The Lab's Quarterly
2025/ a. XXVII / n. 0 – ISSN 2035-5548
Online First: 10 June 2025

A NEW WAVE OF SOCIAL MEDIA APIS RESTRICTIONS
A decade of challenges for digital social research

di Suania Acampa[*]

Abstract

Research on digital phenomena is crucial for understanding the complex social processes unfolding online, particularly on social media platforms. Historically, APIs have enabled large-scale data collection, facilitating studies on critical issues such as disinformation, hate speech, and political polarization. However, since 2018, increasing API restrictions have significantly hindered independent research, raising concerns about academia's ability to conduct replicable, large-scale studies. This paper examines the evolution of API restrictions on significant platforms, focusing on Meta (Facebook and Instagram) and Twitter (now X) and the role of European regulations such as the General Data Protection Regulation (GDPR) and the Digital Services Act (DSA), analyzing their impact on digital social research to balance privacy protection and research accessibility.

Keywords

API Restrictions; Digital Social Research; Digital Data; European Regulations

* SUANIA ACAMPA, post-doctoral researcher Department of Social Sciences, University of Naples, Federico II.
Email: suania.acampa@unina.it

DOI: 10.13131/unipi/rmqe-rp75

INTRODUCTION

Currently, research "on" and "with" digital media is crucial for understanding emerging social phenomena that evolve within online spaces. Social media platforms have become environments where complex phenomena unfold, shaping individuals' daily lives and influencing global sociopolitical processes. In studying these phenomena, social media Application Programming Interfaces (APIs) have long been among the key tools for collecting and analyzing data. They have enabled large-scale data collection to analyze qualitative and quantitative socio-political phenomena.

API access has enabled researchers to explore user interactions, map social networks, track information diffusion, and study phenomena like disinformation (Walker et al., 2019; Taeihagh, 2021), hate speech (Bastos e Mercea, 2019), and political polarization (Farkas et al., 2018). In addition to this, specific API-based data collection tools, such as Netvizz (Rieder, 2013) and 4CAT, have allowed researchers without advanced programming skills to conduct research, effectively democratizing digital research.

Since 2018, API access has become increasingly restricted, particularly affecting academia's ability to conduct large-scale, replicable studies (Bruns, 2019; Perriam et al., 2020). This raises crucial questions about the future of research and the possibility of independently monitoring social media without leaving information control solely to the platforms (Ananny e Crawford, 2018).

In parallel with the restrictions imposed by the platforms, European regulation also plays a decisive role in defining access to data for research. The General Data Protection Regulation (GDPR) and the Digital Services Act (DSA) are two key regulations that influence access to social media data in Europe. The GDPR, in force since 2018, limits using personal data for research, introducing constraints on consent, data collection, and anonymization (Art. 6; Art. 9; Art. 17 GDPR). Although these measures are fundamental to guarantee users' privacy, they represent a challenge for digital social research, which must balance privacy protection with the possibility of accessing relevant data to study phenomena of public interest. The DSA, on the other hand, marks an important step towards greater transparency and accountability of digital platforms (Perriam et al., 2020). It imposes on platforms the obligation to provide regulated access to data to qualified researchers. However, the practical implementation of these rules presents difficulties and challenges related to the cooperation of platforms and the actual

application of the provisions.

This paper explores the impact of APIs restrictions on digital social research, emphasizing the need to balance privacy protection with data access. The restriction in access to data could lead to a phenomenon of data-driven determinism (Kitchin, 2014), in which research questions are adapted to the available data rather than addressing the complexity of social phenomena that develop in digital environments through research guided by solid cognitive and theoretical objectives. This situation reduces the ability of researchers to offer a critical and independent perspective on online social dynamics. It also hinders platform transparency and prevents a deeper understanding of the phenomena shaping our digital societies.

The aims of this paper are threefold: (1) to investigate the role played by APIs in digital social research; (2) to retrace the cyclical closures of access to APIs of the main social media platforms and the alternative solutions they have proposed to researchers; (3) to discuss the role of European regulation in managing access to platforms' data and compare it with the policies of the platforms. Through a critical analysis of current European policies and platform policies, this work aims to promote a discussion on the current state of digital social research in light of the new wave of closures imposed by the platforms and the possible ways of collaboration between academia and proprietary companies to promote ethical and transparent use of data; thus, contributing to a broader reflection on the future of digital social research.

1. DIGITAL SOCIAL RESEARCH AND APIS: PAST OPPORTUNITIES AND FUTURE CHALLENGES

For many years, digital social research has relied on APIs to collect semi-structured and unstructured data from digital platforms, especially social media.

These APIs have equipped researchers with tools to explore socio-technical structures and cultural processes arising from users' daily platform interactions (Venturini, Rogers, 2019), opening new perspectives for analyzing complex phenomena. APIs have facilitated access to large amounts of data that were relatively easy to access before the Cambridge Analytica scandal. This ‘little revolution’ as defined by Caliandro (2021: 226), has allowed researchers to collect detailed and large-scale information using intuitive interfaces created thanks to the collaboration between developers and researchers that have quickly

become a fundamental source of information to understand many of the urgent challenges of global society (Venturini e Rogers, 2019). An important example of this collaboration is represented by a project developed by the Digital Methods Initiative of the University of Amsterdam, under the guidance of Bernard Rieder, called 'Capture and Analysis Tools for Social Media Research' (CAT4SMR), which includes a series of tools designed to collect and analyze data from social media intuitively. This project includes two successful applications developed by Rieder: Netvizz, developed in 2010, which allowed researchers to extract data from Facebook, such as posts, comments, friend networks, and group data. The second app is 4CAT, which was developed in 2019 to support collecting and analyzing data from social media in an open-source, flexible, and user-friendly way. These and many other applications have been progressively limited by the changes made by the companies owning the platforms to their APIs, until the definitive suspension after the Cambridge Analytica scandal. The current cemetery of APIs and interrupted tools demonstrates how these tools have represented a precious resource in the toolbox of the digital researcher.

