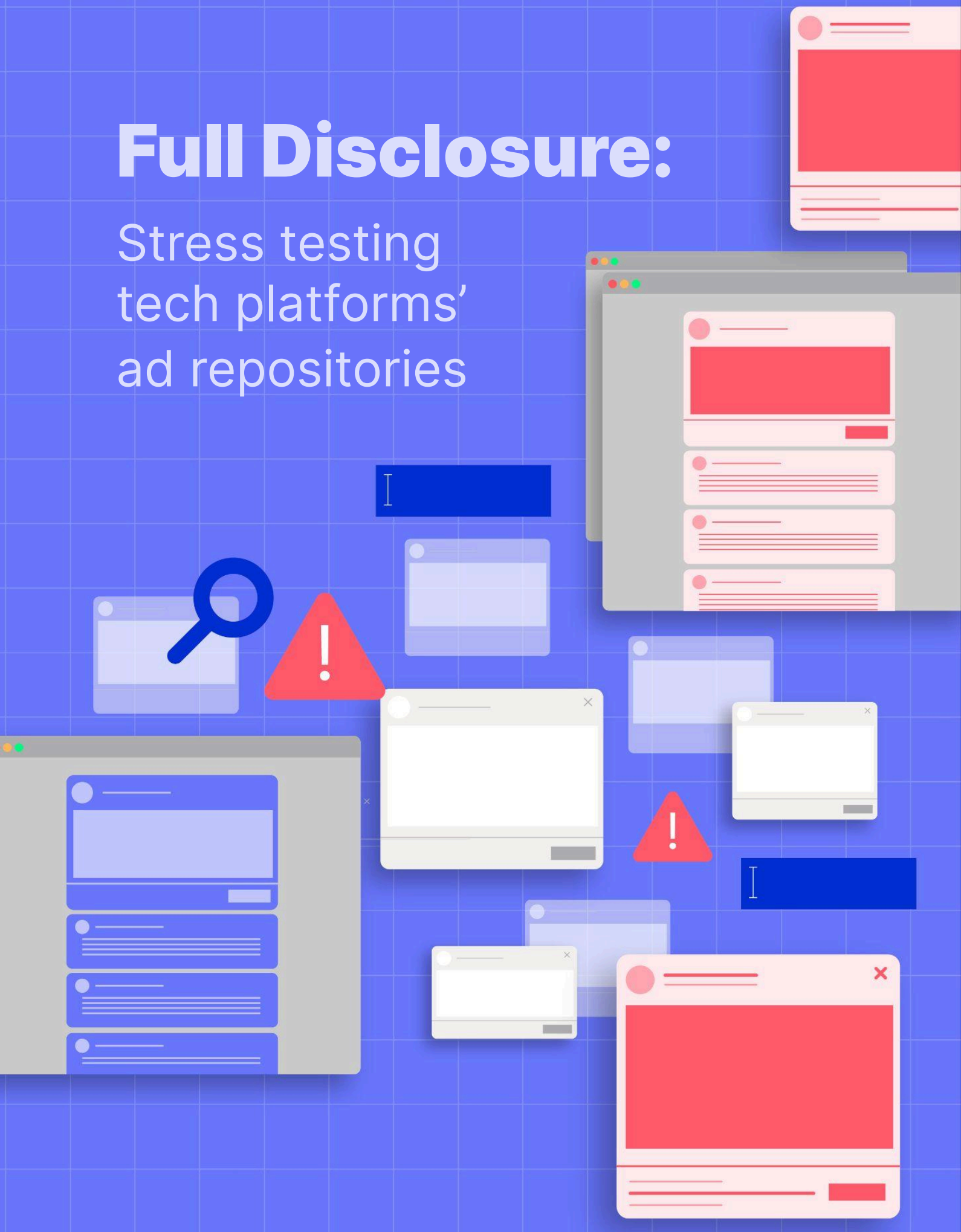
# Full Disclosure:

## Stress testing tech platforms' ad repositories



moz://a

CHECK FIRST

**Executive Summary** **2**
**1. Introduction** **5**
**2. Scope** **7**
a. Platforms Under Review 7
b. Criteria for Assessment 7
**3. Methodology** **11**
a. Limitations of this research 11
b. Data Collection 12
c. Stress Testing 13
d. Usability Testing 14
**4. Some improvements since Mozilla's 2019 Analysis** **16**
**5. Analysis & Testing** **18**
AliExpress 19
Alphabet: Google Search & YouTube 22
Apple App Store 25
Bing 28
Booking.com 30
LinkedIn 32
Meta 34
Pinterest 36
Snapchat 39
TikTok 41
X 44
Zalando 46
**6. Key Takeaways From Comparative Analysis** **49**
**7. Conclusion** **51**
**8. Recommendations** **53**
a. Online Platforms 53
b. Regulators 54
**9. Proposed Standardized Ads Repository Format:** **56**
**10. Annex** **58**
a. Ad Repositories links 58

# Executive Summary

**The ad libraries provided by large tech platforms are critical tools for the public to assess the role of commercial advertising and paid influence on services which are used by billions every day.[1] These libraries collect ad content and information about the intention of ad campaigns, giving researchers and the public a view into how commercial communications are used to influence the information space.**

**The EU Digital Services Act (DSA) requires that the largest online platforms and search engines (those designated by the EU Commission as "Very Large Online Platforms and Search Engines - hereafter VLOs) have ad repositories. This harmonized requirement, set out in Article 39 of the DSA, comes not a moment too soon, with this June's 2024 European elections just around the corner. But the effectiveness of these tools depends on their usefulness for researchers in practice.**

In this independent "stress test", we assessed whether the available ad repositories are truly ready for action. We examined the following services designated by the European Commission in April 2023: **AliExpress, Apple App Store, Bing, Booking.com, Alphabet (Google Search and YouTube), LinkedIn, Meta (Facebook and Instagram), Pinterest, Snapchat, TikTok, X,** and **Zalando.**

Our evaluation is based on a combination of the DSA's requirements for VLOs in Article 39(1-2) and five guidelines written by Mozilla and other experts in 2019. We examine factors such as the depth of information provided regarding the advertisement and its advertiser, the targeting criteria employed, and the ad's reach. Additionally, we evaluate the completeness of the ad repository, the availability of historical data, and the accessibility, consistency, and documentation of the tools provided. Most platforms have a separate web-based ad repository and an API, and we assess these individually.

Our testing was conducted between December 2023 and January 2024. We note, but have not been able to systematically test, developments following the launch of our tests, like the appearance of new APIs. We follow each assessment with individual recommendations, and conclude the report with general observations and suggestions for improvement and harmonization.

---

[1] Vast Networks of Fake Accounts Raise Questions About Meta's Compliance with the EU's New Digital Rulebook ( Reset, Oct 2023), https://www.reset.tech/documents/24102023_Networks-of-Facebook-Fake-Advertisers_Reset.pdf. Facebook Hustles: The Hidden Mechanics of a Scam Machinery Impersonating News Organisations and Creators (CheckFirst, Jun 2023), https://checkfirst.network/wp-content/uploads/2023/06/Facebook-Hustles-The-Hidden-Mechanics-of-a-Scam-Machinery-Impersonating-News-Organisations-and-Creators.pdf.

We find a huge variation among the platforms, but one thing is true across all of them: none is a fully-functional ad repository and none will provide researchers and civil society groups with the tools and data they need to effectively monitor the impact of VLOs advertisements on Europe's upcoming elections. For example, **AliExpress**'s repository provides the bare minimum in terms of data transparency and user interface, and **X (formerly Twitter)** only provides a CSV file. While these platforms provide the worst examples[2], we struggle to tell you which one is best. While we don't test the branded content (influencer) repositories for methodological reasons, we find that only a handful of platforms have this kind of content in their repositories at all, despite many of them having influencer presence. Critically, these repositories are far from compatible, making it difficult for researchers to study paid influence across the platforms systematically. Finally, these transparency tools were often difficult to even locate in the first place. A few are accessible through several clicks from an ad in the interface, but in many cases we had to dig into the platforms' terms and conditions.

## Key findings

- **AliExpress's** repository has no API and a very minimal user interface, and even that requires a user account to access. We encountered loading and display errors and were blocked several times by anti-bot tools.
- **Apple App Store:** The web repository and API lack important details for understanding paid influence, like targeting broken down by country
- **Bing:** We were disappointed by the limited data in the API and limited API documentation.
- **Booking.com:** We found it very difficult to link ads on the platform to ads in the web repository, and we found limited documentation about the API.
- **Alphabet** (Google Search and YouTube): We see progress, but it's been six years and we still can't search by keyword.
- **LinkedIn:** The web repository only allows for limited analysis
- **Pinterest:** We found accuracy issues and missing data.
- **Snapchat:** The search functionality is very limited, and there is currently no API.
- **TikTok:** The ad repository and API look robust, but we encountered gaps and accuracy errors.
- **X:** With a (slow to load) CSV file instead of a web interface, this was a major disappointment.
- **Zalando:** The web repository doesn't allow some basic capabilities like combined searches, and we also found only limited information about targeting.

