

The background of the entire page is a teal-colored image of a computer keyboard. A large metal padlock is attached to the keyboard, with its shackle looping through several keys, including the 'K' and 'O' keys. The padlock is positioned in the lower-left quadrant of the image.

center for
social media, tech
and democracy

 ministry of
digital affairs

Access Denied: When Tech Giants are the Gatekeepers for Researchers' Access to Data

September 2025

**Access Denied:
When Tech Giants are the Gatekeepers
for Researchers' Access to Data**

September 2025

Rounding in tables may cause numbers
not to add up to the total.

Center for Social Media, Tech and Democracy
The Ministry of Digital Affairs
Stormgade 2-6
1470 Copenhagen, Denmark
+45 72 28 24 00
digmin@digmin.dk

ISBN: 97887-85325-16-7

The publication can be downloaded at www.digmin.dk

Design: BGRAPHIC

Photos: iStock

Access Denied: When Tech Giants are the Gatekeepers for Researchers' Access to Data



Contents

Foreword	5
Key findings.....	6
About the study	7
Significant differences in data access over time	8
Danish researchers' experiences with data access	10
X and Facebook are the most frequently researched platforms	11
Facebook and X are also the platforms that researchers most often avoid	12
One in three researchers has redesigned research projects due to data access challenges ...	15
Tech giants' terms of service limit research	17
Where does that leave research regarding the tech giants?	19
References	20



Foreword

More and more of our lives takes place online on the platforms of tech giants. In 2023, more than nine out of ten Danes used social media platforms such as Facebook, TikTok, Instagram, Snapchat, LinkedIn and X^[1]. Democratic discourse is increasingly taking place on digital services^[2], and social media has become a source of information, especially among young people, who often rely on it as their primary source of news^[3].

The more time we spend online, the more data we hand over to the platforms we use. This data determines what content we are presented with online and why some content is profiled more than other. However, we know very little about how the platforms use the data and how the tech giants identify illegal or unwanted content on their platforms.

If we want to understand the online life and behaviour of Danes and thus gain insight into societal development and crucial knowledge about the state of digital democracy in Denmark, we need to understand how the platforms use the data we provide them with. This requires access to the platforms' data on content presentation and profiling.

In this way, access to data is a crucial prerequisite for understanding digital life and its impact on society. Therefore, it is also a democratic issue if we, as a society, do not have sufficient insight into digital services and their impact on society and individuals.

However, the situation seems to be characterised by an asymmetry when it comes to accessing data. While tech giants have almost unlimited access to data about their users and therefore in-depth knowledge about Danes, we as a society generally have limited knowledge about the extent and use of the information and data held by these companies.

New rules in the EU's Digital Services Act^[4] (DSA) address this by giving authorities and researchers easier access to data from tech giants, enabling them to understand the online space, its societal impact, and potential systemic risks.

This report aims to shed light on the experiences of Danish researchers with accessing data from tech giants on the eve of the new European rules taking effect. The report outlines various aspects of the challenges researchers have encountered in their efforts to access data from tech giants. Knowledge that is essential to ensure the full potential of the new regulations is realised.

The report should be viewed in light of the recommendations from the Danish government's expert group on tech giants^[5] and the Nordic Think Tank for Tech and Democracy, which aims to support researchers' access to data from tech giants^[6]. The report is also the first part of Center for Social Media, Tech and Democracy's ongoing focus on practical and efficient researcher access to data.

Caroline Stage
Minister for Digital Affairs



Key findings

- X and Facebook are both the most frequently researched platforms and the most avoided platforms.
 - Almost half of the researchers surveyed have previously used data from X (49 pct.) and Facebook (46 pct.).
 - Half of the researchers surveyed (51 pct.) have avoided specific platforms of the tech giants in their research because they have been unable to access data from the platform or search engine.
 - Among researchers who have avoided specific of the tech giants' platforms in their research, 65 pct. have avoided using data from Facebook, 53 pct. have avoided X, and 36 pct. have avoided Instagram.
- One in three (33 pct.) researchers surveyed have had to redesign an ongoing research project due to challenges with data access from tech giants' platforms and search engines.
- The primary reasons for redesigning research projects are:
 - Data access was changed during the research project.
 - The researchers were unable to access the data they had initially expected.
 - Researchers were unsure if their project (e.g. data collection methodology) violated the platform's terms of service.
- Tech giants' terms of service stand in the way of research. More than two in five (44 pct.) of researchers surveyed say there are methods of data collection and research designs they avoid because they are concerned about not complying with the tech giants' terms of service.