These tools allowed research to be conducted using 'natively' digital data, in line with Rogers's (2013) Digital Methods approach, without the need for advanced programming skills, effectively democratizing access to data for digital social research.

From a methodological point of view, this access to data presented both strengths and weaknesses. On the one hand, according to some scholars, the use of APIs has allowed us to overcome the rigid and traditional distinction between qualitative and quantitative methods (Russell, 2013) highlighted by the fact that digital objects (posts, links, clicks, photos, tweets, etc.), their metrics and their connections, could be interpreted qualitatively but also analyzed quantitatively in the context of more extensive networks. On the other hand, free access to APIs has raised concerns about the risk of researchers' dependence on big social data, thus exposing digital research to a vulnerability that comes from the changes and restrictions imposed by the platforms. Since 2018, companies such as Meta and X (formerly Twitter) have cyclically limited access to these interfaces, weighing heavily on research that depends on these interfaces to collect data continuously, making the future of many long-term research projects uncertain (Perriam et. al., 2020). These restrictions limit academic research in general and directly impact the ability to monitor phenomena such as disinformation or hate speech. Researchers have historically relied on API access to track the spread of misleading content, identify coordinated manipulation campaigns, and

analyze the role of social media in radicalization processes. For example, studies analyzing disinformation during elections (Giglietto et al., 2020) or extremist propaganda networks (Miranda et al., 2022) have examined these phenomena using social media data. As API restrictions tighten, researchers may struggle to conduct similar large-scale real-time analyses. Without systematic access, researchers are forced to rely on smaller datasets, making it harder to detect large-scale trends and draw generalizable conclusions. Moreover, the inability to analyze entire conversations and interaction networks hinders a comprehensive understanding of how hate speech spreads and reduces researchers' ability to assess the effectiveness of platform policies or regulatory interventions to counter it.

This has exposed research to large technology companies' commercial and political logic. Social media data is one of the most important areas of the data market and is constantly growing (Kinder-Kurlanda & Weller, 2020; Chew & Gunasekeran, 2021; Manovich, 2012). After the closure of the APIs, the data has been made available mainly for commercial purposes and is accessible to large marketing and advertising companies (Bruns, 2019). In this regard, a very interesting work by Puschmann and Burgess (2014) analyses the growing control Twitter exercises over users' data through its platform. While initially, Twitter had presented itself as an open platform concerning the closures that Meta was starting, the evolution of its business model led to a subsequent closure of access to data, progressively limiting the amount of data available (for free) through its API and reserving privileged access to commercial partners (Puschmann e Burgess, 2014).

This policy has excluded independent researchers and created a sharp divide between those who have the technical and economic means to access data and those who do not. This process aligns with a broader dynamic of centralizing control over digital data by large technology platforms, which exploit the economic value of data collected through users' daily activities (Marciano et al., 2020; Chen et al., 2022). Freelon (2018: 666) highlights the significance of open API access in digital research, coining the term 'post-API era' to mark the critical turning point caused by platform API closures. This shift has created new challenges for researchers, especially those who do not have the resources to negotiate access or develop alternative data collection methods such as scraping, a data collection method that consists of extracting information from web pages using automated tools or by writing custom scripts in programming languages such as Python or R. The academic world is, therefore, the one that has suffered the heaviest consequences, and

although researchers have been offered alternative data access options, these - as will be explored in the next paragraphs - have not proven adequate to guarantee a supply of data independent of the commercial interests of the platforms themselves (Punziano et. al., 2021).

The end of free API access has triggered several trends in digital social research (Acampa, 2024): the first is migration towards easy data (to collect), therefore websites or platforms that are less restrictive on access to APIs, and the possibility of data scraping. This fuels the (mistaken) idea that any platform is helpful for indiscriminately investigating a specific digital phenomenon. Each platform has its own structural and affordance peculiarities and a different pervasiveness in people's daily lives, which differs by age group and can change from country to country. This triggers the rise of research based on data-driven epistemological approaches in which research questions are adapted to the available data. This situation not only limits the ability of researchers to provide a critical and independent view of online social dynamics but also reduces the transparency of platforms to the detriment of an in-depth understanding of phenomena. This authoritarianism in accessing data could transform the opaque platforms into authentic 'black boxes' (Pasquale, 2015) that escape their responsibilities towards legislators and the public. The second consequence is a return to forms of direct data collection: this forces the researcher to work on small data sets and observe the platforms' dynamics through the same interfaces of the actors he studies (Marres, 2017; Rogers, 2018). In the data collection practice, the risk is that the researcher is influenced by algorithmic recommendations that affect the display of content and profiles (Bucher, 2018; Noble, 2018). The third consequence is the encouragement to deliberately violate the platforms' Terms of Service (TOS) (Freelon, 2018) or to use other unofficial data collection tools, which often clash with the platform's policies. This practice responds to the idea that (in some cases) the benefits for society derived from the violation of the TOS outweigh the damage to the platform (Rogers, 2018; Venturini & Rogers, 2019). A concrete example of these new issues concerns the use of tools such as StorySaver to collect stories from Instagram, which are not accessible via official APIs (Caliandro, 2021) or directly using scraping techniques using HTML code (Bainotti et. al., 2021). The alternative of web scraping on social platforms is problematic, not only from a legal point of view but also from a practical one: the strict controls initiated by the proprietary companies and the implementation of anti-scraping techniques make it difficult (maybe impossible) to complete the collection. Therefore, even if data were collected, their quality and

completeness could be compromised since they are often partial, fragmented, or limited by the technical restrictions imposed by the platforms themselves.

In this 'post-APIcalypse' era, as defined by Bruns (2019: 2), digital social research must face the challenge of finding alternatives that do not bend to the commercial interests of the platforms and that allow researchers to conduct studies on digital phenomena without depending exclusively on direct access to data from platforms or being forced to violate their terms of service (TOS). At the same time, it is necessary to adopt an epistemological approach that is pluralistic, pragmatic, and at the same time critical (Amaturo, Aragona, 2019) that reflects not only on the ethical implications of alternative data collection practices but also looks at the quality of the data collected to ensure reliable results and meaningful interpretations.