[2] During our research Amazon did not even have a public ad repository. It was granted a temporary exemption by the ECJ, but on March 27, Amazon's request to suspend its obligation was rejected. It now has an Ad Library API accessible at http://advertising-api-eu.amazon.com
Order of the Vice-President of the Court in Case C-639/23 P(R) | Commission v Amazon Services Europe (Curia.europa,, March 2024), https://curia.europa.eu/jcms/upload/docs/application/pdf/2024-03/cp240060en.pdf

| Stress Testing Tech Platforms' Ad Repositories | |
|---|---|
| **Platform** | **Evaluation** |
| Aliexpress | Lacks vital data and functionality |
| Bing | Lacks vital data and functionality |
| SnapChat | Lacks vital data and functionality |
| X | Lacks vital data and functionality |
| Zalando | Lacks vital data and functionality |
| Alphabet | Bare minimum data and functionality |
| Booking.com | Bare minimum data and functionality |
| Pinterest | Bare minimum data and functionality |
| Apple App Store | Still has big gaps in data and functionality |
| LinkedIn | Still has big gaps in data and functionality |
| Meta | Still has big gaps in data and functionality |
| TikTok | Still has big gaps in data and functionality |
| No Platforms | Ready for action |

## For Platforms:

1. Platforms should remove access barriers, ensuring the repositories are easily accessible and widely available
2. Repositories should provide more complete data on ad campaigns and more granular information about ad intentions and effectiveness
3. Search functionalities should be improved, with additional options to export data.
4. Platforms should offer better documentation and user support to effectively empower research
5. The web-based repositories and APIs should be more harmonized to facilitate cross-platform research

### For Regulators:

1. The European Commission and the Board for Digital Services Coordinators should develop guidelines on ad repositories, in consultation with the research community, in particular civil society researchers and platform integrity experts.
2. The European Commission should encourage the standardization of APIs across the designated VLOs to increase usability and facilitate cross-platform research
3. Regulators should strengthen disclosure requirements and assurances for branded or 'influencer' content.

# 1. Introduction

**Ad libraries let researchers, watchdogs, and members of the public understand how commercial communications influence the information space and affect society. These tools are particularly important during democratic moments like elections.**

The EU's Digital Services Act (DSA), which entered into force in August 2023, introduces legal obligations on the largest online platforms and search engines ("VLOs"). Article 39 requires VLOs to compile and maintain a public and searchable ad library ("repository") including certain information about the ad content, reach, and targeting criteria.

Mozilla has long advocated for advertising transparency. In 2019, ahead of the last European elections, Mozilla and a cohort of independent researchers identified key traits for an effective ad archive API.[3] These five guidelines - which address comprehensiveness, detail, search functionalities, maintenance and historical access, and public access - were designed to ensure that platforms facilitated independent research and the monitoring of paid disinformation and election influence. This initiative built on the EU Code of Practice on Disinformation, which had set voluntary commitments for industry on ad transparency.

Today, most of these platforms have public ad repositories; but the DSA has set an important legal standard for the basic functionality and data they need to provide. This harmonized transparency requirement has come just in time for the June 2024 European Parliamentary elections. But the effectiveness of these repositories in helping researchers investigate harms like disinformation hinges on the usefulness of these tools in practice.

[3] "Facebook and Google: This is What an Effective Ad Archive API Looks Like" (Mozilla, March 2019), https://blog.mozilla.org/en/mozilla/facebook-and-google-this-is-what-an-effective-ad-archive-api-looks-like/

This research aims to assess the overall readiness of these repositories for researchers, journalists, and other watchdogs as we head not only into this June's European elections, but into elections around the world in 2024, in which roughly half the world's population are eligible to vote.

While we find there is still much room for improvement with these repositories, we are encouraged to see many of the points from our guidelines were taken up by the platforms, following the passage of the DSA. When Mozilla and partners released the 2019 guidelines, Facebook and Google had just pledged to launch public ad archive APIs. Now, 11 of the world's largest tech companies have ad repositories. The road here has not been smooth, but this is progress.

For this research, we combined our understanding of the DSA's requirements in Article 39 and Mozilla's five guidelines from 2019 into a single set of criteria for assessment. As before, our aim is to ensure that these transparency tools will be useful in practice for researchers and the public. It is up to the regulator to assess legal compliance with the DSA's requirements - we cannot and do not aim to do this. We hope that this report can help to guide regulators and policymakers, and support the wider research community - as well as to encourage VLOs to continue to improve their transparency over their advertising practices.

# 2. Scope

## a. Platforms Under Review

We examine the ad repositories of the following VLOs:[4] AliExpress, Apple App Store, Bing, Booking.com, Alphabet (Google Search and YouTube), LinkedIn, Meta (Facebook and Instagram), Pinterest, Snapchat, TikTok, X, and Zalando.

## b. Criteria for Assessment

Our criteria were inspired by two sources: Article 39 (1-2) of the DSA, which obliges VLOs to compile a searchable ad library, and Mozilla's expert analysis on how to build effective APIs for these ad libraries.

---

[4] We did not examine Amazon Store because Amazon had been granted an exemption from making its ad repository publicly available by the European Court of Justice. See "Order of the President of the General Court" (Case: T-367/23 R), 27 September 2023: https://curia.europa.eu/juris/document/document.jsf?text=&docid=277901&pageIndex
We did not do a separate evaluation for Alphabet-designated services Maps, Play Store or Shopping.
Wikipedia does not have ads so it is not studied.
We do not study the platforms designated by the EU Commission later than April 2023. See " Digital Services Act: Commission designates first set of Very Large Online Platforms and Search Engines", European Commission, 25 April 2023: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413

## DSA Article 39:[5]

Article 39 requires VLOs to compile and make publicly available a repository containing detailed advertisement information. This repository should include:

- **Content of The Advertisement**: The repository should include the content of the advertisement, the name of the product, service, or brand, and the subject matter of the advertisement.
- **Advertisement Entities**: Information about the natural or legal person on whose behalf the advertisement is presented and the person who paid for the advertisement.
- **Duration**: The period during which the advertisement was presented.
- **Targeting Parameters**: Whether the advertisement was intended for specific groups of recipients and the main parameters used for that purpose.
- **Commercial Communications**: The content containing commercial communications published on the platforms as per Article 26(2) (ie self-declared influencer content or branded content).
- **Recipient Data (audience reach data)**: The total number of service recipients reached, with aggregate numbers broken down by EU Member State for targeted groups.

## Mozilla's Ad Archive API Recommendations:[6]

**In March 2019, Mozilla and a cohort of independent researchers published five guidelines that ad archive APIs must meet in order to support election influence monitoring and independent research. The experts were based at Oxford University, the University of Amsterdam, Vrije Universiteit Brussel, Stiftung Neue Verantwortung, and other institutions. The recommendations are:**

- **Comprehensive Content**: The API should provide access to both paid political ads and issue-based ads without restrictions based on pre-selected topics or keywords.
- **Detailed Ad Information**: This includes the content of the advertisement, targeting criteria, impressions, engagements, payment details, and microtargeting specifics.
- **Functionality**: The API should support advanced research and analysis, allowing for trend analysis, content downloads, and search functionalities.
- **Data Accessibility**: Advertisements should be available within 24 hours of publication, with historical data access going back at least 10 years. The API should be maintained and promptly fixed if issues arise.

---

[5] Mainly 39.1 and 39.2 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065.
[6] Facebook and Google: This is What an Effective Ad Archive API Looks Like (Mozilla, Mar 2019), https://blog.mozilla.org/en/mozilla/facebook-and-google-this-is-what-an-effective-ad-archive-api-looks-like.

- **Public Access**: Both the API and the data derived from it should be accessible to the general public.

We arrived at the combined and simplified criteria list below, first combined and then differentiated for the web-based repository and for the API:

| **Web-based Ad Repository & API Ad Repository Analysis (Combined)** | |
|---|---|
| **Public Availability** | The extent to which platforms have made an ad repository publicly accessible. |
| **Content of Advertisement** | Verifying if the repositories included detailed information on the content, product, service, brand name, and subject matter of the advertisements. |
| **Advertisement entities (payer and beneficiary details)** | The disclosure of information about the entity presenting the advertisement and the payer behind it. |
| **Duration Reporting** | Assessment of the accuracy and comprehensiveness of the information about the duration of the advertisements. |
| **Targeting Parameters** | Whether the ad repository provides clear indications of whether ads were targeted and the main parameters used for targeting. |
| **Commercial Communications** | The inclusion of a tool for commercial communications content like user-reported ads, influencer content or “branded content” (Corresponding with DSA Article 26(2)). An important difference between this content and “traditional” ads is that one category is self-disclosed. |
| **Recipient Data (Reached Audience Data)** | The availability and accuracy of data regarding the total number of recipients and the breakdown by EU Member State for targeted groups. |
| **Detailed Ad Performance** | Detailed targeting criteria, impressions, engagements, payment details, and specifics on microtargeting. |
| **Search Functionality** | Whether the repository (web and API) supports advanced research capabilities such as trend analysis, content downloads, and enhanced search functionalities |

| | |
|---|---|
| **Data Accessibility** | The availability of ads within 24 hours of publication and historical data access for at least 1 year. |

| **Additional Web Repository Analysis** | |
|---|---|
| **Accuracy** | Ensuring that the data presented in the web-based ad repository mirrors the actual ads displayed on the platform. |
| **Reliability (Data Consistency)** | Observing consistency in performance, both in terms of data availability and search results over time, by repeated browsing of the ad repository over several days to search for new ads and to reiterate previous searches. |
| **Filtering And Sorting Capabilities** | Measuring the efficiency and effectiveness of filtering functions (ability to filter the results of a search, e.g. by recency, performance) and sorting (ability to rank the results of a search) capabilities. |
| **Tooltips And Help Features** | The availability and effectiveness of tooltips (tools on the interface), Help / FAQ sections, and other guidance features |
| **Feedback Mechanism** | Assessment of the ease with which users can provide feedback or report issues. |
| **“Distance” to The Repository** | Whether users can reach the ad repository from an ad displayed on the platform and how many clicks it takes. |

| Additional API Repository Analysis | |
|---|---|
| **Reliability** | Consistency of the API's performance over time including during periods of high query volumes. |
| **Completeness** | Assessing whether data points described by DSA Article 39 and Mozilla's criteria are consistently available for each advertisement. |
| **Responsiveness** | Responsiveness of the API, focusing on its ability to provide timely feedback and results to queries under different load conditions. |
| **Data Integrity** | Ensuring repeated and varied queries did not lead to data corruption or loss. |