About the study

Center for Social Media, Tech, and Democracy (CSTD) has conducted a survey among 477 researchers from Danish universities to determine whether they have worked with data from tech giants' platforms or search engines and what challenges they have encountered in doing so. 222 of the researchers surveyed have worked with data from these platforms. These researchers form the basis of the inquiry's results.

The questionnaire only asked about researchers' experiences working with data from platforms that were appointed as very large online platforms by the European Commission at the time the questionnaire was sent out, i.e. platforms and search engines with more than 45 million monthly active users in the EU. When the report refers to tech giants, it is referring to these very large online platforms and search engines. The survey results do not shed light on data access to smaller online platforms and search engines. The report refers to the platforms by their current names.

However, since the questionnaire focuses on previous experiences, some of the platforms may have changed names since the researchers used data from them. For example, the report refers to "X", even though the results can refer to experiences with the platform both when it was called Twitter and after it changed ownership and name to X. Prior to developing the questionnaire, CSTD conducted a qualitative interview study with five selected researchers. The interviews were used as background knowledge and to qualify the development of the questionnaire. The question framework has been pilot tested twice before sending out the questionnaire. The collection period ran from 26 August 2024 to 16 September 2024.

Following the survey, CSTD invited three researchers to provide quotes for the study.





Significant differences in data access over time

Research based on data from tech giants is already happening on a large scale today. Both in Denmark and abroad.

However, the conditions that determine whether and how researchers can access data from tech giants' platforms, as well as what data researchers can access, have varied between platforms before the European Digital Services Act came into force and have also changed over time^[7].

In practice, this has meant that data access has varied over time.

Changes in data access can present new opportunities for data access, but can also mean new restrictions that make previously accessible data unavailable to researchers.

Box 1: Digital Services Act (DSA), Art. 40

Article 40 of the DSA aims to provide researchers with access to data from very large online platforms and search engines (with more than 45 million monthly active users in the EU) for research that sheds light on systemic risks in the EU.

There are two different ways of accessing data:

1. Access for vetted researchers to non-publicly accessible data, see Art. 40(4).
2. Access for researchers, including those working for non-profit organisations, associations, and similar organisations (it is not a requirement here that the researcher be affiliated with a research institution), to already publicly accessible data, see Art. 40(12).

Source: European Centre for Algorithmic Transparency
FAQs: DSA data access for researchers - European Commission

Box 2: About changing data access for researchers



Researchers' access to social media data has been in decline for several years. There are a few and often temporary exceptions. For example, access to certain types of data was improved shortly after the revelation of Russia's interference in the 2016 US presidential election.

Twitter also further improved access shortly before Elon Musk took over the platform by giving researchers free access to historical data on past tweets. However, even these measures have been rolled back under Musk's leadership.

– Yevgeniy Golovchenko, Assistant Professor at the Department of Political Science and SODAS, University of Copenhagen



Such changes in data access can create uncertainty for researchers using data from tech giants' platforms and search engines, as they cannot be sure that the data they use to examine specific topics will always be available.

For example, the so-called "Cambridge Analytica scandal"^[8] in 2018 is reported to have led to a severe restriction of access to data from Facebook, among others, which immediately after the scandal introduced significant restrictions on access to data from Facebook groups and events, which researchers previously used to study phenomena, such as political mobilisation^[9].

So far, there have also been considerable differences in the practices and conditions for data access that different digital services have had in

terms of providing access to data for research purposes. This applies, for example, to both the scope of access and the process of access^[10].

The varying access to data and changes in practice can impact the way data is collected and subsequently utilised in research. For example, it is stated that some platforms have been "over-studied" because it has actually been possible to access data from these platforms. This has been the case with X, for instance, when the platform was called Twitter, which was once one of the platforms with the easiest access to data and thus one of the most frequently researched platforms^[11]. Other platforms, on the other hand, have been avoided by researchers because access to their data has been limited.

Box 3: About restricted researchers' access to data



In addition to significantly restricting researchers' access to data, there are also examples of platforms increasingly making it harder for researchers to use the data they do gain access to, citing concerns about user privacy. The reality is that the quality of the data provided has deteriorated, making it more challenging to utilise the data in research.

