

From the Wild West to the Walled Garden

The Evolution of Twitter/X Data Access for Research

Fabio Giglietto · University of Urbino · orcid.org/0000-0001-8019-1035

Cornelius Puschmann · ZeMKI, Universität Bremen · orcid.org/0000-0002-3189-0662

Published in: *M/C Journal*, Vol. 29, No. 2 (2026) — Issue theme: twitter.

How to cite: Giglietto, F., & Puschmann, C. (2026). From the Wild West to the Walled Garden: The Evolution of Twitter/X Data Access for Research. *M/C Journal*, 29(2). <https://doi.org/10.5204/mcj.3257>

DOI: <https://doi.org/10.5204/mcj.3257>

Original publication date: 21 April 2026.

© 2026 Fabio Giglietto & Cornelius Puschmann. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Licence (CC BY-NC-ND 4.0): <https://creativecommons.org/licenses/by-nc-nd/4.0/>. Authors retain copyright and are permitted to self-archive the published version in institutional repositories and on personal sites, with acknowledgement of initial publication in *M/C Journal*.

Introduction

The proliferation of large online platforms has fundamentally reshaped societal processes, from political communication and civic interaction to the mundane organisation of everyday social life (Jungherr et al.). The large volumes of digital traces and interactions generated on these platforms offer researchers an unprecedented opportunity to study both contemporary society and how these digital environments shape it, enabling analysis at a scale and granularity that would otherwise be unattainable. However, accessing such data has been a persistent challenge, as a result of competing interests involving user privacy, platform commercial incentives, and the public's right to scrutiny (Weller). The development of academic data access has been uneven, evolving from informal, ad hoc collaborations to the present more regulated era (Darius et al.).

It is important to regard questions of data access in relation to large online platforms against the background of their outsized societal importance and the risks and benefits associated with their widespread use. The present policy initiatives that have transformed how researchers gain access to and work with platform data are a reaction to prior modes of working with such data, which were largely unregulated. A number of widely publicised scandals, such as the Cambridge Analytica (Hu) scandal and the OkCupid data breach, prompted public discussion about how to counter potential misuse, including in the context of scientific research (Puschmann and Bozdag). At the same time, because so many societal processes, political communication, and civic interactions take place through platforms, they are of key importance for a functioning democracy. Therefore, a middle ground between preserving user privacy and allowing researchers to analyse data that is of key importance for understanding society and safeguarding against systemic threats had to be found.

This article traces the evolution of the technical and regulatory frameworks for accessing social media data for research purposes, with a particular focus on the trajectory of Twitter (now X). While we centre our analysis on Twitter/X, its trajectory is far from unique. Meta's shuttering of CrowdTangle in 2024—a tool that had become indispensable for researchers studying the spread of content on Facebook and Instagram—removed a key transparency resource with little prior consultation of the research

community (Bruns). Reddit’s controversial decision in 2023 to impose steep API pricing effectively ended many third-party research tools, while TikTok has never offered meaningful programmatic data access for researchers outside of limited partnerships. These parallel developments underscore that the story we trace here is not idiosyncratic to one platform but reflects a structural pattern across the platform economy. The history of Twitter serves as a powerful case study, illustrating the interplay between technological infrastructure, corporate strategy, and regulatory oversight. We provide a brief account of the early, unregulated “Wild West” of data access—a term used to describe the largely ungoverned landscape of social media research prior to major platform restrictions (Puschmann, “An End”)—, the rise and fall of platform-led initiatives, and the recent, transformative impact of comprehensive legislation like the European Union’s Digital Services Act (DSA). The closure of open APIs, which Bruns termed the “APIcalypse”, signalled a turning point in this trajectory. Ultimately, we argue that the path to data transparency is not a linear progression but uncertain terrain, where progress can be fragile and subject to the shifting priorities of powerful platform owners.