# 3. Methodology

## a. Limitations of this research

**We want to be clear on the scope and limitations of this research.**

**We designed our tests in November 2023 and conducted our tests between December 2023 and January 2024. VLOs have made changes after the launch of our tests. We monitor and note developments following the start of our testing period (December 1st), like the appearance of new APIs, but we have not been able to systematically test these developments.**

We have tested the web-based ad repositories and APIs, but we have not scrutinized the content of the information provided. In other words, we look at things like accessibility, functionalities, and completeness (whether or not most ads on the platform appear in the web repository and API) - but we do not assess the accuracy or truthfulness of the information, for example, whether the beneficiary information is correct. We therefore do not assess whether the ad repos and APIs accurately include branded or influencer or political content, also because allowance and policies vary.

We do not assess the quality of the VLO repositories with respect to influencer or branded content, corresponding with the obligations under DSA Article 26(2), which states that "*providers of online platforms shall provide recipients of the service with a*

*functionality to declare whether the content they provide is or contains commercial communications*."[7]

Assessing the accuracy of the repositories with respect to influencer content would require a different methodology(See Mozilla's 2021 research into paid political influence on TikTok[8]). More recently, independent researchers at the German think tank Stiftung Neue Verantwortung released in-depth research into TikTok's DSA ad repository and its inclusion of "other commercial content". They outline several issues, including that "only a very small number of users use this functionality" and that there is limited information about the ads and interests behind them.[9]

We note that only Meta, LinkedIn, YouTube, and Snapchat have separate branded content repositories and that TikTok has a dedicated branded content API, in addition to having a separate branded content web repository. Pinterest has a field indicating yes/no with respect to "commercial content" in its main web repository. The other platforms do not have this kind of content in their repositories at all. This may be because they do not allow influencer content on their platforms (likely the case for Apple App Store, Booking.com, Bing, Google Search), but we would expect Zalando, X, and AliExpress to have this content.[10]

Scrutinizing the inclusion and accuracy of influencer content is a critical area of future research, as is clear from the results of the European Commission sweep which found that out of 97% of influencers publishing posts with commercial content, only 20% systematically disclosed this.[11]

Similarly it was beyond our scope to assess the accuracy and completeness of the repositories and APIs with respect to "political content". Currently, policies on including political content, and definitions of political content, vary across the VLOs. During our testing, we found that Meta, Alphabet, and Snapchat allow political advertising on their platforms in the EU and distinguish these ads in their ad repositories and APIs. Others prohibit it.[12] The European Institutions have recently agreed on regulation on the transparency and targeting of political advertising which may change this situation.[13]

---

[7] DSA Article 26.2(2)
[8] These Are "Not" Political Ads › About the methodology (Mozilla, 2021), https://foundation.mozilla.org/en/campaigns/tiktok-political-ads/about-the-methodology/
[9] Tik-Tok, DSA O'Clock? (Auditing TikTok - Alexander Hohlfeld, Anna Semenova, Martin Degeling, Greta Hess, Kathy Meßmer, Feb 2024) https://tiktok-audit.com/blog/2024/Tik-Tok-oclock/
[10] It is our understanding that AliExpress, X, and Zalando have influencer programs of some kind.
[11] Investigation of the Commission and consumer authorities finds that online influencers rarely disclose commercial content (Europa Press Corner, Feb 2024), https://ec.europa.eu/commission/presscorner/detail/en/ip_24_708.
[12] When contacted prior to publication, Pinterest emphasized that they prohibit political advertisements.
[13] Transparency and targeting of political advertising: EU co-legislators strike deal on new regulation (European Council Press Releases, Nov 2023), https://www.consilium.europa.eu/en/press/press-releases/2023/11/07/transparency-and-targeting-of-political-advertising-eu-co-legislators-strike-deal-on-new-regulation/.

## b. Data Collection

We conducted our testing between December 2023 and January 2024.
Our data collection process was structured as follows:

### 1. Accessing Repositories

- **API Integration**: Use the application programming interfaces (APIs) provided by each VLO to access their ad repository.
- **Automated Retrieval**: In cases where APIs might have limitations or are not provided, employ automated retrieval techniques to extract data directly from the platform's ad repository interface.

### 2. Data Points Collected

- **Advertisement Content**: Retrieve the content of the advertisement, including visuals, text, and any multimedia elements.
- **Product/Service Details**: Extract the name of the product, service, or brand advertised.
- **Advertisement Entities**: Identify the natural or legal person on whose behalf the advertisement is presented and the one who financed it.
- **Advertisement Duration**: Record the period during which the advertisement was active on the platform.
- **Targeting Information**: Gather data on whether the advertisement was intended for specific groups and the parameters used for such targeting.
- **Recipient Data**: Collect data on the total number of service recipients reached and any available breakdown by EU Member State or other demographics.

### 3. Data Storage and Organization

- **Database Integration**: Collected data is stored in a structured database, ensuring easy retrieval, analysis, and assessment.
- **Data Categorization**: Data is organized by platform.

### 4. Data Quality Assurance

- **Data Verification**: A subset of the collected data was cross-checked against the actual advertisements displayed on the platforms to ensure accuracy.
- **Handling Inconsistencies**: Algorithms were used to detect and handle any inconsistencies or anomalies in the data, ensuring the integrity of the dataset.
- **Data Update Mechanism**: We updated this data once per day, capturing any new advertisements and changes to existing ones.

- **Data Retention Policy:** The collected data will be maintained for the duration of the project, followed by archiving on CheckFirst's server for a one-year period.

## c. Stress Testing

These stress tests aimed to uncover any potential vulnerabilities, inefficiencies, or shortcomings in the ad repositories and identify areas for improvement for real world use. Stress testing was used to evaluate the robustness, reliability, and effectiveness of the ad repositories under various conditions and loads, to mimic real-world demands and challenges.

### Here is our approach:

#### 1. Complexity Test

**Objective**: Evaluate the repository's capability to handle multi-criteria queries, which are complex and demand more processing power.
**Execution**: Craft queries that combine multiple search criteria and filters, then measure the accuracy of the returned results and the time taken to process these queries.

#### 2. Accuracy Test

**Objective**: Ensure that the data retrieved from the repository matches the actual ads displayed on the platform.
**Execution**: We randomly captured 100 ads on each platform between December 2023 and January 2024. 24 hours after capture we tried to find the same ads in the repository, searching by keyword or advertiser.

#### 3. Reliability Test

**Objective**: Determine the consistency of the repository's performance over time.
**Execution**: Query the repository at different intervals (e.g., peak hours, off-peak hours) and compare response times, data accuracy, and system stability.

#### 4. Completeness Test

**Objective**: Ensure that all required data points, including VLOs obligations under DSA Article 39(2) and Mozilla's API recommendations, are available for each ad.
**Execution**: For a random set of advertisements, check for the presence of all required data points and note any omissions.

#### 6. Data Integrity Test

**Objective**: Ensure that repeated and varied queries do not lead to data corruption or loss.

**Execution**: After each test, verify the integrity of the data in the repository, ensuring no inadvertent changes or losses have occurred.

## d. Usability Testing

A system can be technically sound, but if it's not user-friendly, its effectiveness is limited. This usability testing aims to ensure that the ad repositories are also user-centric, since these transparency tools are designed for the public - the DSA specifically requires VLO to make these tools publicly available, searchable and reliable.

### Here is our approach to usability testing:

#### 1. Search Functionality Test

**Objective**: Evaluate the effectiveness and accuracy of the search function.
**Execution**: We performed a series of searches using various criteria and filters, then measured the relevance and accuracy of the returned results.

#### 2. Filter and Sorting Test

**Objective**: Determine the efficiency and effectiveness of filtering and sorting options.
**Execution**: Users applied various filters and sorting options to the ad data, assessing the speed, accuracy, and relevance of the results.

#### 3. Tooltips and Help Features

**Objective**: Evaluate the availability and effectiveness of tooltips, help sections, and other guidance features.
**Execution**: Users access these features and give feedback on their clarity, usefulness, and accessibility.

#### 4. Error Handling

**Objective**: Determine how the system handles errors, such as invalid queries or system overloads.
**Execution**: Intentionally introduce errors like misspellings and observe the system's responses, ensuring they are clear, informative and guide the user towards a resolution.

#### 5. Feedback Mechanism

**Objective**: Evaluate the ease with which users can provide feedback or report issues.
**Execution**: Users attempt to use any available feedback mechanisms, assessing their accessibility, clarity, and responsiveness.

### 6. Documentation and User Guides

**Objective**: Assess the availability and quality of user documentation, tutorials, and guides.
**Execution**: Review any provided documentation for clarity, comprehensiveness, and relevance.