– Anja Bechmann, Professor and Head of DATALAB, Aarhus University



Danish researchers' experiences with data access

Center for Social Media, Tech and Democracy has asked 222 researchers who have worked with data from tech giants about their experiences with data access. On the following pages, you can read more about their experiences.

Box 4: Why researchers' access is important



Access to data from the tech giants is crucial for researchers to understand the organisation, structure and development of society.

For example, data about what content and actors the user encounter on the platforms and what content the platforms restrict users' access to. Information about this can provide insights into the systematic highlighting of some actors over others - actors who thereby gain an invisible power in our society.

It is therefore a major societal problem, that it has become increasingly difficult for researchers to gain access to data from tech giants. With the amount of work and time we spend on their platforms, it can even be said to have a detrimental impact on our society that there is such an asymmetry in that the tech giants have almost unlimited knowledge about us, while we as a society have very little knowledge about what information the tech giants have about us and how they use the information they have about us.

– Anja Bechmann, Professor and Head of DATALAB, Aarhus University

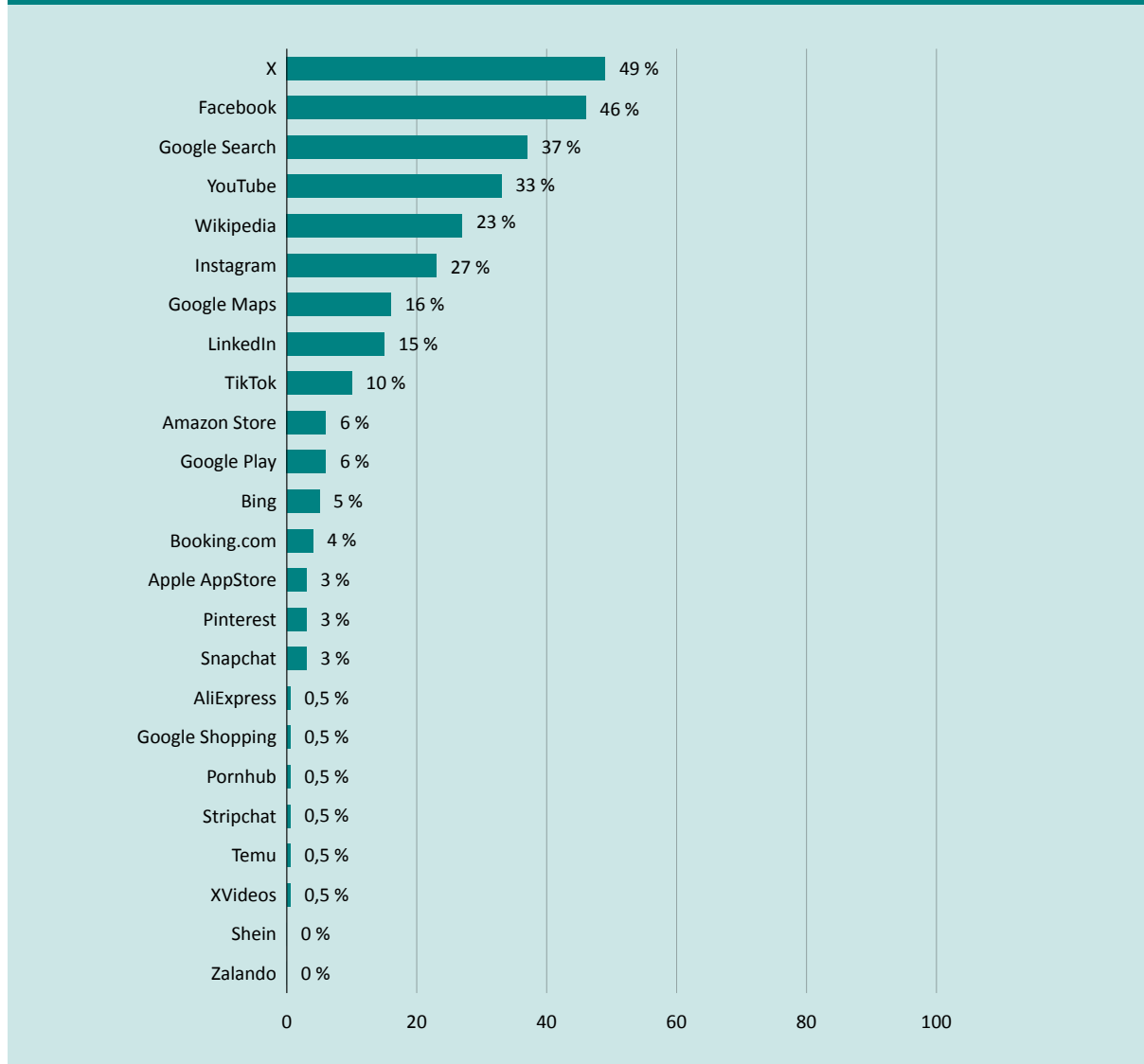


X and Facebook are the most frequently researched platforms

X and Facebook are the platforms that most researchers report having used data from, as shown in Figure 1. Almost half of the researchers have used data from X (49 pct.) and Facebook (46 pct.) respectively. This result indicates that Facebook and X have been among the most popular platforms for using data in Danish research environments. This may be because X and Facebook have had relatively open data access