1.1 Overcoming the API Challenge: Methodological Approaches for Digital Research

Although not about APIs, some scholars have discussed epistemological and methodological approaches that could become practical and alternative solutions to overcome APIs' limitations in digital research in the future. Each approach exploits the potential of digital environments to collect data, understand real-time behaviors, and directly involve users in the research process.

According to Rogers (2018), with the closure of APIs, it is necessary to rethink digital methods, shifting the focus from APIs to new, more independent ways of collecting data, such as direct observation of users' practices or collaboration with alternative networks. Caliandro (2021: 235)., in this regard, proposes to shift the focus from 'follow the medium (Rogers, 2018) to 'follow the natives'. This approach suggests observing and leveraging users' natural methods to manage their data. Researchers can adopt everyday tools and practices digital users use, such as content-saving apps or tracking software, to explore new digital dynamics.

One methodological solution may be that of 'interface methods' (Marres & Gerlitz, 2016), which focus on analyzing interactions and structures of digital platforms, directly observing how interfaces (web pages, applications) influence users' behaviors. This approach suggests that the interface of a platform is not just a passive tool but an active mediator that organizes and guides users' interactions with data and information. The basic idea of interface methods is that digital platforms

reflect social behaviors and shape them through algorithms, ranking mechanisms, automatic suggestions, and other processes. Therefore, studying how interfaces influence such behaviors can offer crucial information on how digital platforms produce social effects. In this context, we do not access the structured data provided by APIs but analyze how platforms manage and present content. For example, analyzing Facebook news feeds can reveal important information on the mechanisms determining online content's visibility and circulation. Added to these approaches are "trace" (Salganik, 2019: 75) methods, which focus on studying digital traces left by users during their online interactions. Users' actions, such as a like, a comment, or a search on the platform, generate a digital trace that can be analyzed to understand behaviors and social dynamics. Salganik (2019: 47) explores how these traces can be used, for example, for the practice of 'search query': when people use search engines, they leave valuable traces in the form of search queries by analyzing the most frequent search terms, researchers can obtain indications about collective concerns, interests, and behaviors in real-time, offering a vast database from which to draw information. However, a central problem in this approach is related to interpretation: user searches do not always represent concrete behaviors or stable opinions. However, they can be performed in an exploratory or random manner. Similarly, one of the main opportunities the Semantic Web offers is the ability to integrate data from different sources and create interconnections between them, improving the analysis of digital social dynamics (Halford et al., 2013). Using ontologies and metadata, the Semantic Web allows us to give context to data and enrich them with additional information: it could help, for example, to better understand network dynamics or emerging social phenomena since data are organized in ways that allow the discovery of hidden patterns. However, implementing the Semantic Web also brings challenges, such as collaborating with computer science experts to manage technological complexity and correctly interpret the results of advanced analyses.

Digital experiments (Salganik, 2019: 151) are also interesting. They allow manipulating variables in a controlled online environment and observing users' reactions and behaviors in real time. A classic example is the Emotional Contagion Experiment conducted by Facebook in 2012 to study the influence of emotions in the news shown to users (Kramer et al., 2014). This experiment, which involved millions of users, manipulated the news feed to show more positive or negative content to observe how it affected the users' mood.

Finally, crowdsourcing could be a valid strategy for collecting data in

a participatory and decentralized way. It foresees that many people voluntarily contribute their data, such as information about their online activities, consumption habits, or opinions. Platforms that integrate these approaches can invite users to participate by providing data voluntarily, thus making it possible to collect valuable information more democratically and openly.

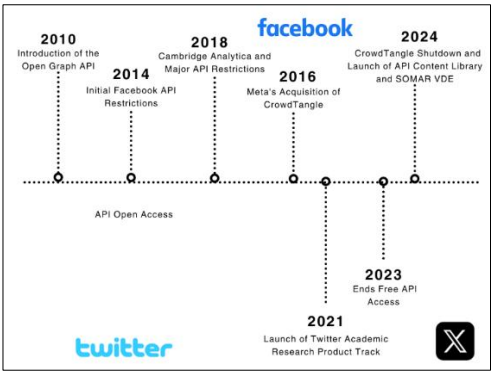
The current context of digital research, therefore, requires a rethinking of methods (Caliandro, 2021). The future of research will necessarily have to integrate innovative approaches that guarantee ethical and transparent access to data while promoting the ability to adapt to the continuous changes in digital platforms and the global technological landscape.

According to Rogers (2018), although a series of proposals have been put forward to change the landscape of AI-based digital research, to date, academic production based on the use of data from proprietary platforms is larger than that which uses alternatives.

2. THE WAVES OF MAIN SOCIAL MEDIA PLATFORMS' API RESTRICTIONS

In recent years, major social media platforms have imposed increasingly strict API restrictions, citing privacy protection and data misuse concerns. This section examines how these restrictions evolved on Meta (Facebook and Instagram) and Twitter (now X), highlighting their impact on digital social research. Figure 1 summarizes the main moments of the restriction's evolution.

Figure 1. Timeline of API restrictions



1.2 The Meta Case

The introduction of Facebook's Open Graph API in 2010 marked a significant turning point in accessing social media data. This API allowed developers and third-party companies to access a wide range of Facebook user data to improve integration and interoperability between Facebook and other web services; thus , it transformed social media into a global social network and allowed developers to connect their applications directly with the platform. Through the API, developers could access different types of user data, including profile information (name, age, gender, location, friends list, etc.); interests and activities (pages "liked", events attended by the user, groups to which the user belongs); interactions (published posts, comments, reactions, shares, and messages; connection data (social graphs showing connections between users, such as mutual friends and shared interactions). This global network has offered social researchers a powerful tool to study various digital phenomena. Facebook began restricting API access in 2014, with severe limitations following the Cambridge Analytica scandal in 2018. This turning point significantly reduced the availability of social media data for research, raising concerns about the ability to monitor online dynamics independently.