Of Data Philanthropy and Ad Hoc Partnerships

Prior to the late 2010s, the landscape of academia-industry collaboration was characterised by a model best described as “data philanthropy”. In this paradigm, platforms selectively shared data with researchers based on their own criteria, often through informal relationships or one-off initiatives. These early partnerships were not governed by standardised rules but by the goodwill and strategic interests of the companies themselves, leading to a system that was often criticised for its lack of transparency and potential for bias. This occurred either as the result of a formalised call for participation—often in reaction to a critical event, such as suspected election manipulation or the spread of hate or misinformation—or without a formalised call, as the result of mutual contact and shared academic pedigree (Darius et al.). Both variants were criticised as being intransparent and subject to biases and conflicts of interest on the part of platforms.

Over time, the informal model, in which informal ties between industry researchers and academics—often on the basis of mutual acquaintance or shared academic pedigree—were the basis for collaboration, gave way to the more formalised variant in which a process was designed that resulted in data sharing based on a selection process of project applications. This shift over time mostly occurred as a reaction to public objections, as the collaborations built mostly on social ties and shared research interests were widely criticised. In some cases, these initiatives of company founders or senior management were viewed as favouring specific individual researchers. However, the model that replaced this, which involved some level of scrutiny through an open call, was still often regarded as quite intransparent because the selection criteria were not clear.

This era was marked by several high-profile, yet selective, collaborations that highlighted both the promise and the peril of platform data. Facebook, for instance, engaged in deep collaborations with social scientists to conduct massive field experiments. A prominent example is the “61-million-person experiment” on social influence and political mobilisation, published in *Nature* in 2012, which manipulated the News Feeds of millions of users to test the effects of social messages on voter turnout (Bond et al.). A similar, more controversial initiative was the 2014 “Emotional Contagion” study, where the algorithm was manipulated for nearly 700,000 users to test whether emotions could spread through a social network (Kramer et al.). These projects, while yielding significant scientific insights, were relationship-based and not open to the broader research community, raising concerns about preferential access and the power of platforms to shape the research agenda. The model of “independence by permission”, where research operates within parameters set by the studied entity, became a recurring

theme.

Other platforms also experimented with structured data sharing. In 2014, Twitter launched its “Data Grants” program, a pilot initiative that provided a small number of selected research projects with free access to historical public tweet data, which were otherwise prohibitively expensive via resellers. The selected projects were highly issue-specific, with grants awarded to study influenza trends and social media use during flood disasters, marking one of the first times a major platform acknowledged the public utility of their data by formalising a donation mechanism for academia. Similarly, Microsoft Research collaborated with academics to use anonymised Bing search logs for public health research, such as identifying adverse drug reactions missed during clinical trials. Yahoo’s Webscope program, established in 2006, operated like a reference library, offering curated, static datasets—such as language models and user interaction graphs—to university faculty who signed data use agreements, providing a critical resource for information scientists who lacked access to the internal data of search engines.

A case underscoring the limits of digital trace data sharing was the 2006 AOL Research search data release. In an attempt to support the academic community, AOL released a dataset of 20 million search queries from 650,000 users. Despite attempts at anonymisation, journalists at *The New York Times* were able to re-identify individuals based on their search histories. The incident reinforced a move away from open data releases and toward more controlled, selective access models, and underscored the profound risks to individual privacy.

Evolution of Data Access on Twitter/X

The history of Twitter as a data source for academic research provides a microcosm of the broader trends and reversals in the field. Its evolution can be divided into three distinct phases, tracing a path from openness to more formalised partnerships and, ultimately, to a precipitous retreat behind a commercial paywall.

Phase 1: The ‘Wild West’ of the Open API (2006–2020)

In its early years, Twitter was extremely popular as a source of social media data. Its public-by-default nature, combined with a robust and relatively permissive set of developer APIs, created a “Wild, Wild West” environment for data acquisition. The developer APIs were originally intended to foster a third-party ecosystem of clients and apps, but academics quickly discovered their utility for research. During this phase, a large body of scholarship emerged, characterised by experimentation and description, often only loosely grounded in theoretical assumptions. Researchers explored everything from breaking news and political debate to social movements and crisis communication. As researchers identified the platform’s strengths—its simplicity, large data volumes, and focus on public, real-time communication—access remained largely unregulated, and those with the technical skills could collect large datasets with relative ease. This period established Twitter as an invaluable, if chaotic, real-time sensor for social and political life.