in the past^[12]. However, the result cannot conclude that Facebook and X have had the best data access, as it does not, for example, shed light on how many people have failed to access data from these platforms. X and Facebook are also among the longest-standing online platforms, while other platforms, such as TikTok and Temu, have been around for a shorter time which this result should be seen in light of.

Figure 1: Have you worked with data from any of the following very large platforms or search engines in previous or current research projects? (Select all the platforms and search engines you have worked with data from)



Note: n = 222
Respondents were able to select multiple answers.

Source: Center for Social Media, Tech and Democracy, 2024



Facebook and X are also the platforms that researchers most often avoid

Data access and data reliability are important prerequisites for research. An expectation of inadequate or non-existent data access can therefore influence which platforms are researched.

Of the researchers surveyed, just over half (51 pct.) have avoided specific of the tech giants' platforms or search engines in their research because they have been unable to access data from the platform or search engine, as shown in Figure 2.

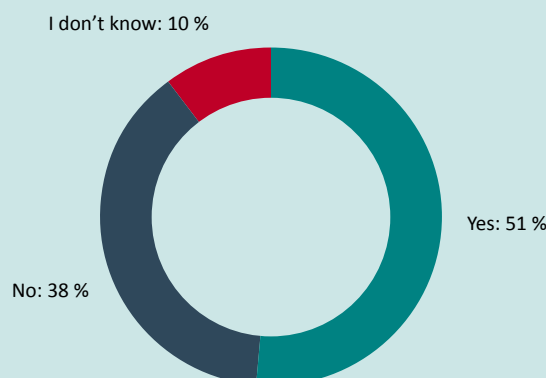
Box 5: On the consequences of platforms' different data access practices



It is evident that the different practices among the platforms have consequences when you look at the data on which research in this area is based. There used to be a lot of research based on data from Twitter (X), but there is significantly less research based on data from Instagram, for example. As a researcher, you have to be pragmatic and look for data wherever you can get it, so you have to organise your research accordingly. There are therefore platforms that researchers avoid using in their research."

– Daniel Lundgaard, Postdoc, Department of Management, Society and Communication, Copenhagen Business School

Figure 2: Have you avoided specific, very large online platforms or search engines in your research because you couldn't access data from the platform/search engine?