Some scholars criticized the platforms for using user privacy as a justification for restricting access to data, arguing that these decisions may also be motivated by a desire to reduce external control and accountability over their operational practices (Freelon, 2018).

In response to these criticisms and the growing debate over platform transparency, Facebook has undertaken several initiatives to facilitate access to data for academic research purposes. It is trying to balance the research needs and user privacy protection. Among these initiatives are the partnership with Social Science One and the integration of CrowdTangle[1].

Social Science One is an initiative launched in 2018 by Facebook in collaboration with high-profile researchers and academics from Harvard University and Stanford University. The project aimed to create a model for collaboration between academia and platform companies that would allow transparent and controlled access to social media data for research. SSO proposed providing researchers

[1] The information on the two initiatives and the data collection process comes from the official websites (now closed) and from the actual use of the services in data collection for many studies conducted by the author.

with secure and controlled access to large, pre-packaged datasets, such as user interactions with political content and campaign ads during elections. This access was subject to strict protocols to prevent abuse and ensure the data was used ethically. An independent review board was responsible for evaluating research proposals, ensuring that studies were conducted according to rigorous ethical standards and that the results were published independently of Facebook's wishes. However, the project encountered numerous difficulties in meeting its promised goals: many researchers complained about delays in accessing the data and limitations in the quantity and quality of information available. Some datasets have never been made available or only partially available (King, Persily, 2019). This has produced a certain mistrust on the part of the academic community regarding the control exercised by the platform on the project, particularly about the selection of the available data and, therefore, the influence on the research results (Bruns, 2019). Furthermore, using pre-packaged datasets, even very large ones, to investigate different digital phenomena raises many doubts linked to data accuracy and reliability problems[2].

Data collection from platforms is one of the most controversial points to address in digital research because this data is not always stable over time. Some elements vanish, and others appear. This refers to the concept of 'interactive complexity' (Rogers, 2018: 560) since some objects collected at a precise moment (for example, likes) can be influenced by new elements that are introduced later (for example, reactions). Therefore, if we wanted to examine likes over time to determine feelings or preferences, the drops or increases in this metric could depend on changes made to the platform's algorithm and the actual change in users' opinions. Furthermore, data can vary due to settings desired by legal and political contingencies or restrictions imposed by national regulations: national authorities could ask platforms to eliminate or limit the circulation of particular content, for example, political content with extremist positions. Therefore, to comply with government requests, the platform does not make particular content available in the country that made the request, but this does not mean that they cannot be seen elsewhere. These examples make one clear: Social Science One's proposal to provide researchers with prepackaged datasets to study digital phenomena

[2] There is no talk of representativeness as it is taken for granted that social media platforms "don't represent all people" (Boyd, Crawford, 2012), but rather a highly specific subset.

raises numerous problems. The critical issues concern not only the quality and stability of the data but also the methodological implications that influence the research that can be conducted on such datasets. As seen, the data collected can be influenced by external factors such as algorithmic changes or government restrictions, which complicates a reliable and coherent reading of the digital phenomena under examination.

The second initiative was CrowdTangle (CT). This tool was born as an independent startup and later acquired by Meta in 2016. CT was designed to monitor and analyze content that spreads on social media, focusing on platforms such as Facebook, Instagram, Reddit, and, to a certain extent, Twitter. CT was created to help publishers, journalists, and researchers investigate how content spreads on the platforms mentioned and the engagement it gets, providing detailed data and metrics on user interaction. Unlike SSO, CT did not provide access to pre-packaged datasets but provided direct access to public and aggregated data, useful for monitoring trends and analyzing large-scale phenomena, but with limitations regarding the depth of the possible analyses. The data researchers could obtain through CT was limited to public content, such as posts published by pages, public groups, and verified profiles, but the text data of comments to these posts was excluded. Regarding comments, the only information available was the number of comments that a given post had obtained. With CT, Facebook's goal was to respond to the growing demand for data access by researchers since CT provided aggregated and public data that allowed (in addition to monitoring engagement) also historical tracking and, therefore, visualizing how certain posts or themes performed over time, facilitating longitudinal data analysis. The tool's novelty was the ability to perform searches by keywords, hashtags, specific pages, and other variables, allowing customized queries for detailed studies. CT has been particularly valuable for all those studies that are based on monitoring and analyzing the spread of disinformation content and tracking its impact on the public (Giglietto, 2020), but also to monitor the spread of hate speech by identifying the communities most active in promoting extremist content (Miranda et. al, 2022); to trace the evolution of social movements, such as Black Lives Matter (Kim, Lee, 2021) or climate change (Alperstein, 2021); to studies on online political polarization by identifying interaction patterns that contributed to the creation of information bubbles (Soares & Recuero, 2021).

The shutdown of CrowdTangle in August 2024 further strained the

relationship between social media platforms and the academic community. A report by the Coalition for Independent Technology Research[3] (2024) found that 88% of surveyed researchers feared that such closures would significantly hinder their work, forcing many to revise or abandon their projects.

Meta has recently launched a new initiative called API Content Library[4], in collaboration with the Inter-University Consortium for Political and Social Research (ICPSR) in collaboration with the Inter-University Consortium for Political and Social Research (ICPSR) which is responsible for ensuring, through an independent review process, that data access is granted to projects that have passed an ethical review process and to accredited researchers who are part of recognized academic institutions. This provides two data access possibilities: the API Content Library and the Social Media Archive Repository (SOMAR) Virtual Data Enclave[5] (VDE).