Phase 2: The ‘Golden Age’ of the Academic API (2020–2023)

As the field matured and calls for more structured access grew, Twitter took a step towards more structured data access specifically for academic research. In 2020, it launched the Academic Research product track, a dedicated API designed specifically for the needs of researchers. This was a transformative development. Vetted academics were granted free, high-volume access to both real-time and historical Twitter data, a resource that had previously been either technically inaccessible or

financially prohibitive. This change in the access regime was paired with technical advancements in indexing historical data, making the platform's archive searchable. As a result of both the higher volume of data and the sophisticated search capabilities, research based on Twitter data became more advanced. The Academic API represented a new, sanctioned regime for data access, one that explicitly recognised the public interest value of independent research. It streamlined the process, removed significant financial barriers, and enabled a new wave of more sophisticated and methodologically rigorous research. For a brief period, the Twitter Academic API was hailed as a best-practice model for industry-academia collaboration, a seemingly stable bridge between platform resources and public-interest inquiry.

Phase 3: The Musk Takeover and the Commercial Curtain (2023–Present)

This period came to an abrupt end with Elon Musk's acquisition of Twitter in late 2022. The subsequent overhaul of the company, rebranded as X, involved a radical shift in its approach to data access. In early 2023, the Academic API was shuttered, and the teams that supported the research community and policy distillation were disbanded, along with most employees involved in policy and content moderation. Access to the platform's data was consolidated under a new, expensive commercial API pricing structure, with no special provisions for academic or non-profit researchers.

This move effectively returned the state of Twitter data access to one determined by financial resources, erecting a paywall that placed the data out of reach for the majority of the academic community. The decision was likely driven by a combination of aggressive cost-cutting, a desire to monetise the platform's data assets, and a strategic goal to control data for use in Musk's own AI ventures. As a result, research into the platform has declined, a loss compounded by the non-trivial share of users who abandoned the platform following the takeover. The story of X demonstrates that progress in data access is not guaranteed; a system built on platform benevolence can be dismantled overnight by a change in ownership or corporate strategy.

The Regulatory Turn: The Digital Services Act and the Mandate for Access

Against the background of the unpredictability we have described, comprehensive government regulation presents itself as an obvious alternative to access that is a privilege bestowed on researchers by platforms, rather than a right. The most significant development in this domain is the European Union's Digital Services Act (DSA). A key component of the DSA is Article 40, which establishes a legal framework for researcher access to platform data, shifting the paradigm from one of discretionary philanthropy to one of regulatory obligation.

The DSA's approach is a novel attempt to regulate from the outside. It empowers vetted researchers from academic institutions to request access to data from Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) for the specific purpose of conducting research into "systemic risks" in the Union. This includes risks related to the dissemination of illegal content, negative effects on fundamental rights, and the spread of disinformation. The act creates a tiered system, distinguishing between data that are already publicly accessible (Article 40.12) and more sensitive, non-public data (Article 40.4) which require more stringent safeguards.

However, the implementation of this regulatory framework also faces challenges. The gap between regulatory intent and platform execution is substantial. Platforms retain significant control over the data infrastructure, and there is a "structural asymmetry of knowledge" where platforms possess unique, internal insights that researchers lack. As a recent EDMO report highlights, platforms store information

in complex data warehouses with “hundreds or thousands of tables”—structures that even newly hired platform employees struggle to navigate (Motyl et al.). Any tool presenting platform data to external researchers is necessarily designed by the platforms themselves, creating an irreducible dependency: platforms mediate what researchers can see.