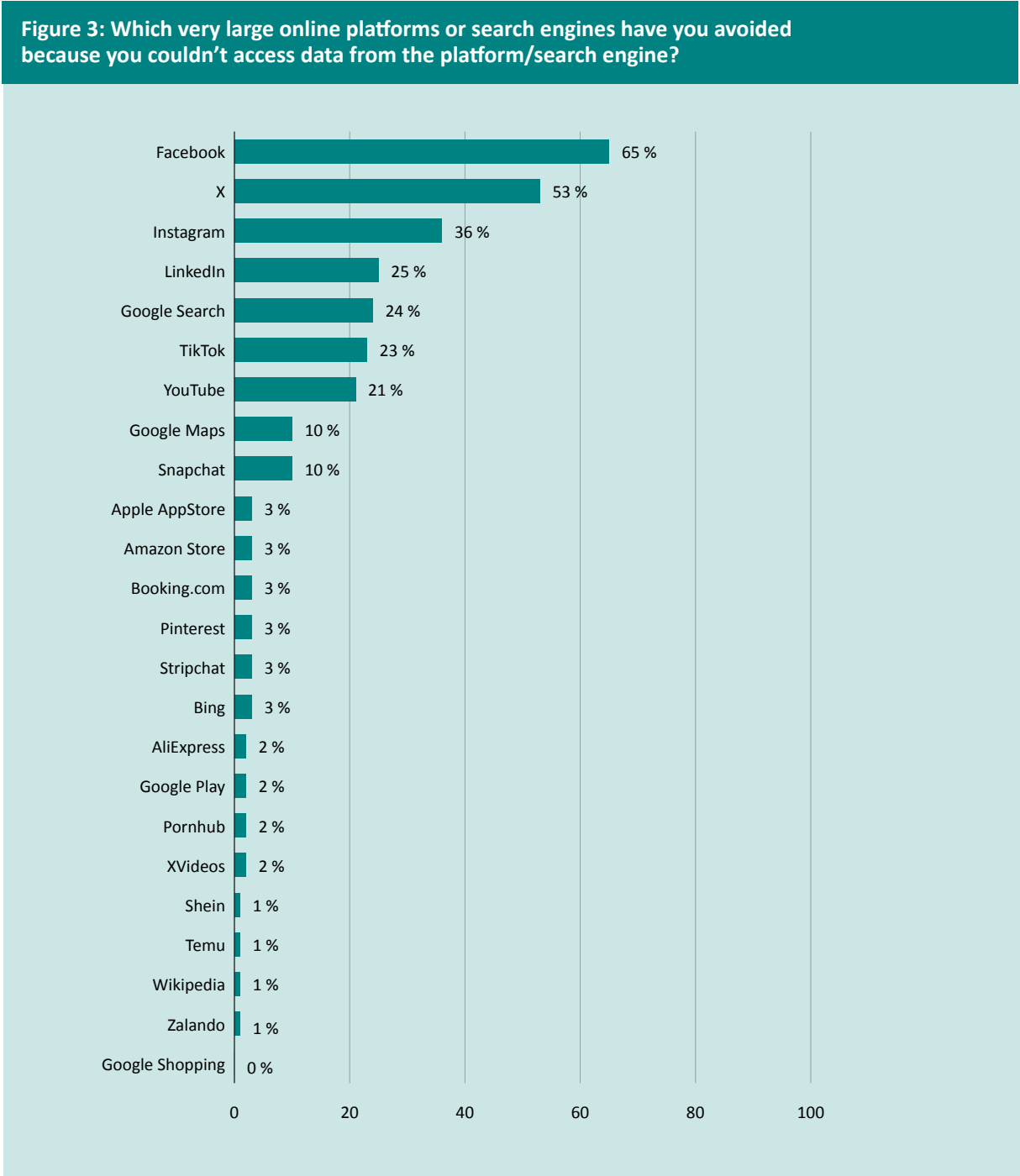
Note: n = 222

Source: Center for Social Media, Tech and Democracy, 2024



Among the researchers who have avoided specific of the tech giants' platforms and search engines, 65 pct. say they have avoided using data from Facebook, 53 pct. have avoided X, and 36 pct. have avoided Instagram, as shown in Figure 3.

This means that some of the most used platforms by the Danish population (Facebook and Instagram^[13]) are also among the platforms that researchers have previously avoided in their research due to a lack of data access.



Note: n = 115
Respondents were able to select multiple answers.

Source: Center for Social Media, Tech and Democracy, 2024



Box 6: About the consequences of X's restrictive data access



Twitter, now X, has become one of the most closed platforms when it comes to data access. It is still possible to access datasets, but the cost has increased significantly, making it unaffordable for many researchers, which is particularly concerning given that a substantial portion of global research has relied on Twitter data. In some cases, this has not only affected research but also the teaching of courses where students had to learn how to work with social media data. I have been in a situation myself where the transition to a payment system has created a lot of turbulence and uncertainty about exactly what data students can work with and when their access ends.

– Yevgeniy Golovchenko, Assistant Professor, Department of Political Science and SODAS, University of Copenhagen

As previously described, Facebook and X are also among the online platforms that most of the researchers surveyed say they have used data from. It may seem contradictory that these platforms are both among the most frequently used by researchers and, at the same time, among those that most of the researchers surveyed have avoided in their research. However, these findings may reflect that data access to the platforms has changed over time and has been more or less open at different times.