# 4. Some improvements since Mozilla's 2019 Analysis

**In 2019, Mozilla and dozens of civil society organizations urged Facebook to launch an ad archive API to let researchers, journalists, and users understand the advertising on the platform. At the same time, Google pledged to launch a similar tool, also under the EU's Code of Practice on Disinformation. These commitments were critical ahead of the June 2019 EU Parliamentary elections. In March of that year, Mozilla gathered a cohort of experts, including experts from the University of Amsterdam and Stiftung Neue Verantwortung, and published five guidelines for Facebook and Google's APIs, which were then endorsed by more than 70 researchers.**

When Google and Facebook released their APIs in 2019, Mozilla scrutinized them carefully. We found Facebook's API to be inadequate,[14] since it did not provide sufficient data to facilitate the work of researchers monitoring disinformation. In particular, we found that Facebook's ad API provided no information on ad criteria and engagement, preventing researchers from understanding who advertisers were paying to reach, and how successful they were in their attempt to influence. We also found that Facebook's API constrained researchers through arbitrary and unnecessary restrictions such as search rate limits. We weren't able to assess how complete the content was at the time since the ad library didn't allow for filtering, bulk downloads, or the identification of ads via a unique identifier. We found that Google's 2019 API was much better compared to Facebook's,[15] but it was still far from sufficient. We found it usable, since it allowed for filtering and bulk downloads, but like Facebook, it didn't include targeting criteria or engagement data. (See Mozilla's 2019 log documenting design limitations, data errors, and bugs in the APIs that made consistent analysis essentially impossible).[16]

Since 2019, platforms have made a number of positive changes to their ad archives and APIs. Many of the things we asked for and that were first set out in the EU Code of Practice on Disinformation have become enshrined in law through the DSA. We

[14] Facebook's Ad Archive API is Inadequate (Mozilla, 2019), https://blog.mozilla.org/en/mozilla/facebooks-ad-archive-api-is-inadequate/
[15] Google's Ad API is Better Than Facebook's, But… (Mozilla, May 2019), https://blog.mozilla.org/en/mozilla/googles-ad-api-is-better-than-facebooks-but/
[16] Data Collection Log — EU Ad Transparency Report (Mozilla, 2019), https://adtransparency.mozilla.org/eu/log/

welcome the fact that both Google and Facebook have included ad targeting criteria and engagement data, which represents a significant improvement from the tools they had in 2019. They also have kept their promises to include historical data, going back six years so far. Facebook (now Meta) also allows for filtering, and this time we were able to run a test for comprehensiveness.

But these tools still fall well short of what they are intended to do: allow researchers and the general public to comprehensively assess the role of commercial advertising and paid influence on their services. While Meta and Google's ad transparency tools may be among the most advanced of those we study, they've had a considerable head start, and they still have a long way to go. In our view, neither one provides an example to follow. We also hope they maintain historical data, since under the letter of the law, they would only be required to maintain this for one year.

# 5. Analysis & Testing

**Both web-based ad repository (hereafter Web Repository) and API-based ad repository (hereafter API) are mentioned in the DSA. In our experience they sometimes have different data and functionalities, so we test them separately.**

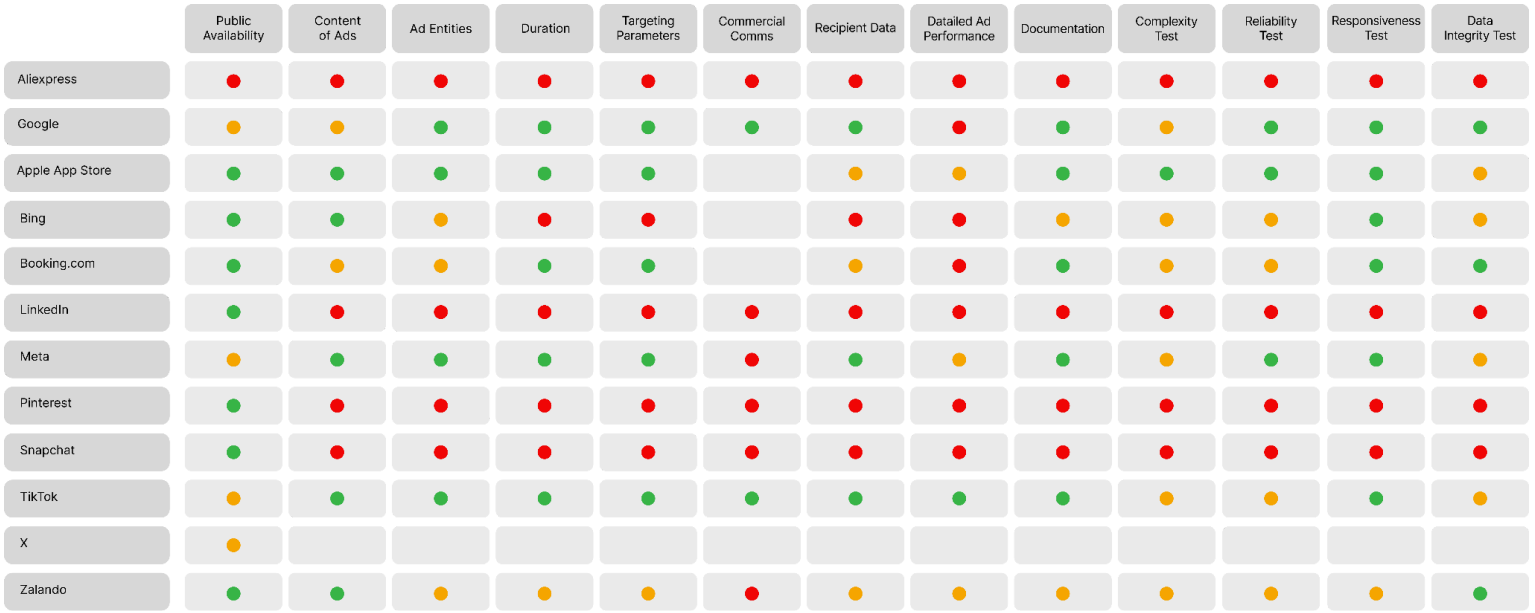
This analysis is structured in five parts:

1) an overview analysis of the web-based ad repository.
2) a discussion of the web-based repository stress test results.
3) an overview analysis of the API (if available during our testing period).
4) a discussion of the results of testing the API (if available during our testing period).
5) a list of recommendations.

*[Our evaluation scheme]*

| | | | | |
|---|---|---|---|---|
| Does not exist | Lacks vital data and functionality | Bare minimum data and functionality | Still has big gaps in data and functionality | ready for action! |

## Find our comparative table here:



# AliExpress

***The absolute bare minimum. AliExpress's repository has no API,a very minimal user interface and requires an account to access. During our tests we encountered loading and display errors and were blocked several times by anti-bot tools.***

## 1. Web Repository Analysis

AliExpress's ad repository requires a user account, limiting public accessibility. It presents only basic ad elements like images and product description.

The repository includes information about ad duration, targeting parameters and recipient/audience data, including geographic targeting and aggregate user numbers.

While AliExpress covers comprehensive content, it lacks more detailed information on targeting criteria, impressions, and payment modalities, which are crucial for a complete understanding of its advertising practices.

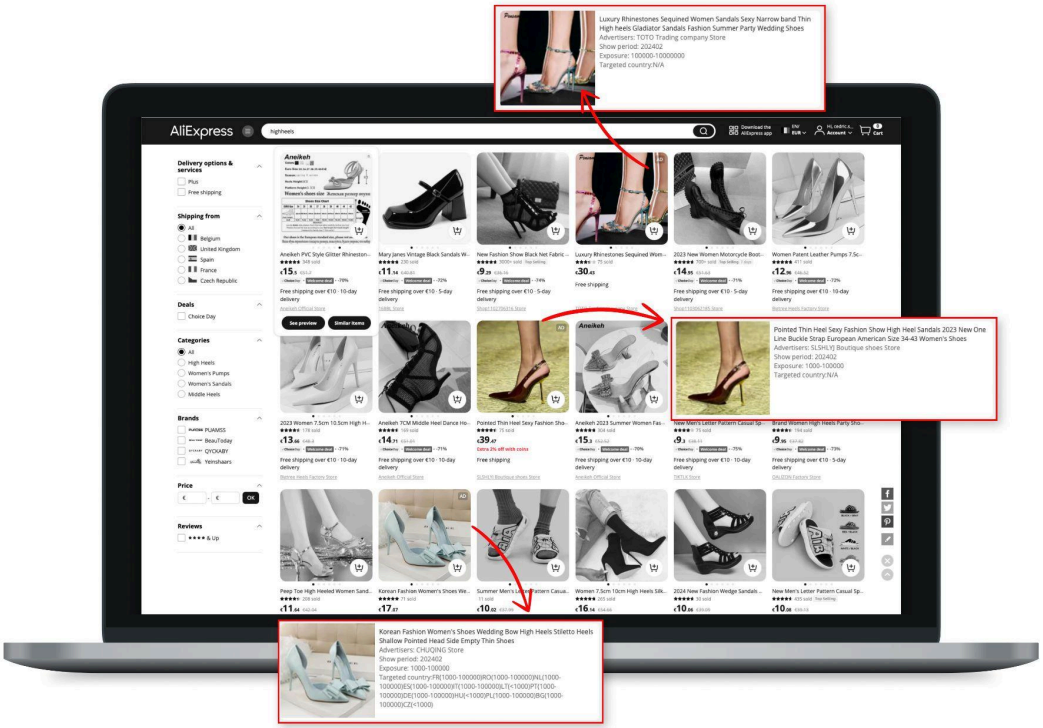
Historical access is available from February 2024.

