order to provide a structured framework to help funders, regulators and researchers translate the report's proposals into concrete action in support of social media data access, transparency and democratic resilience.

Matrix A: Supporting Underlying Data Access Infrastructure

Workstream	Top-Line Indicators	Suggested Funding Model	Review Cycle
Open-Source Tooling	Existing repository lists, usage dashboards	Sliding-scale cost-sharing across EU member states	Quarterly review
Initial Infrastructure Setup	% of EU member states integrated, infrastructure uptime, request latency	Sliding-scale cost-sharing across EU member states	3-5 years to complete
Researcher Access Rates	# of access requests submitted, approval rates, median response time, infrastructure access logs (e.g., GESIS, ICPSR)	Public–platform co-funding	Quarterly review
Data Coverage	% of major platform datasets included, frequency of updates, metadata schema completeness	Co-funded through public + platform agreements	Bi-annual review
Scaling Efficiency	Average cost per researcher per dataset, duplication rate across infrastructure	Public-philanthropic consortia	Annual review
Legal Risk Mitigation	# of legal consultations provided, # of incidents handled, # of researchers covered by indemnity	University legal clinics + pooled liability fund	2-3 years to complete
Strategic Litigation	# of precedent-setting legal cases, # of settlements, areas of legal ambiguity resolved	Law school partnerships; public-interest legal organizations	Needs-based

Matrix B: Building Best Practices for the Research Community

Workstream	Top-Line Indicators	Suggested Funding Model	Review Cycle
Participation Growth	# of researchers engaged annually, year-over-year participation growth rate, membership rosters	Academic networks, platform-neutral grants	Annual review
Cross-Disciplinary Engagement	% representation from diverse disciplines, # of interdisciplinary co-authored outputs for publication	Institutional support, ERC collaboration	Bi-annual review
Training Outcomes	# of fellowships and grants awarded, post-fellowship employment in field, satisfaction and impact scores	Co-funded fellowships (e.g., Ofcom, UKRI)	1-2 years to complete
Event Engagement	# of events held, average attendance, post-event survey data	Rotating institutional sponsorship	Quarterly review
Guideline and Toolkit Creation	# of toolkit downloads, frequency of access over time, citation in research outputs	Open-access funders, institutional sharing	Quarterly review
International Standards for Ethics	# of ethics templates adopted, % of applications citing standardized documentation, # of vetting panels trained	Multi-stakeholder governance	1-2 years to complete
Researcher Directory or Registry	# of vetted researchers listed, average time to verification, stakeholder use frequency	Consortium-managed	Year 1 pilot, quarterly review

Matrix C: Fostering Cooperative Researcher-Regulator Relations

Workstream	Top-Line Indicators	EU Regulators	Non-EU Regulators	Suggested Funding Model	Review Cycle
Collaboration Instances	# of joint research initiatives launched, # of advisory appointments, # of recurring regulatory–academic working groups	Digital Services Coordinators (DSCs), EC Expert Groups	Ofcom, US/CAN bodies	Public institutional co-support	Medium term (1–3 years)
Researcher Training and Fellowships	# of researchers placed in regulatory bodies, satisfaction & retention rates, post-placement collaborations	DSA authorities	UKRI, US Intergovernmental Personnel Act, Canadian equivalent	Co-funding with host agencies	Short-term (0–12 months)
Standardized Legal Documentation	# of legal templates published, # of uses in access applications, % of regulator alignment with templates	EC, national authorities	US/UK/CAN authorities	Grant-funded then institutionally maintained	Short-term (0–12 months)
Cross-Border Convenings	# of convenings held, % of new participants, regional and disciplinary diversity	Hosting rotation	Global Online Safety Regulators	Rotating sponsorship model	Quarterly or biannual
Feedback Loop Channels	# of structured feedback channels launched, update frequency, % usage by researchers, platform analytics, survey data (e.g., EC landing page for engagement; site visits as indicators)	EC transparency boards	N/A	EC-led with external platform integration	Ongoing
Trust and Satisfaction Metrics	% of researchers and regulators reporting satisfaction with engagement, independent evaluations, # of disputes or delays	European Commission & DSCs	US/UK/CAN equivalents	Independent oversight (nonprofit evaluator)	Annual review