One in three researchers has redesigned research projects due to data access challenges

One in three (33 pct.) of the researchers surveyed report having had to redesign a research project that was underway due to various challenges with data access from platforms and search engines, as shown in Figure 4.

Among researchers who have had to redesign a research project due to challenges with data access, half (51 pct.) state that they changed their research design because the data access changed during their project, as shown in Figure 5. Almost half (46 pct.) have also found that they were unable to access the data they initially expected.

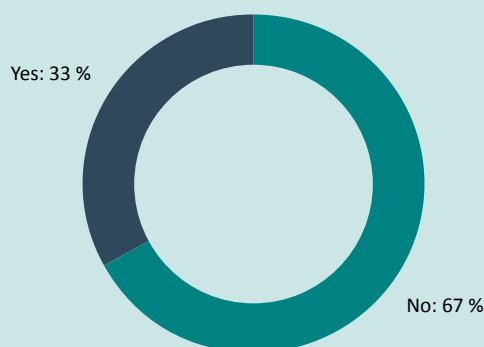
Box 7: On the consequences of data access challenges



The consequence of the fact that it has become significantly more difficult to access data from the majority of platforms is limitations in understanding the role platforms play in society.

– Daniel Lundgaard, Postdoc, Department of Management, Society and Communication, Copenhagen Business School

Figure 4: Have you ever had to redesign a research project that was already underway due to data access challenges?



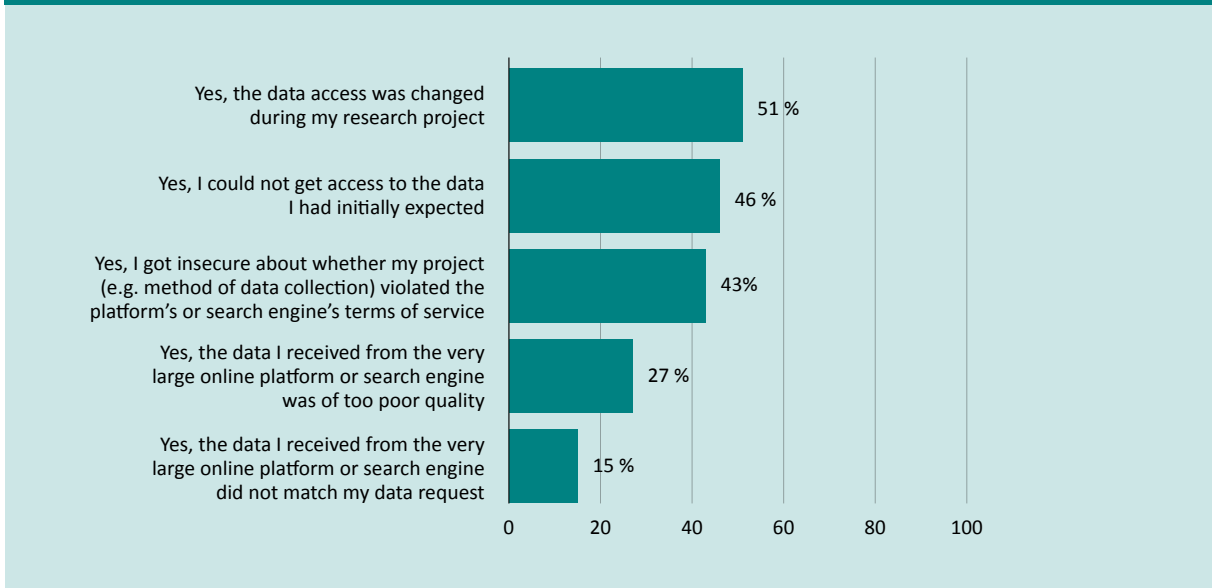
Note: n = 222

The original wording of the question was: "Have you ever had to redesign a research project due to any of the following reasons regarding data access from very large online platforms or search engines? (Select all relevant reasons)". Respondents have been able to answer "No" or select the relevant reasons related to data access for redesigning a research project. "Yes" depicts all respondents who did not answer "No".

Source: Center for Social Media, Tech and Democracy, 2024



Figure 5: Have you ever had to redesign a research project that was already underway due to any of the following reasons related to accessing data from very large online platforms and search engines? (Select all relevant reasons).