The absence of a public API limits accessibility and usability.

## 2. Web Repository Testing

The AliExpress ad repository offers satisfactory search functionality, allowing users to search by country, advertisement titles, and advertiser, with the possibility of combining two of these search criteria.

We were able to find 99 of the 100 ads we studied in the interface in the repository. However, for 65% of our set , targeting criteria were either not available or missing country-level data. This suggests there may be a need for more precision in the repository's data accuracy.




Reliability was another area of concern: we found inconsistencies in performance over several days. For instance, we encountered display and loading errors, and were even blocked from using the repository because of scrolling too fast and setting off a bot-detection tool. This variability affects the usability of the tool and the trustworthiness of the data, and suggests possible implications for data management and update processes.

The repository does not have filtering and sorting functionalities. While testing, we were often hindered by rate limits (limiting the number of queries we could conduct), which we feel reduce the usefulness for someone requiring immediate access to the library.

There are no support features or tooltips (tools on the interface to help researchers using the repository), which help new or detailed-oriented users. Another drawback is the lack of feedback and report mechanisms, which are important for continuous improvement and addressing user needs effectively.

## 3. API Analysis

No API as of March 18, 2024

## 4. API Testing

No API as of March 18, 2024

## 5. Recommendation

| Web Interface | API |
|---|---|
| Ensure complete disclosure of ads, ad entities, including detailed information about micro-targeting strategies and funding sources. | Create a public API for transparent and accessible ad data, integrating detailed information like ad content, targeting, and historical access up to 10 years. |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | |
| Ensure easy and free access to ad data for all users without an account. | |
| Improve the user interface for better navigation and user experience. | |
| Establish systems for collecting and incorporating reports and user feedback. | |
| Incorporate tooltips and a help section for immediate assistance and improved user understanding and navigation. | |
| Extend the availability of historical ad data to at least 10 years. | |

# Alphabet: Google Search & YouTube

***We see progress, but it's been six years and we still can't search by keyword in the web -based repository.***

Note: Google Search and YouTube ads are stored in the same repository.[17] It's not possible to filter by Google Product using the repository interface although it is possible to search ads by type (text, image or video).

[17] We had difficulties studying ads related to Google Play Store, Maps, and Shopping, so we excluded them from this research. We believe that Google Play Store ads are also stored in this repository, but we found no way to filter the ads to see only the ads from Play Store.

## 1. Web Repository Analysis

Google's ad repository is accessible to the public (no account is required). The repository, focused solely on advertisement content, provides essential information about the advertisers, including their legal names, locations, and verification statuses. Google also discloses ad duration.

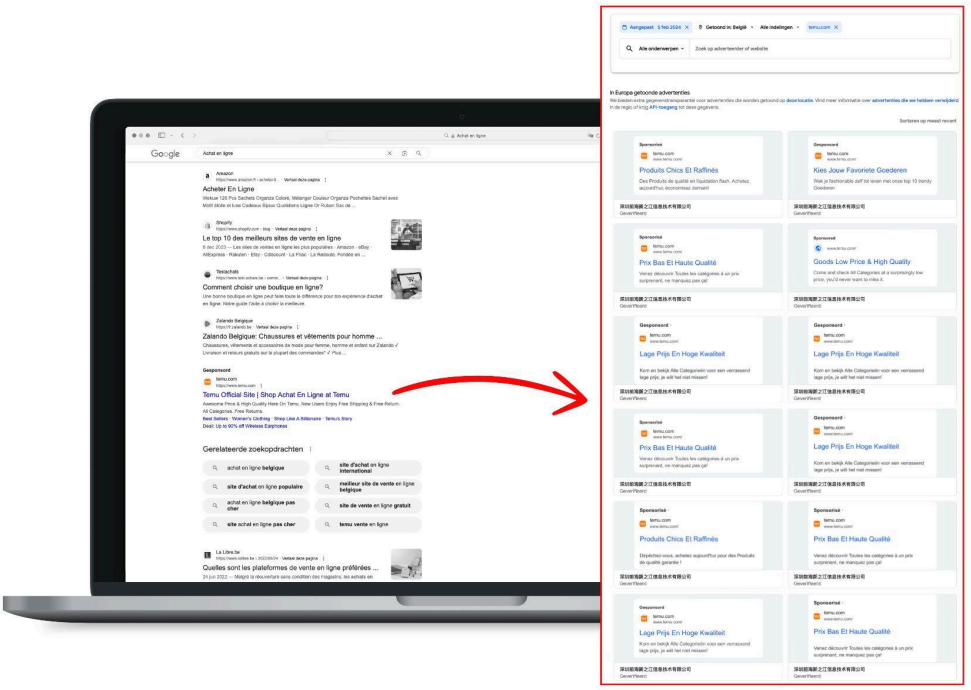
Although targeting parameters and recipient data are somewhat limited and sometimes totally absent, Google's inclusion of country-specific targeting and aggregate recipient data offers some insight into ad reach and targeted demographics. The repository offers a search functionality allowing for multiple search parameters and historical data access back to 2018 (six years).

In the case of YouTube, we note that this is one of the few repositories that includes branded or user-declared commercial communications content like influencer content - though this is not scrutinized in our report.

## 2. Web Repository Testing

Google's ad repository search functionalities performed well on our tests. Users can also reach the repo directly from an ad they have been exposed to on the platforms in three clicks.

Our accuracy test showed mixed results: for YouTube, we found 100% of our ads in the repository, but for Google Search, only 67%. This figure reflects our ability to match an exact ad captured the previous day with a single ad in the repository. It's possible this lower accuracy figure for Google Search could be due to time delay (some ads taking 48-72 hours to appear) or the ads being a responsive search ads, "one of many variations served to users and can include different combinations of text, images, and links that employ slight variations", as explained in their Ads Transparency Center FAQ.



The repository's reliability is high and filter options are adequate, but the lack of sorting options impede usefulness. Critically, there are no keyword-based searches possible on the web repository - users may only search by domain and advertiser name.

The repository includes a help page and tooltips, and users can provide feedback or report ads directly. We also note that the repository mixes advertisements from the different platforms owned by Alphabet without providing a way to easily filter by platform.

## 3. API Analysis

Google's API is accessible exclusively to users with a Google account. While it provides an extensive range of fields, it does not disclose the actual content of the advertisement; instead, it offers a URL that directs to the web repository. The dataset is comprehensive, including details about the advertiser, the duration of the ad, recipient data, and a limited set of targeting parameters. The search functionality allows for a wide range of queries since each field is searchable. Data access extends beyond a one-year period.

## 4. API Testing

Google's repository stands out as the only one designed using BigQuery, Google's proprietary service. Although the dataset can be queried through numerous fields, as with the web repo, we could not conduct keyword searches. We found the API to be

stable and reliable over time. There is thorough documentation available in multiple languages, complete with practical code examples.

### 5. Recommendations:

| Web Interface | API |
| --- | --- |
| Ensure complete disclosure of ads, ad entities, including detailed information about micro-targeting strategies and funding sources. | Enhance ad data by providing extensive details on ad content, ad duration, targeting strategies, audience metrics, and advertiser disclosures. |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options, and allow keyword searches | Remove access restrictions, and loosen the strict rate limits to facilitate more efficient data extraction. |

## Apple App Store

*We found the web repository and API lack important details for understanding paid influence, like targeting broken down by country.*

### 1. Web Repository Analysis

The Apple App Store ad repository offers unrestricted access to its ad repository, featuring a range of content and visual elements like “screenshots and app previews”. The platform discloses the advertisement entities, including app details (e.g. app name, app subtitle, app label) and developers' legal names. It also discloses ad duration.

However, the targeting parameters are limited and recipient data numbers are not provided. While the Apple App Store offers comprehensive content - it provides extended detailed ad information, such as placement and format - it lacks information on useful aspects like reach and engagement.

The platform offers one year of historical data.

## 2. Web Repository Testing

The repository has a useful search functionality, allowing users to search by developer or app, country or region, and date range, and to combine these parameters for more targeted queries.

However, in some cases we could not be sure that we located the specific ad we saw due to a problem with ad content differentiation, i.e. the presence of multiple identical ads for the same app in search results. This suggests a need for a more unique presentation or search capability.

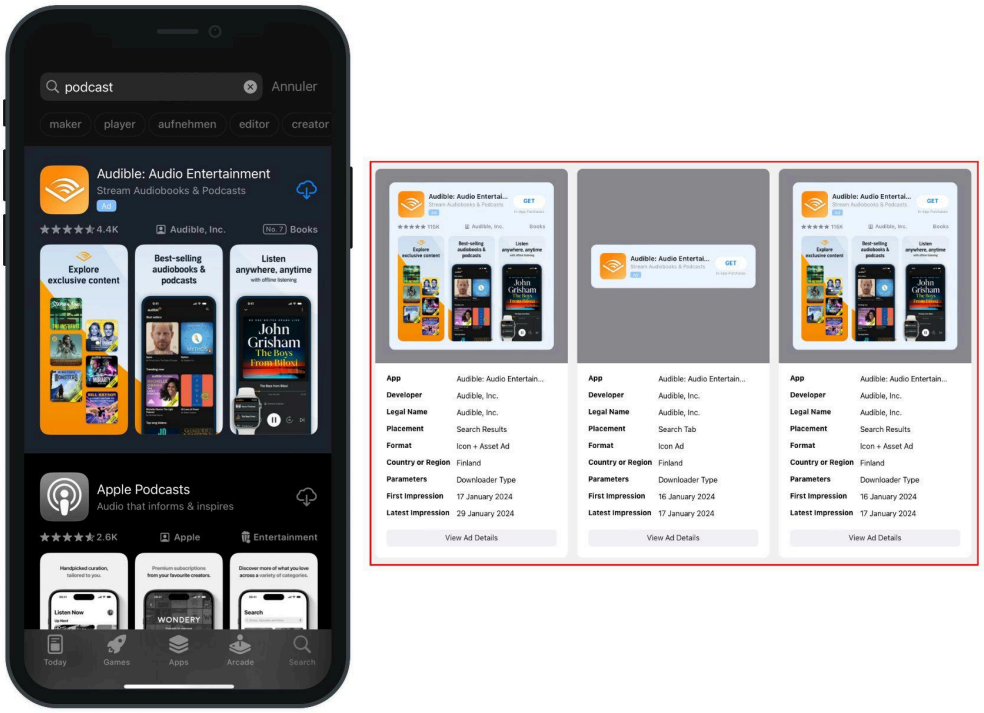
The repository's performance is reliable and consistent, but lacks filtering and sorting capabilities. Support features are also somewhat lacking: there is a help page but there are no tooltips or feedback and reporting mechanisms.