These solutions address different needs of academic research. The API Content Library has a user-friendly interface. It offers more flexible access to public data. It is useful for macro studies of social dynamics and public discourses but significantly restricts the number of cases that can be extracted. However, access to the Content Library is limited, especially for journalists, and only a tiny percentage of the tens of thousands of CT users have been granted access (Coalition for Independent Technology Research Report, 2024). SOMAR's VDE, on the other hand, allows for more complex and detailed research that also involves the use of non-public data. With access to historical data, the VDE enables longitudinal studies that track changes in user behavior over time, analyzing the evolution of interactions and preferences individually. SOMAR's Virtual Data Enclave (VDE) introduces a novelty in using platform data for research purposes: access is regulated and controlled. Accessing the data exclusively within the virtual enclave, a protected and isolated environment where researchers must conduct their analyses is possible. Every interaction within the VDE is monitored to ensure compliance with security guidelines. Data cannot be downloaded or exported from the VDE, and the review committee must approve any results to be disclosed before being included in a research report; they cannot be shared with

[3]Available at the link: https://independenttechresearch.org/survey-metas-shutdown-of-crowdtangle-bad-for-the-public-interest/

[4]Available at the link: https://transparency.meta.com/it-it/researchtools/meta-content-library/

[5]Available at the link: https://socialmediaarchive.org/pages/?page=FAQs

anyone until approved. Restrictions on data analysis also concern the use of advanced analysis techniques[6] that could, for example, generate results at the individual level, compromising the privacy of users and increasing the risk of personal identification.

The regulated and monitored nature of data access through the VDE entails numerous constraints that can reduce the autonomy of researchers: the fact that each interaction is monitored and the results must pass a review process before being disclosed limits the analytical freedom of researchers and increases the risk of implicit censorship. One of the most worrying aspects concerns the impact on the replicability of studies. This is a crucial principle in scientific research, allowing other scholars to verify, reproduce, and deepen the results obtained. The restrictions imposed by the protected environment of the VDE – such as the prohibition on data export and the limited use of advanced analysis techniques – hinder the access by independent researchers to the data necessary to replicate a study, exacerbating an already existing problem in the field of research on digital environments.

1.3 The Twitter Case

Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt.

Twitter has historically been a valuable source of data for researchers interested in studying online social dynamics, political interactions, and information diffusion (Zimmer, Proferes, 2014). Due to its public nature and ease of access to data via the API, Twitter has been used to monitor events in real-time, analyze user behavior, and study phenomena such as content virality and the propagation of fake news (González-Bailón, Wang, 2016). In 2018, Twitter began to limit the number of API requests that could be made to prevent platform abuse, such as the phenomenon of BOTs and the manipulation of public conversations (Zubiaga, Ji, 2014). Although justified by

[6] It is prohibited: a) any attempt to combine VDE data with other external datasets to identify specific individuals; b) the use of spatial analysis techniques that operate on geolocation data at a very detailed level (e.g. GPS coordinates); c) analyses that use temporal data with a high level of resolution (e.g. exact timestamps of each interaction); d) the use of unsupervised or supervised machine learning algorithms on sensitive data (e.g. predictive analytics to determine personal or behavioral attributes); e) user profiling models based on behavior, interaction and preference data; f) social network analyses with connections at the individual level.

security needs, these restrictions represented the first sign of the platform's distancing from the academic community. A direct response to researchers' criticism of the restrictions was the 'Twitter Academic Research Product Track' project, launched in January 2021 as part of a new initiative to make data easier for academic research. Compared to standard API accounts, the Academic Research Product Track program allowed accredited researchers to send significantly more API requests—up to 10 million monthly tweets, versus the 500,000 limits for standard users. This allowed them to collect and analyze large amounts of data about specific events, social and political trends, or other phenomena of interest. Access was still limited to public tweets, meaning researchers could not analyze direct messages, protected tweets, or other forms of non-public interaction. However, it did have a temporal advantage: researchers could access the entire archive of public Tweets going back to the platform's inception in 2006. This was particularly valuable for longitudinal and historical studies requiring long-term data analysis. To access the program, researchers had to submit a detailed application that included a description of their research project, institutional affiliation, and the purpose of the study. Twitter evaluated these applications to ensure that the projects met ethical standards and that the data would not be used for commercial or non-scientific purposes. The project was open only to researchers affiliated with academic institutions, excluding independent researchers and investigative journalists, who often conduct critical studies on social media. Despite this, the 'Twitter Academic Research Product Track' project represented an important step towards greater transparency and support for academic research; however, with Twitter's change of ownership and corporate policies that began in 2023, the project's future soon became uncertain. In fact, in 2023, Twitter announced the end of free API access, introducing data collection fees.

By introducing high API costs[7] , Twitter has significantly restricted independent research, disproportionately affecting scholars with limited financial resources. This decision has made data access increasingly exclusive, forcing many research projects to be scaled down or abandoned (Freelon, 2018; Tufekci, 2014).

While both companies (Meta and X) still have to find a balance

[7] The basic package costs $100 per month and gives you access to 10,000 posts, while the Pro package costs $5,000 per month and gives you access to up to a million posts. Available at the link: https://developer.x.com/en/docs/x-api/getting-started/about-x-api#item1

between privacy protection and support for academic research, the relationship between the scientific community and the platforms reflects distinct priorities and business models: while Facebook has tried to create institutional partnerships and structures for data access, albeit with many limitations and criticisms about the slowness and selectivity of the data provided; Twitter - which initially offered more open access - has since introduced significant restrictions and high costs, drastically limiting independent research. In the end, while representing a step forward in data access for digital social research, the platforms' proposals could ultimately accentuate the gap between privileged researchers (who manage to gain access to data and overcome review processes) and those excluded from these opportunities. Furthermore, they could push researchers to increasingly focus on studies that respect constraints imposed by the platforms, thus limiting the independence of research and orienting the topics of study towards what is permitted or easy to analyze within the technical and regulatory restrictions imposed by the platforms.

Though platform-imposed restrictions have limited data access for research, European regulations like the Digital Services Act and GDPR have introduced tools to balance transparency and privacy, setting clear guidelines for researchers accessing user-generated data.