Early experiences under the DSA have revealed a highly uneven landscape. Response times to researcher data access requests vary dramatically—from as little as two days to over one hundred—and there is no common standard for how platforms should handle such requests. Some platforms have established dedicated portals with clear documentation, while others lack any identifiable channel for Article 40(12) submissions, forcing researchers to navigate generic email addresses and automated replies. In several documented cases, platforms have denied access using boilerplate justifications invoking necessity and proportionality, with no meaningful pathway for appeal or dialogue—effectively precluding independent scrutiny. Others have conflated the different access tracks under Article 40, redirecting researchers from one mechanism to another without clear justification. Perhaps most strikingly, some platforms have responded to structured data requests by simply whitelisting researchers for Web scraping, framing publicly visible Web pages as adequate “access” while placing the entire burden of data collection, infrastructure, and quality assurance on the researchers themselves. Even when data are provided, they frequently diverge from researcher specifications in structure, completeness, and format, requiring extensive preprocessing that undermines reproducibility and cross-platform comparison. The European Commission has already opened formal proceedings against several platforms, including X, TikTok, and Meta, for suspected breaches of their Article 40 obligations, signalling that enforcement will be a contentious and ongoing battle. As of late 2025, preliminary findings determined that both TikTok and Meta were in breach of their obligations to provide researcher access to public data under Article 40.12. Moreover, the DSA’s reach is inherently limited to the European Union, meaning that researchers based outside the EU have no equivalent legal right to demand platform data. This geographic asymmetry risks creating a two-tier system in which EU-based scholars gain regulated access while colleagues in the Global South and other regions remain dependent on platform goodwill or expensive commercial APIs.

Ideally this creates a productive (if difficult) tension, because while regulation creates the legal right to scrutiny, the quality of that scrutiny still depends on a degree of platform cooperation. An overly adversarial approach risks pushing platforms toward minimal compliance, while a purely collaborative model risks co-optation. The future of data access under the DSA will likely be a negotiated middle ground, a continuous process of requests, refusals, and regulatory interventions that slowly carves out a space for meaningful independent research (Abdo et al.). The framework proposes a triangulation of Proactive Data Interfaces (APIs), Custom Data Requests, and Independent Data Collection (scraping) to maintain this balance, where platforms provide infrastructure while researchers maintain independence through multiple pathways.

Despite its limitations, regulation has produced meaningful infrastructure. The Meta Content Library is a direct outcome of DSA Article 40.12 provisions, providing capabilities that earlier tools lacked. Progress should be measured against what came before: while earlier access tools offered valuable capabilities with significant limitations, the new infrastructure provides greater depth within a more controlled environment—a trade-off rather than unambiguous progress. Regulation also creates accountability structures that shift platform incentives, creating counter-pressure against further restrictions on data access.

The relationship between regulation and platform cooperation is neither pure imposition nor pure voluntarism. The complexity of governing socio-technical systems means no single actor—platform,

regulator, or researcher—can fully predict systemic effects. Academic research has documented effects that internal analyses missed, such as asymmetric impacts of content moderation on different actor types and timing discrepancies between announced policies and actual implementation. External research thus serves not merely accountability but legibility—helping platforms understand what they themselves cannot see. A collaborative orientation requires recognition that understanding algorithmic effects on democratic discourse serves partially overlapping interests across stakeholder communities, even as fundamental tensions remain. As frameworks for researcher data access continue to develop, lessons may also be drawn from adjacent domains: Morten, Nicholas, and Viljoen have argued that the governance of clinical trial data sharing offers instructive parallels for designing sustainable, rights-based models of platform data access.

Beyond regulatory mandates and platform-provided APIs, researchers have increasingly turned to alternative modes of data collection that do not depend on platform cooperation. Data donations, in which users voluntarily share data packages they can request from platforms under the EU General Data Protection Regulation (GDPR)'s right of access, have emerged as a promising if methodologically complex approach. Browser-based data collection tools and citizen science initiatives offer further pathways, enabling distributed data gathering without relying on centralised platform infrastructure. Web scraping occupies a legally contested space: while the US ruling in *hiQ Labs v. LinkedIn* affirmed the legality of scraping publicly accessible data in certain circumstances, the practice remains fraught with jurisdictional variation and platform opposition. As Freelon et al. have argued, the post-API age demands a diversification of methods, and Perriam, Birkbak, and Freeman have shown how digital methods must adapt when the technical conditions that originally enabled them no longer hold. These alternatives are not without their own limitations—data donations suffer from self-selection bias, scraping raises ethical and legal questions, and browser-based tools require significant participant engagement—but together they represent a necessary broadening of the researcher's toolkit in an era of diminishing platform generosity.