## 3. API Analysis

Apple's API is freely accessible, with no need for registration or restrictions. It provides a variety of endpoints, each enabling different types of search queries. In the scope of this report, we have concentrated on the 'ad-repository-ads' and 'ad-repository-entities' endpoints to gather information. The API presents a more detailed view of advertisements than the web repository. The data encompasses the advertisement content, the entities involved, and the ad's duration.

While targeting parameters are available, there's room for further detail, like impression data. The API does not provide a detailed breakdown by member states in recipient data.

The data remains accessible for a period of one year.




## 4. API Testing

Apple's repository API is designed using REST architecture and equipped with RSQL endpoints offering comprehensive granularity. This advanced functionality permits users to conduct searches with a range of parameters including keywords, advertisers, country, and date range, either individually or in combination. Throughout the testing phase, the API operated without any errors. However, some fields, such as 'adBanner.promotionalText' and 'adAssets.videoUrl', were absent from the dataset on multiple ads.

The API's documentation is comprehensive and includes examples, although it is available solely in English.

## 5. Recommendations

| Web Interface | API |
| --- | --- |
| Ensure complete disclosure of ads, ad entities, including detailed information about | Enhance ad data by providing extensive details on ad content, ad duration, targeting |

| | |
|---|---|
| micro-targeting strategies and funding sources. | strategies, audience metrics, and advertiser disclosures. |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | |
| Incorporate tooltips and a help section for immediate assistance and improved user understanding and navigation. | |
| Extend the availability of historical ad data to at least 10 years. | |

# Bing

***We were unable to search the web repo with special characters like "é è" and were disappointed by the limited data in the API and limited API documentation.***

## 1. Web Repository Analysis

Bing provides public access to its ad repository, presenting the ads themselves as they appear on the platform. Bing discloses advertiser information including the entity paying for ads and their registered locations. The platform displays ad duration, targeting parameters for recipient data and aggregate numbers broken down by EU Member State.

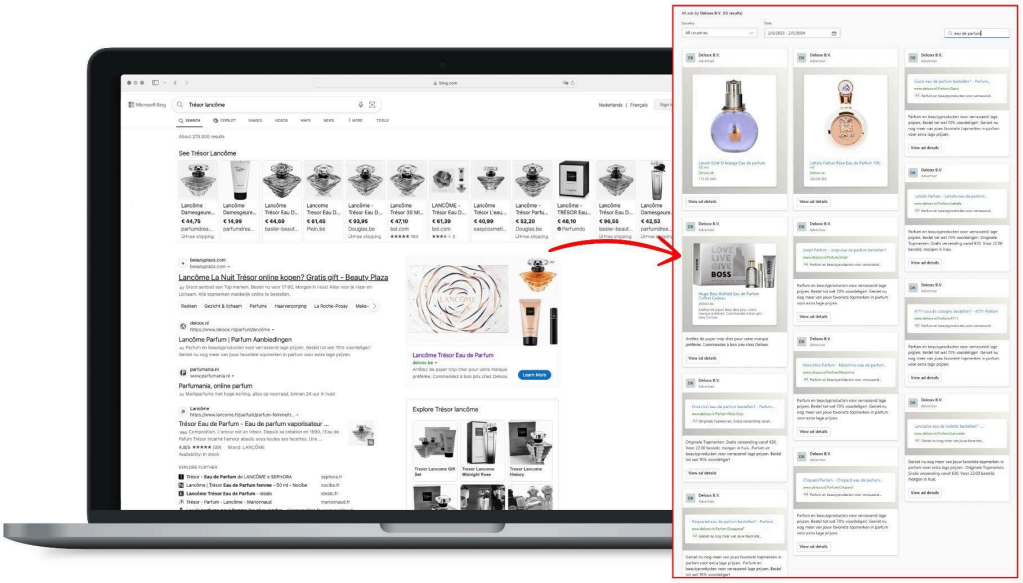
While Bing offers some detailed advertisement information, it lacks comprehensive details on crucial aspects like ad performance (engagements), and payment.

Data is available for up to one year.

## 2. Web Repository Testing

In our view Bing's ad repository's search functionality could be improved. Currently, we can only search by keyword or advertiser, not by country or date. Despite having advanced filtering options (advertiser, country, date, ad content), the repository lacks sorting capabilities, which would help users order the results.

The repository is reliable - providing consistent data and performance over time. But in our accuracy test, we only found about 75% of our ads viewed on Bing.com in the repository.



*Example of ad not found on the repository*

We find that data retrieval within the repository is fast, but the repository does not allow searches by special characters like “ é à è é ï ”.

The repository does not offer tooltips or help features and lacks feedback and report mechanisms.

## 3. API Analysis

Bing's API is open to the public and features a single endpoint for repository searches. The fields it provides match those found in the web repository. While the API discloses advertisement entities by name, it lacks crucial details such as the duration of the ads, targeting parameters, and recipient data.

The data remains available for a period of one year to date.

## 4. API Testing

Bing's repository API is crafted on a REST architecture. Users can query data using filters such as keyword, advertiser, country, and date range, either singly or in conjunction. In terms of content, the fields accessible through the API are consistent with those available on the web repository. We found that the 'AssetJson' field frequently lacks data, so we could not always see the images used in the ad campaign.

The documentation for the API is somewhat minimalistic (a single page with only one curl example of a request) and is only provided in English.

## 5. Recommendations

| Web Interface | API |
|---|---|
| Ensure complete disclosure of ads, ad entities, including detailed information about micro-targeting strategies and funding sources. | Enhance ad data by providing extensive details on ad content, ad duration, targeting strategies, audience metrics, and advertiser disclosures. |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | Upgrade the API and repository functionalities to support advanced research, with simplified access and updated documentation for researchers. |
| Improve the user interface for better navigation and user experience. | |
| Incorporate tooltips and a help section for immediate assistance and improved user understanding and navigation. | |
| Extend the availability of historical ad data to at least 10 years. | |

# Booking.com

***We found it very difficult to link ads on the platform to ads in the web repository, and we found insufficient documentation about the API.***

## 1. Web Repository Analysis

Booking.com offers public access to its ad repository, but the content related to the advertisements within it is limited.

The platform discloses limited information about the entities funding ads. It displays ad duration and aggregate reach numbers broken down by EU Member State, but it does not provide specific details about targeting criteria, using simply a template sentence description for each ad.

Despite comprehensive content coverage, Booking.com lacks detailed ad information, such as the particular groups of recipients targeted.

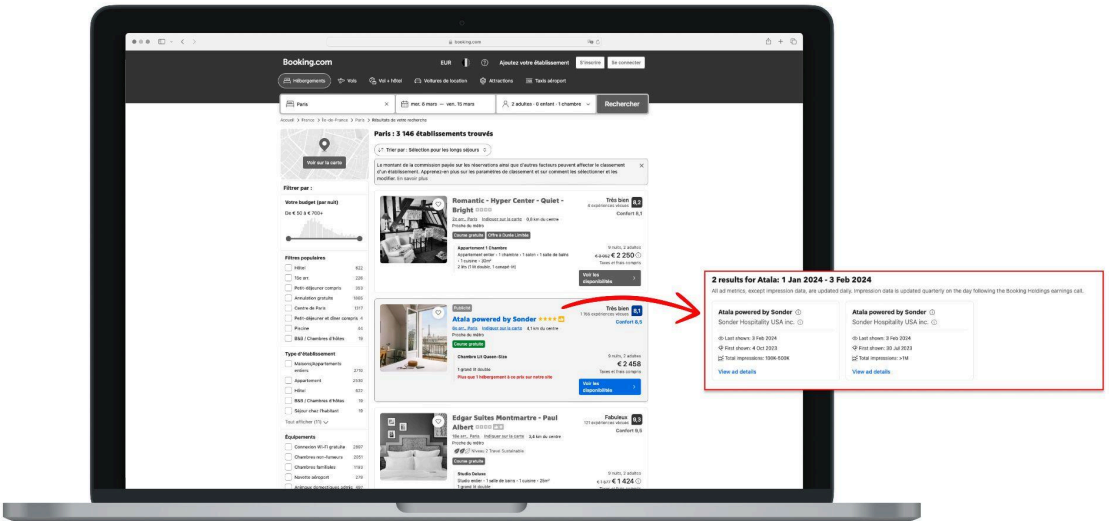
The available data seems to date back to one year (based on the selector) but currently we can access only ads dating from August 2023.