3. The European Regulation on Access to Social Media Data: Digital Services Act and GDPR

In recent years, European regulation has taken on an increasingly central role in regulating access to data from digital platforms, including academic research. In particular, the General Data Protection Regulation (GDPR), but most importantly, the most recent Digital Services Act (DSA), introduces rules for researchers who want to access user-generated data on social media. These regulatory tools aim to ensure transparency, security, and respect for privacy while imposing rules on how the data can be used for investigative purposes.

Although there are no specific articles for social media data in the GDPR (which came into force on 25 May 2018), several provisions impact the use of this data for research. Article 6 requires that the processing of personal data must take place with the explicit consent of users, which complicates data collection without direct authorization. Article 9 establishes that sensitive data, such as political opinions, are subject to restrictions. Restrictions on consent

and processing of sensitive data are particularly relevant for using social media data, as many pieces of information, such as political interactions or personal preferences, may fall into protected categories, forcing researchers to be extremely cautious. In the end, Article 17 on the right to erasure allows users to request the removal of their data, influencing the availability of datasets over time. At the same time, Article 15 (right of access) grants individuals the right to access their personal data. In line with the principle of transparency, this provision allows users to obtain a copy of the data held by platforms, allowing them to understand how such data is used. Article 20 (right to data portability) grants individuals the right to receive data in a structured, commonly used, machine-readable format. At the same time, these rights provide an opportunity for researchers to analyze anonymized data while respecting users' privacy. Indeed, thanks to the mechanism of *data donation*—a practice in which individuals voluntarily share their personal data with researchers for scientific purposes—researchers could collect anonymized datasets for research purposes, thus contributing to the advancement of knowledge without compromising the privacy and protection of the individuals involved.

The Digital Services Act (DSA) - approved in 2022 - is one of the most important documents in terms of regulating researchers' access to social media data. The document was created to regulate the activities of online platforms and search engines, promoting a safe and transparent digital environment. It mainly applies to providers of very large online platforms and search engines (VLOP and VLOSE) to create a uniform and clear regulatory framework to improve accountability, transparency, and security in the online environment while protecting users' fundamental rights. In this document, the figure of the researcher, on the one hand, is recognized as a key subject for analyzing online phenomena and promoting greater transparency in the digital ecosystem; on the other hand, his activity is strictly regulated. The regulation introduces an obligation for large online platforms to provide “qualified researchers” with access to relevant information necessary to study online phenomena such as disinformation, systemic risks, and the impact of algorithms. This provision, provided for in Article 40, facilitates independent analysis of platforms, helping to reduce information asymmetries between large digital companies and the public and between platforms and regulators. The DSA allows researchers to monitor the influence of algorithmic systems, offering the possibility of better understanding

how content moderation and recommendation mechanisms work and how these can affect systemic risks online. Furthermore, the DSA allows researchers to access publicly available data on platforms without obstacles. It allows them to monitor user interactions with public content, such as pages and groups, improving the understanding of collective behavior online.

Platforms themselves are encouraged to actively collaborate with researchers through voluntary commitments and codes of conduct to provide broader access to data. This approach increases the transparency of platforms and allows for a more comprehensive analysis of the risks to society arising from online activity in a framework that promotes responsibility and security in the digital world. In parallel, the DSA ensures strict privacy protection, thanks to the integration with the GDPR, establishing that any access to data, including for research purposes, must comply with personal data protection regulations. The data that researchers can access must be anonymized or pseudonymized unless this compromises the purpose of the research, thus ensuring that users' privacy is always protected. The regulation prohibits researchers from accessing information that allows the direct identification of users, protecting the most sensitive personal data. Furthermore, platforms can limit access to data if sharing compromises trade secrets or the security of the service, introducing protection for the legitimate interests of digital companies. The DSA also requires researchers to take rigorous measures to ensure the security and confidentiality of the data they access, requiring them to adopt technologies and procedures that comply with data protection regulations. This includes implementing technical systems such as data vaults, which allow for the secure sharing of information without compromising the confidentiality of personal data or commercial information. In the end, researchers must obtain approval from the competent authorities to access data, which introduces additional control to ensure that only legitimate and proportionate research requests are granted access to sensitive data. In this way, the DSA seeks to balance the promotion of transparency of online platforms with the protection of users' privacy. A key aspect of this balance is the proportionality of data access requests: they must be proportionate to the research purposes and not exceed what is necessary to achieve those objectives. This principle, set out in Recital 97 and Article 40, ensures that access to data is controlled and targeted, reducing risks to users' privacy. However, the DSA also imposes strict control on researchers, who must comply with high

confidentiality and security standards, limiting the use of data to specific research purposes.

Following the definitive closure of CrowdTangle in August 2024, the European Commission asked US tech giant Meta[8] to provide more details on the measures the company intends to take to continue to comply with its obligations under the DSA and allow researchers to access the data. It also requested more information on the new content library and application programming interface, including its eligibility criteria, application process, data that can be accessed, and functionalities. Already in April 2024[9], the Commission had opened proceedings against Meta regarding the company's failure to provide effective tools for real-time third-party monitoring of elections and public debate. This requirement is particularly relevant given the European elections.

The European Commission has also opened a formal investigation into X[10] (formerly Twitter) for possible breaches of the DSA, in particular in areas such as risk management, moderation content, misleading interface design (in particular "blue checks"), and advertising transparency. Particular attention is paid to the effectiveness of X's measures to counter the spread of illegal content and information manipulation, as in the case of recent controversies related to disinformation spread on the platform about global conflicts. As part of this investigation, the Commission also examined the platform's data access policies, which fail to comply with the DSA's requirements regarding transparency and researchers' access to data.

CONCLUSIONS

Access to social media data will continue to be a central theme for digital social research. As highlighted in the previous paragraphs, the so-called 'post-APIcalypse' era (Bruns, 2019: 2) requires rethinking research methods and adopting new approaches to overcome the limitations imposed by platforms while maintaining compliance with

[8]Available at the link:https://digital-strategy.ec.europa.eu/en/news/commission-sends-request-information-meta-under-digital-services-act-2

[9] Available at the link:https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373

[10] Available at the link:https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709

current regulations and privacy protection. The increasing difficulty in accessing data limits researchers' ability to study digital dynamics independently, highlighting the urgency of finding alternative solutions.