Note: n = 74
Percentages do not add up to 100, as respondents were able to select multiple response categories. Respondents could also select the response category “No, I have not had to redesign a research project due to reasons regarding data access from very large online platforms and search engines.” Respondents who selected this response category are excluded from this figure. They are depicted in Figure 4.

Source: Center for Social Media, Tech and Democracy, 2024

Just over one in four (27 pct.) say they have redesigned a research project because the data they received from the platform or search engine was of too poor quality. 15 pct. state that the data they received from the platform or search engine did not match their data request.



Tech giants' terms of service limit research

To grant access to data, tech giants generally require researchers to comply with the digital service's terms of service. Such terms differ for each service but include, for example, specific terms regarding how data may be used, such as the duration of data storage, the purposes for which data may be used, etc. Often, reference is also made to compliance with other regulations in this area, such as GDPR.

As depicted in Figure 5 above, 43 pct. of researchers who had to redesign a research project stated it was due to their concern that their project violated the platform's terms of service, specifically regarding the method of data collection.

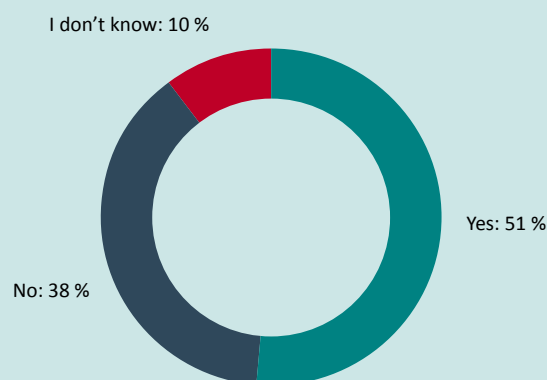
More than two out of five (44 pct.) of the researchers surveyed also indicated that there are methods of data collection and research design that they avoid completely because they are concerned about not complying with the platforms and search engines' terms of service, as shown in Figure 6 below.

In addition to researchers having to redesign research projects that are already underway due to concerns about non-compliance with the platform's terms of service, the same concerns can also prevent researchers from incorporating specific data collection methods into their research projects altogether.

Concerns have even been raised among researchers that the invocation of terms of service by digital services may have been done to limit research on data, for example, by imposing various requirements, including technical ones, that may be difficult to fulfil^[14].

There is a general demand among researchers for greater clarity in the content and scope of such terms of service to minimise the risk of individual researchers acting contrary to them^[15].

Figure 6: Have you avoided data collection methods or research designs because you were worried about not complying with a very large online platform or search engine's terms of service?



Note: n = 222

Source: Center for Social Media, Tech and Democracy, 2024



Box 8: About tech giants setting conditions for data access



Tech giants can restrict both access to and use of data, for example by imposing conditions on how data can be used. In practice, these conditions can be difficult for researchers to fulfil. Allowing tech giants to control research in this manner can significantly compromise the freedom of research. The freedom of research also suffers if tech giants have a say in which researchers and research teams can access data.

Setting up different conditions for researchers' access to data can ultimately mean that research into data from tech giants risks being done at the tech giants' terms and mercy.

– Anja Bechmann, Professor and Head of DATALAB, Aarhus University



Where does that leave research regarding the tech giants?

Shifting data access and unclear or unfair terms of service can influence which platforms researchers choose to investigate and how they conduct their research. Researchers are to some extent dependent on what data the tech giants provide access to and how this data can be stored, used, and so on.

This has implications for the knowledge society can gain about the conditions on tech giants' platforms and search engines, as well as the impact of tech giants on society. It also has implications for researchers' ability to reproduce existing studies and thus qualify existing knowledge. This is because you cannot expect to get access to the same data that has been used in previous studies, as several tech giants don't allow researchers to share the data they have accessed with each other^[16].

Center for Social Media, Tech and Democracy considers a lack of or poor data access as a fundamental democratic problem. The consequences of poor data access are that areas of great importance to our society, such as the protection of children and young people online, or the spread of misinformation and disinformation on online platforms, end up not being adequately

investigated. Ensuring robust access to data from tech giants' platforms and search engines is therefore crucial for society. It will play a key role in creating a knowledge base to address societal problems, such as the spread of misinformation and disinformation, and the exposure of harmful content to children and young people, as well as enforcing the European Digital Services Act (DSA).

Article 40 of the DSA aims to ensure that researchers have access to data from tech giants' platforms and search engines for research in systemic risk in the EU. It is essential to assess whether the rules work as intended and actually ensure that researchers get access to the necessary data.