Conclusion: Navigating an Uncertain Future for Research

Social media data access for research has been a non-linear and often frustrating challenge. We have moved from an era of informal, ad hoc “data philanthropy”, dependent on the goodwill of platforms, to a brief, promising period of formalised partnership, exemplified by the Twitter Academic API. Today, the landscape is fractured. The closure of academic access to X demonstrates the inherent fragility of voluntary systems, while the DSA represents a powerful but largely untested attempt to mandate transparency through law.

Researchers now find themselves in a more complex and uncertain environment. Access is now bifurcated: for those with financial resources, high-cost commercial APIs provide one path; for others, the legal and bureaucratic channels of the DSA provide another. Neither is a perfect solution. The former excludes the vast majority of the research community, while the latter promises a right to access but is dependent on the slow and contested processes of regulatory enforcement. This means researchers will continue to be dependent on the cooperation of platforms, who, while no longer in the privileged position of philanthropists, will still exert considerable influence over what research can be conducted.

The implications of this bifurcation extend well beyond technical inconvenience. De Vreese and Tromble have described the emerging situation as a “data abyss”—a condition in which the lack of reliable access to platform data leaves both research and democratic society unable to understand the very systems that shape public discourse. This abyss is not evenly distributed. Researchers in the Global

South, early career scholars without institutional resources to fund commercial API subscriptions, and civil society organisations working on platform accountability are disproportionately affected. The current landscape risks entrenching a system in which only well-funded research groups at major institutions in wealthy countries can meaningfully study the platforms that billions of people use daily, while those who may have the most at stake are excluded from generating the knowledge needed to hold these platforms to account.

The lessons from the past two decades extend beyond technical or legal frameworks. When Twitter was launched, it was not taken particularly seriously as a source of data by most academics because its uptake was limited and salient societal debates took place elsewhere. With growing uptake, both by ordinary users and by political, social, cultural, and economic elites, Twitter became increasingly interesting as a source of research data. At the same time, regulatory regimes shifted from a state of unregulation to a more streamlined process, very much as a result of the Cambridge Analytica scandal and assumptions around Russian election manipulation in the context of the 2016 US presidential elections and beyond. For researchers, the regime for accessing Twitter data became markedly better as a result of the formalisation of researcher access: a sanctioned regime that explicitly permitted access to data for research purposes without commercial exploitation, tied to the merit of research projects and adherence to ethical guidelines.

The central lesson from the past two decades is that there is no one-way street toward greater transparency. The technical, commercial, and political whims of platform owners can quickly undo years of progress. As the case of Twitter's sale to Mr Musk shows, such changes can have radical and unforeseen consequences for data access. Researchers must be ready to navigate this complex legal and regulatory landscape at a time when platforms grow ever more politicised as a result of rising geopolitical uncertainty.

Looking ahead, several avenues merit concerted attention from researchers, policymakers, and institutions alike. First, academic consortia for collective bargaining with platforms could provide researchers with greater leverage than individual institutions can achieve alone, establishing shared standards for access agreements and reducing duplication of effort. Second, public investment in independent research data infrastructure—including data trusts or intermediaries that sit between platforms and researchers—could reduce the dependency on platform goodwill that has proven so fragile. Third, researchers should embrace multi-method approaches that combine API-based collection with data donations, browser-based tools, and, where legally permissible, independent data collection, building resilience against the loss of any single access pathway. Finally, regulatory frameworks like the DSA must be extended, strengthened, and replicated beyond the EU, with enforcement mechanisms robust enough to ensure that the right to data access is not merely theoretical. The path forward requires not just better tools or stronger laws, but a fundamental recognition that independent research into platform effects is a public good—one that demands sustained institutional support rather than reliance on the shifting generosity of the platforms themselves.