The introduction of the Digital Services Act (DSA) and the GDPR represents a step towards a more transparent regulatory framework; however, these regulations fail to fully address the challenges that digital social research faces. While European regulations aim to balance transparency and data protection, platforms continue to impose severe restrictions, controls, and high costs, creating a significant gap between regulatory intent and real-world implementation.

This gap hinders independent research, limiting opportunities for replicable and in-depth analysis of digital phenomena. It is, therefore, essential to think about new strategies that ensure fair and responsible access to platform data, overcoming the barriers created by recent restrictions. Although the DSA and the GDPR recognize the importance of social research in understanding digital phenomena and improving public policies, these regulations still fail to ensure adequate recognition of the researcher's figure by large platforms. Strengthening the implementation of these regulations is necessary to support academic research more effectively.

To address this challenge, legislators must collaborate with platforms; this cooperation could include agreements that facilitate structured and transparent access to data reserved for recognized research institutions in compliance with European regulations. Although Social Science One and CrowdTangle's experiences have shown limitations, they demonstrate that a dialogue is possible if platforms actively collaborate and that public policies create an adequate structure to support such collaboration.

It is therefore urgent to explore, on the one hand, new forms of partnership between academia and the technology industry, promoting models of mutual trust and transparency, as suggested by Venturini and Rogers (2019); on the other hand, research areas should focus on the ethical implications of broader access to data, underlining (if there was still a need) the public value of the knowledge generated by this type of research. It is also crucial to investigate new methodological approaches that ensure data anonymization and address the risks associated with unethical use, as well as to outline strategies that respect users' rights and ethical responsibility in digital research. For example, the GDPR's potential to enable data donation

practices represents an opportunity to bridge the gap between privacy protection and research needs. By empowering users to share their data voluntarily under ethical and legal frameworks, the GDPR facilitates independent, privacy-preserving research.

Data donation has already been successfully implemented in other sectors, such as health research, where individuals voluntarily provide medical data for studies on diseases and treatments. A similar model could be adapted to social media research, allowing users to donate their interaction data (e.g., posts, comments, engagement history) in a controlled and anonymized manner. Platforms could integrate user-friendly consent mechanisms, enabling individuals to contribute to research projects while maintaining privacy safeguards. Implementing these solutions would require collaboration between researchers, regulators, and technology companies to ensure ethical compliance while maximizing the availability of valuable data for social research. In conclusion, although researchers are called upon to answer many of the questions that arise in studying digital phenomena, the current scenario often makes this task impossible. Academia still does not have adequate access to the data needed to provide complete answers, leaving many questions only partially investigated. Phenomena such as misinformation, hate speech, and political polarization, which develop on social platforms, should be a wake-up call for legislators to pressure technology companies to open their data and make it freely and transparently available for research.

In addition, it is increasingly urgent to understand the algorithmic functioning of these environments, a need that clashes with a further element: that of corporate secrecy. What we know today about the algorithmic functioning of social media and its distorting consequences we owe to those rare cases of whistleblowers (Acampa, 2022). This need becomes increasingly difficult to achieve at this rate, despite the impositions of the Digital Services Act and the recognition of the researcher's role in this algorithmic transparency process. The lack of adequate data to evaluate risks and benefits often leads to incomplete results, which raises more questions than answers.

If access to data remains under the total control of Big Tech, the future of independent digital research—and, consequently, the public's ability to understand and regulate social platforms—will be at serious risk.

REFERENCES

ACAMPA, S. (2022). I profitti dell'odio: come l'algoritmo di Facebook alimenta rabbia e contenuti controversi. *Riskelaboration*, 3(1): 47-54.

ACAMPA, S. (2024). *From Dezinformatsiya to Disinformation. A Critical Analysis of Strategies and Effect on the Digital Public Sphere*. Cham: Springer Nature.

ALPERSTEIN, N. (2021). The present and future of performing media activism. In *Performing Media Activism in the Digital Age* (pp. 7-28). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-73804-4_7

AMATURO, E., ARAGONA, B. (2019). Per un'epistemologia del digitale: note sull'uso di big data e computazione nella ricerca sociale. *Quaderni di Sociologia*, 81(81-LXIII): 71-90.

ANANNY, M., CRAWFORD, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3): 973-989.

BAINOTTI, L., CALIANDRO, A., GANDINI, A. (2021). From archive cultures to ephemeral content, and back: Studying Instagram Stories with digital methods. *New Media & Society*, 23(12): 3656-3676. https://doi.org/10.1177/1461444820960071

BASTOS, M. T., MERCEA, D. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review*, 37(1): 38-54.

BOYD, D., CRAWFORD, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5): 662-679.

BRUNS, A. (2019). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 22(11): 1544-1566.

BRUNS, A. (2021). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. In S. Walker, D. Mercea, M. Bastos (eds). *Disinformation and Data Lockdown on Social Platforms* (pp. 14-36). New York: Routledge.

BRUNS, A., HIGHFIELD, T., LIND, R. A. (2012). Blogs, Twitter, and breaking news: The produsage of citizen journalism. In L. Goode & A. M. Nightingale (Eds.), *Citizen Journalism: Global Perspectives* (pp. 92-104). New York: Peter Lang.

BUCHER, T. (2018). *If... Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.

CALIANDRO, A. (2021). Digital methods for ethnography: Analytical

concepts for ethnographers exploring social media environments. *Journal of Digital Social Research*, 3(2): 21-40.

CHEW, A. M., & GUNASEKERAN, D. V. (2021). Social media big data: the good, the bad, and the ugly (un) truths. *Frontiers in big Data*, 4: 623794.