This report is the first step in an ongoing initiative by Center for Social Media, Tech, and Democracy to investigate the impact of Article 40 of the DSA on Danish researchers' access to data. The next step will follow Danish researchers' requests for data access in the early days of the DSA to assess whether the goal of ensuring robust researcher access to data is being realised in practice.



References

- [1] ICT usage in households and by individuals, Statistics Denmark, 2023, p. 25 (in Danish)
- [2] ICT usage in households and by individuals, Statistics Denmark, 2023, p. 25 (in Danish)
- [3] Cultural Habits Survey, Statistics Denmark Q3 2023
- [4] Digital Services Regulation, European Commission. EU's forordning om digitale tjenester. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en
- [5] Democratic Control Over Big Tech Business Models. Interim report from the Danish government's expert group on big tech (2023) p. 21. <https://www.em.dk/aktuelt/udgivelser-og-aftaler/2023/jun/demokratisk-kontrol-med-tech-giganternes-forretningsmodeller>
- [6] A Nordic Approach to democratic debate in the age of Big Tech - Recommendations from the Nordic Think Tank for Tech and Democracy (2023). <https://pub.norden.org/nord2023-004/>
- [7] Davidson, B.; Wischerath, D.; Racek, D.; Parry, D. A.; Godwin, E.; Hinds, J.; van der Linden, D.; Roscoe, J. F.; Ayravainen, L. & Cork, A. 2023, "Platform-controlled social media APIs threaten Open Science", Nature Human Behaviour. <https://doi.org/10.1038/s41562-023-01750-2>
- [8] After the 'APIcalypse'.pdf (snurb.info), page 1. Consultancy firm Cambridge Analytica had used more than 50 million Facebook profiles to target adverts during the 2016 US presidential election, among other things. The company was accused of not taking good enough care of users' data. The digital services' response to the case was to exclude users and tighten the terms of use, including restrictions on researchers' access to the services' data, see, e.g. Jessamy Perriam et al. "Digital methods in a post-API environment", page 28
- [9] Perriam, J.; Birkbak, A.; & Freeman, A. (2019). Digital methods in a post-API environment. International Journal of Social Research Methodology, 23(3), 277-290.
- [10] Davidson, B.; Wischerath, D.; Racek, D.; Parry, D. A.; Godwin, E.; Hinds, J.; van der Linden, D.; Roscoe, J. F.; Ayravainen, L. & Cork, A. 2023, "Platform-controlled social media APIs threaten Open Science", Nature Human Behaviour. <https://doi.org/10.1038/s41562-023-01750-2>
- [11] Trezza D. (2023) To scrape or not to scrape, this is the dilemma. The post-API scenario and implications on digital research. Front Sociol. 8:1145038. doi: 10.3389/fsoc.2023.1145038
- [12] Perriam, J.; Birkbak, A.; & Freeman, A. (2019). Digital methods in a post-API environment. International Journal of Social Research Methodology, 23(3), 277-290. <https://doi.org/10.1080/13645579.2019.1682840>
- [13] ICT usage in households and by individuals 2023, Statistics Denmark
- [14] Response to the European Commission's call for evidence for a Delegated Regulation on data access provided for in the Digital Services Act. "The DSA must ensure public data for public interest research", Algorithmic Watch, 2023
- [15] Access Granted: Why the European Commission Should Issue Guidance on Access to Publicly Available Data Now | Democracy Reporting International. <https://democracy-reporting.org/en/office/global/publications/access-granted-why-the-european-commission-should-issue-guidance-on-access-to-publicly-available-data-now>
- [16] Davidson, B. I.; Wischerath, D.; Racek, D.; Parry, D. A.; Godwin, E.; Hinds, J.; van der Linden, D.; Roscoe, J. F.; Ayravainen, L.; Cork, A. G. (Forthcoming/2023). Platform-controlled social media APIs threaten Open Science. Nature Human Behaviour

About Center for Social Media, Tech and Democracy

Center for Social Media, Tech and Democracy was established as part of the Media Agreement 2023-2026. The center is located in the Ministry of Digital Affairs.

The center's tasks include contributing knowledge about the mental well-being of users, the impact and consequences of tech giants on society, and the impact of the spread of misinformation and disinformation on digital platforms on democratic discourse.