References

Abdo, Alex, et al. *Better Access: Data for the Common Good*. Knight Georgetown Institute, 2025.

<https://kgi.georgetown.edu/research-and-commentary/better-access/>.

Bond, Robert M., et al. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489.7415 (2012): 295–298. <https://doi.org/10.1038/nature11421>.

- Bruns, Axel. "After the 'APIcalypse': Social Media Platforms and Their Fight against Critical Scholarly Research." *Information, Communication & Society* 22.11 (2019): 1544–66. <https://doi.org/10.1080/1369118X.2019.1637447>.
- Darius, Philipp, et al. "Election Research in the Age of Regulated Data Access under the EU Digital Services Act." *Internet Policy Review* 15.1 (2026). <https://doi.org/10.14763/2026.1.2085>.
- De Vreese, Claes, and Rebekah Tromble. "The Data Abyss: How Lack of Data Access Leaves Research and Society in the Dark." *Political Communication* 40.3 (2023): 356–60. <https://doi.org/10.1080/10584609.2023.2207488>.
- Freelon, Deen, et al. "The Post-API Age of Social Media Data Access: Past, Present, and Future." *The Annals of the American Academy of Political and Social Science* 715.1 (2024): 16–37. <https://doi.org/10.1177/00027162251372557>.
- Hu, Margaret. "Cambridge Analytica's Black Box." *Big Data & Society* 7.2 (2020). <https://doi.org/10.1177/2053951720938091>.
- Jungherr, Andreas, et al. *Retooling Politics: How Digital Media Are Shaping Democracy*. Cambridge: Cambridge UP, 2020.
- Kramer, Adam D.I., Jamie E. Guillory, and Jeffrey T. Hancock. "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks." *Proceedings of the National Academy of Sciences* 111.24 (2014): 8788–90. <https://doi.org/10.1073/pnas.1320040111>.
- Morten, Christopher J., Gabriel Nicholas, and Salomé Viljoen. "Researcher Access to Social Media Data: Lessons from Clinical Trial Data Sharing." *Berkeley Technology Law Journal* 39.1 (2024): 109–204.
- Motyl, Maciej, et al. "Platform Datasets: Challenges, Insights, and Examples for Researchers under Article 40 of the Digital Services Act." *European Digital Media Observatory*, 2025. <https://hdl.handle.net/1814/93101>.
- Perriam, Jessamy, Andreas Birkbak, and Andy Freeman. "Digital Methods in a Post-API Environment." *International Journal of Social Research Methodology* 23.3 (2020): 277–90. <https://doi.org/10.1080/13645579.2019.1682840>.
- Puschmann, Cornelius. "An End to the Wild West of Social Media Research: A Response to Axel Bruns." *Information, Communication & Society* 22.11 (2019): 1582–9. <https://doi.org/10.1080/1369118X.2019.1646300>.
- Puschmann, Cornelius, and Engin Bozdog. "Staking Out the Unclear Ethical Terrain of Online Social Experiments." *Internet Policy Review* 3.4 (2014). <https://doi.org/10.14763/2014.4.338>.
- Venturini, Tommaso, and Richard Rogers. "'API-Based Research' or How Can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach." *Digital Journalism* 7.4 (2019): 532–40. <https://doi.org/10.1080/21670811.2019.1591927>.
- Weller, Katrin. "Accepting the Challenges of Social Media Research." *Online Information Review* 39.3 (2015): 281–9. <https://doi.org/10.1108/OIR-03-2015-0069>.

Author Biography

Fabio Giglietto, University of Urbino. Fabio Giglietto, PhD, is Professor of Internet Studies at the Department of Communication Sciences, Humanities and International Studies at the University of Urbino Carlo Bo. His research explores the intricate relationships between social systems, media, and digital technologies, with a focus on the theory of information, communication, and society.