COALITION FOR INDEPENDENT TECHNOLOGY RESEARCH. (2024). Independent technology research and transparency: Challenges and future directions. https://www.coalitiontechresearch.org/report2024

EUROPEAN UNION. (2016). Regulation (EU) 2016/679 of the European Parliament and Council of April 27, 2016 General Data Protection Regulation (GDPR). Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj

EUROPEAN UNION. (2022). Regulation (EU) 2022/2065 of the European Parliament and Council of October 19, 2022, Digital Services Act (DSA) and amending Directive 2000/31/EC. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065

FARKAS, J., SCHOU, J., NEUMAYER, C. (2018). Platformed antagonism: Racist discourses on fake Muslim Facebook pages. *Critical Discourse Studies*, 15(5): 463-480.

FIESLER, C., PROFERES, N. (2018). "Participant" perceptions of Twitter research ethics. *Social Media + Society*, 4(1): 1-14.

FLORIDI, L. (2017). *The Ethics of Information*. Oxford University Press.

FREELON, D. (2018). Computational research in the post-API age. *Political Communication*, 35(4): 665-668.

GIGLIETTO, F., RIGHETTI, N., ROSSI, L., MARINO, G. (2020). It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication & Society*, 23(6): 867-891.

GONZÁLEZ-BAILÓN, S., WANG, N. (2016). Networked discontent: The anatomy of protest campaigns in social media. *Social Networks*, 44: 95-104.

GORWA, R., GUILBEAULT, D. (2020). Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*, 12(3): 281-301.

HALFORD, S., POPE, C., WEAL, M. (2013). Digital futures? Sociological challenges and opportunities in the emergent semantic web. *Sociology*, 47(1): 173-189.

HELBERGER, N., PIERSON, J., POELL, T. (2018). Governing online platforms: From contested to cooperative responsibility. *The*

Information Society, 34(1): 1-14.

KIM, S., LEE, A. (2021). Black lives matter and its counter-movements on Facebook. *Available at SSRN 3980259*.

KINDER-KURLANDA, K. E., & WELLER, K. (2020). Perspective: acknowledging data work in the social media research lifecycle. *Frontiers in big Data*, 3: 509954.

KING, G., PERSILY, N. (2020). A new model for industry–academic partnerships. *PS: Political Science & Politics*, 53(4): 703-709.

KITCHIN, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE Publications.

KRAMER, A. D., GUILLORY, J. E., HANCOCK, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24): 8788-8790.

KREISS, D., MCGREGOR, S. C. (2018). Technology firms shape political communication: The work of Microsoft, Facebook, Twitter, and Google with campaigns during the 2016 US presidential cycle. *Political Communication*, 35(2): 155-177.

MANOVICH, L. (2012). Trending: The promises and the challenges of big social data. In M. K. Gold (Ed.), *Debates in the Digital Humanities* (pp. 460-475). Minneapolis: University of Minnesota Press.

MARCIANO, A., NICITA, A., RAMELLO, G. B. (2020). Big data and big techs: understanding the value of information in platform capitalism. *European Journal of Law and Economics*, 50: 345-358.

MARRES, N. (2017). *Digital Sociology: The Reinvention of Social Research*. Cambridge: Polity Press.

MARRES, N., GERLITZ, C. (2016). Interface methods: Renegotiating relations between digital social research, STS and sociology. *The Sociological Review*, 64(1): 21-46.

MIRANDA, S., MALINI, F., DI FATIMA, B., CRUZ, J. (2022, May). I love to hate! The racist hate speech in social media. In I. Lupa-Wójcik, M. Czyżewska, *Proceedings of the 9th European Conference on Social Media* (pp. 7-20). Krakow: Academic Conferences International (ACI).

NG, L. H., TAEIHAGH, A. (2021). How does fake news spread? Understanding pathways of disinformation spread through APIs. *Policy & Internet*, 13(4): 560-585.

NOBLE, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

PERRIAM, J., BIRKBAK, A., FREEMAN, A. (2020). Digital methods in a

post-API environment. *International Journal of Social Research Methodology*, 23(3): 277-290.

PUNZIANO, G., MARRAZZO, F., ACAMPA, S. (2021). An application of content analysis to Crowdtangle data: The 2020 constitutional referendum campaign on Facebook. *Current Politics & Economics of Europe*, 32(4): 371-397.

PUSCHMANN, C., BURGESS, J. (2014). The politics of Twitter data. In *The Ethics of Studying Social Media* (pp. 43-54). Peter Lang Publishing.

RIEDER, B. (2013). Studying Facebook via data extraction: The Netvizz application. In *Proceedings of the 5th Annual ACM Web Science Conference* (pp. 346-355), Paris. New York: Association for Computing Machinery.

RIEDER, B., SIMON, F. (2016). Dataviz for social media research: Challenges and opportunities. In *The Ethics of Studying Social Media* (pp. 89-109). Springer.

ROGERS, R. (2013). *Digital Methods*. Cambridge: MIT Press.

ROGERS, R. (2018). Social Media Research After The Fake News Debacle. *Partecipazione e Conflitto*, *11*(2): 557-570.

RUSSELL, M. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. Sebastopol: O'Reilly Media.

SALGANIK, M. J. (2019). *Bit by Bit: Social Research in the Digital Age*. Princeton: Princeton University Press.

SOARES, F. B., RECUERO, R. (2021). How the mainstream media help to spread disinformation about COVID-19. *M/C Journal*, 24(1).

TUFEKCI, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Proceedings of the 8th International Conference on Weblogs and Social Media* (pp. 505-514). Michigan: ICWSM.

VENTURINI, T., ROGERS, R. (2019). API-based research or how can digital sociology and journalism studies learn from the Facebook and Cambridge Analytica data breach. *Digital Journalism*, 7(4): 532-540.

WALKER, S., MERCEA, D., BASTOS, M. (2019). The disinformation landscape and the lockdown of social platforms. *Information, Communication & Society*, 22(11): 1531-1543.

ZIMMER, M., PROFERES, N. (2014). A topology of Twitter research: Disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66(3): 250-261.

ZUBIAGA, A., JI, H. (2014). Tweet, but verified: How journalists retweet, filter and annotate tweets. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 241-242). New York: ACM.