## 2. Web Repository Testing

The repository's search functionality is effective, allowing searches by date, keyword, and the entity funding the ad, and for combined searches for more precise results.

However, in our accuracy test, for all (100%) of the ads studied, we had trouble matching the ads on the platform to the ads in the repository, highlighting a need for more clarity in ad identification. We also found that 23% of our studied ads had no targeting breakdown, which may be because this information is updated on a quarterly basis.

The ad repository is reliable, demonstrating consistent and dependable performance. But we find the lack of filtering and sorting options to limit the usefulness.




*Example of similar ads making it complicated to differentiate*

The repository does not offer a help page and tooltips on some fields. There is no feedback or report mechanism.

## 3. API Analysis

Booking's API is freely accessible to everyone, with no need for registration. Its dataset consists of seven fields, which include the ad's name, the legal entity funding the advertisement, and its duration. However, the API does not include detailed targeting information. Recipient data is available and is broken down by country. The search functionality of the API allows basic queries based on date, ad name, or legal entity. The data remains accessible for a period of one year.

### 4. API Testing

Booking's ad API is built using GraphSQL architecture and features a single endpoint that handles search options. However, it's not possible to search results by date or country. The API experienced frequent downtime during our testing, but overall we found it efficient and consistent. Documentation is quite limited, consisting only of a single curl (command) example.

### 5. Recommendations

| Web Interface | API |
|---|---|
| Ensure complete disclosure of ads, ad entities, including detailed information about micro-targeting strategies and funding sources. | Enhance ad data by providing extensive details on ad content, ad duration, targeting strategies, audience metrics, and advertiser disclosures. |
| Extend the availability of historical ad data to at least 10 years. | Upgrade the API and repository functionalities to support advanced research, with simplified access and updated documentation for researchers. |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | |
| Incorporate tooltips and a help section for immediate assistance and improved user understanding and navigation. | |

## LinkedIn

*Web version is ok but should improve search functionalities*

Note: We were not able to locate the LinkedIn Ad API when we began our testing and for this reason we are not able to include it in our research. This analysis is limited to the web repo.[18]

### 1. Web Repository Analysis

LinkedIn's ad repository is publicly accessible and focuses on displaying advertisement content. We find this helps understand the ads' nature and messaging. The platform also discloses advertiser names and when the ad was active.

[18] After reaching LinkedIn before publication they declared to have launched their API prior to the DSA deadline. Authors of this report were not able to locate any documentation or link to this API at the time of our testing. The API is available here.

LinkedIn's targeting parameters, such as country, language, location, and company, provide a detailed view of audience targeting. The inclusion of recipient data offers clear insights into the reach and impact of its ads.

The search functionality is robust, and historical data is available from June 2023.

## 2. Web Repository Testing

Our accuracy test found 100% of ads studied in the repository: we found accurate listings and detailed information on the ads during our testing period.

The repository's reliability is strong, indicating consistent and dependable performance. However, we find the repository lacks both filtering and sorting capabilities.

The repository offers an FAQ on the homepage for assistance, but there are no tooltips and no feedback mechanism.

## 3. & 4. API Testing and Analysis

We were not able to locate the LinkedIn ad API at the start of our testing period so we did not include it in our systematic tests.

## 5. Recommendations:

| Web Interface | API (available but not tested) |
| --- | --- |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | |
| Incorporate tooltips and a help section for immediate assistance and improved user understanding and navigation. | |
| Extend the availability of historical ad data to at least 10 years. | |

# Meta

*We see progress since 2019 but are disappointed by accuracy errors and missing data fields in the web repository.*

Note: Meta's repository includes ads on both Facebook and Instagram - to identify on which of these platforms an ad was shown with the web repository: you first have to search for an ad before you can then filter by platform. This is easier with the API, where you can filter your search by platform.

## 1. Web Repository Analysis

Meta's ad repository is publicly accessible and focuses on the ads themselves as they're shown to users. Meta discloses advertisement entities, including both the beneficiary and the payer, as well as the ad duration.

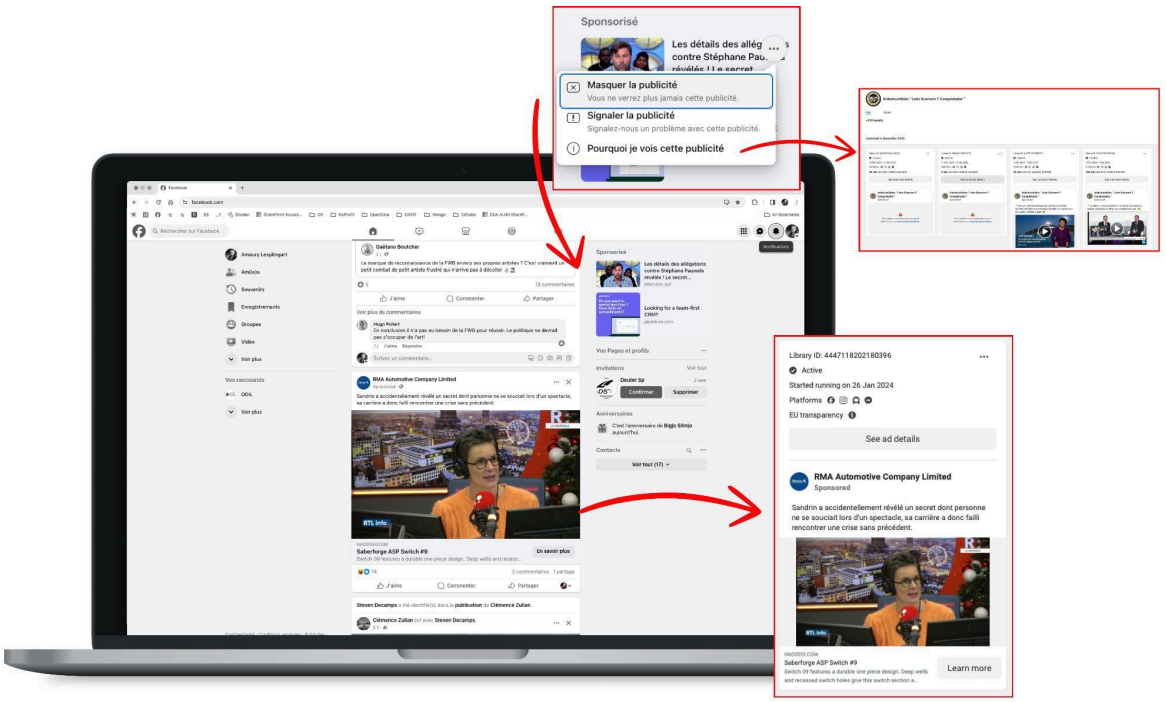
The platform's detailed targeting parameters, like detailed location information, age, and gender, provide a nuanced understanding of ad targeting. The repository includes recipient data, and it is possible to search by country, ad category, keyword, or advertiser, and to combine searches. It holds historical data up to six years.

Meta is one of the few with a separate branded content repository containing content with commercial communications of influencer content - though it is not scrutinized here.

## 2. Web Repository Testing

Our testing finds the search functionality works well. It is also possible to reach the repository directly from an ad the user has been exposed to on the platform in five clicks.

The accuracy test showed mixed results: during our test period, we found 83% of our studied Instagram ads in the repository, and only 65% for Facebook. On Facebook we found beneficiary information lacking on 13% of ads.




*Example of ads not found on the repository*

The repository's reliability is strong and filtering options are comprehensive, covering various criteria like language, advertiser, platform, media type, active status, dates, and keyword. However, we found the absence of sorting options limiting.

The repository offers tooltips and help features, but the feedback mechanism appears to be less useful. There is a “Report as unlawful” option - (which we find confusing and could actually have a chilling effect on reporting if people fear any error on their part) - but no direct way for users to provide general feedback.

## 3. API Analysis

Unlike the web repo, Meta's API is accessible exclusively to users with a Facebook account, a developer account, and - to access political advertisements - also identity verification. This limits accessibility. Once accessed, the dataset is rich, encompassing the content of the advertisement and its associated entities. It reveals the duration of the advertisement, as well as targeting parameters, including age, gender, and location. Moreover, recipient data is detailed, offering a breakdown to the regional level. The data can be accessed for up to one year.

### 4. API Testing

Meta's repository is built on Facebook's internal GraphSQL architecture, which incorporates tokens and user identification for access. This setup allows for queries using various criteria, including keywords, ad type, and reach by country. The API is stable, but during our testing period, we observed discrepancies in the dataset, including fields that were occasionally missing , like bylines, creative link descriptions and delivery by region. This inconsistency may limit the API's usefulness. While there is documentation available in English, complete with code examples, it remains somewhat basic in its scope and depth.

### 5. Recommendations:

| Web Interface | API |
|---|---|
| Ensure complete disclosure of ads, ad entities, including detailed information about micro-targeting strategies and funding sources. | Enhance ad data by providing extensive details on ad content, ad duration, targeting strategies, audience metrics, and advertiser disclosures. |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | Upgrade the API and repository functionalities to support advanced research, with simplified access and updated documentation for researchers. |
| Incorporate tooltips and a help section for immediate assistance and improved user understanding and navigation. | Remove access restrictions, loosen the strict rate limits to facilitate more efficient data extraction. |

## Pinterest

***We found accuracy issues and missing data.***

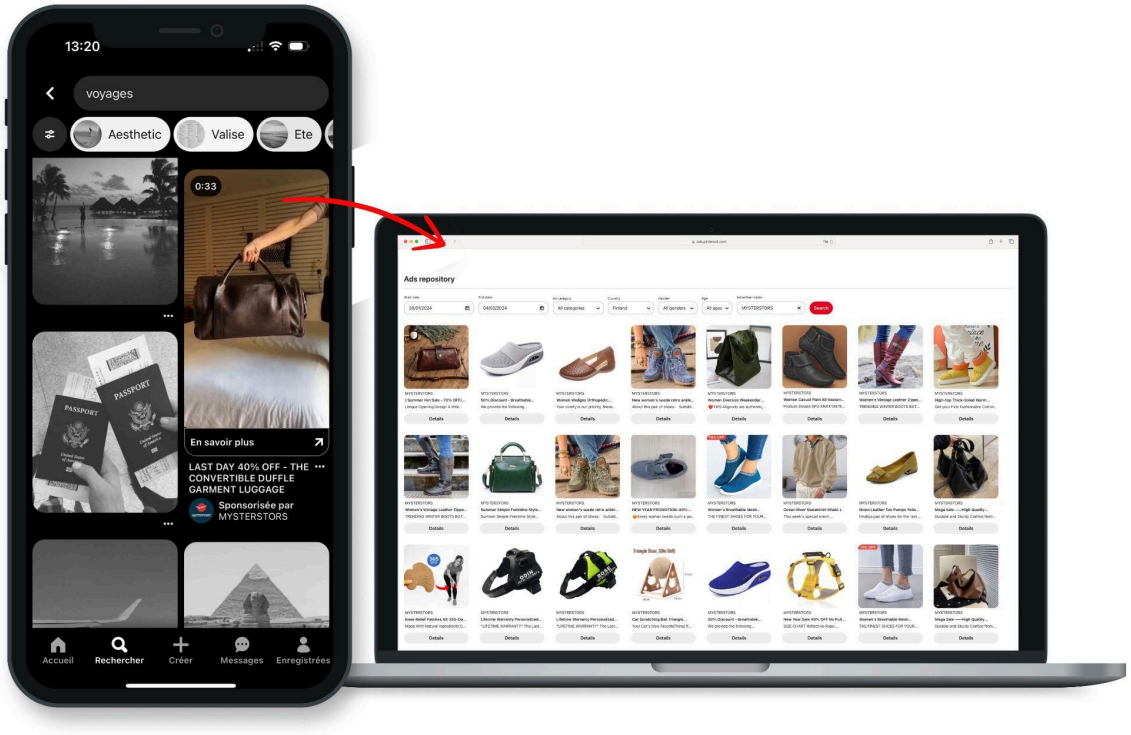
### 1. Web Repository Analysis

Pinterest's web ad repository is publicly available. It features images, titles, and descriptions of ads and discloses advertiser names and ad durations. It offers comprehensive targeting parameters, including age, country, gender, interests, and keywords, and it includes recipient data. Pinterest's search functionality offers users the ability to search by date, category, country, gender, age, and advertiser name, and to conduct combined searches. Historical data is available for up to one year.

Pinterest includes branded or “influencer” content: the repository includes a field called “Commercial Content”, but with a "Yes" or "No" value only without further context.

## 2. Web Repository Testing

The repository's search functionality tests well for us. We encountered only a small accuracy gap: We located 87% of ads in the repository.




*Example of ads not found*

The repository showed consistent and reliable performance over time. However, the repository lacks adequate filtering and sorting capabilities. There is a minimal help page, but no tooltips, and no devoted feedback mechanism.

## 3. API Analysis & Testing

We did not identify an ad repository API during our testing. Prior to the publication of our report, we reached out to Pinterest. They informed us that they had an Ad Repository API and provided us the link. The link provided, however, was for documentation related to now-deprecated version 4 of the API, which, as of April 2022, has been replaced with version 5. As far as we know, there is no manner in which the documentation related to version 4 can be accessed through the developer portal on their website as the aforementioned portal only displays documentation related to version 5, which currently does not include an Ad Repository endpoint.

## 5. Recommendations:

| Web Interface | API |
|---|---|
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | Ad API not assessed |
| Ensure complete disclosure of ads, ad entities, including detailed information about micro-targeting strategies and funding sources. | |
| Incorporate tooltips and a help section for immediate assistance and improved user understanding and navigation. | |
| Establish systems for collecting and incorporating reports and user feedback. | |

# Snapchat

***Search functionality is very limited, and there is currently no API.***

## 1. Web Repository Analysis

Snapchat's repository is publicly available. Visually, it focuses primarily on the ads themselves (showing them as they appear to users on the platform). Snapchat discloses information about advertisers and the organizations charged for each ad, as well as ad durations. Targeting parameters include country, basic demographics, and devices, which enable a detailed understanding of audience-specific ad tailoring. Historical data is available for over one year.

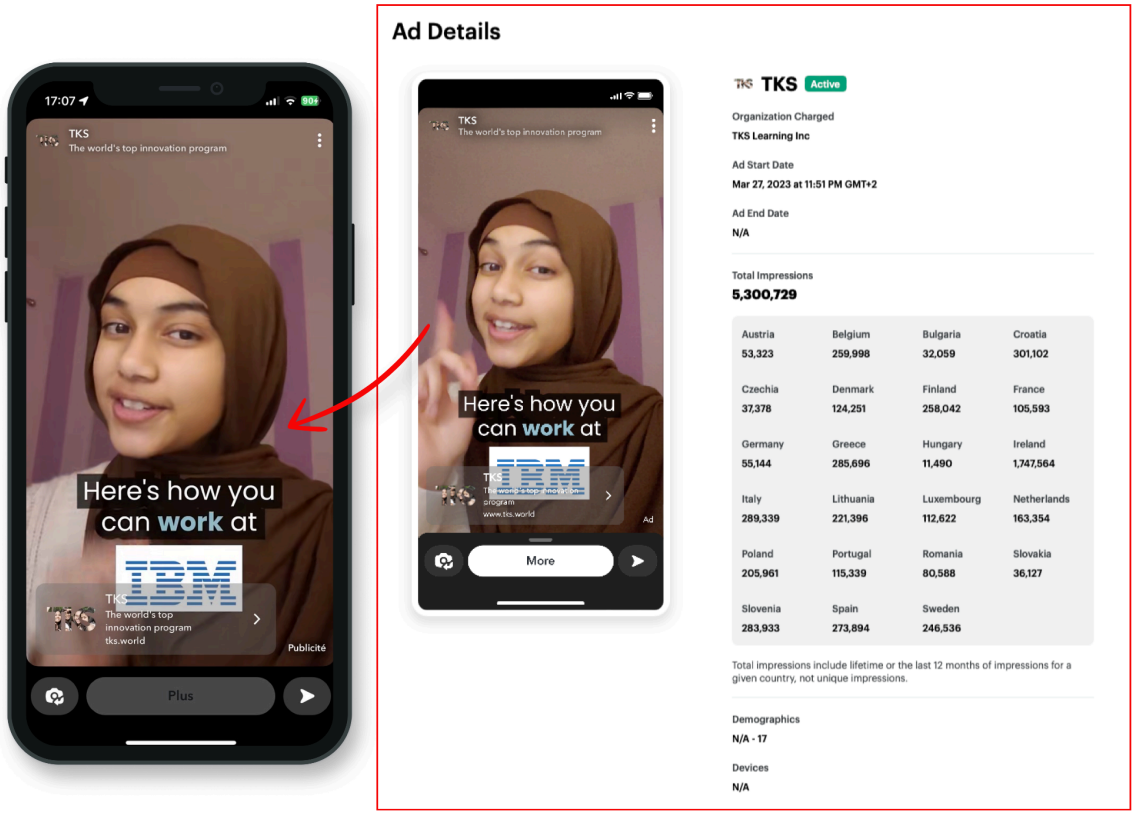
The search functionality allows users to search by country, ad status, date, and publisher, with the additional capability of combined searches.

Snapchat's repository includes content with commercial communications (branded or influencer content) - but we have not scrutinized this.

The absence of a public API limits accessibility and usability.

## 2. Web Repository Testing

Snapchat's repository did well in our accuracy test, correctly reflecting the ads captured in a user feed 100% of the time. The repository's reliability is also strong.




However, the repository's search lacks the ability to look up ads by keywords or anything beyond the exact publisher name as it appears on their Snapchat profile. It also does not offer filtering and sorting capabilities.

We also find the repository lacks user-oriented resources like tooltips, guidance or a feedback mechanism.

## 3. API Analysis

No API as of March 18, 2024.

## 4. API Testing

No API as of March 18, 2024.

## 5. Recommendations

| Web Interface | API |
|---|---|
| Extend the availability of historical ad data to at least 10 years. | Create a public API for transparent and accessible ad data, integrating detailed information like ad content, targeting, and historical access up to 10 years. |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | |
| Establish systems for collecting and incorporating reports and user feedback. | |

# TikTok

*The ad repository and API look robust but we encountered gaps and accuracy errors.*

## 1. Web Repository Analysis

TikTok provides public access to its ad repository. Visually, the repository focuses primarily on the ads themselves. TikTok discloses advertiser information, including the party paying for ads and their registered locations. It also discloses ad durations. It has detailed targeting parameters, such as gender, age, audience, interest and creator interactions.

The search functionality is robust, allowing users to search by country, ad type, date, and advertiser name or keyword, with the additional capability of combined searches.

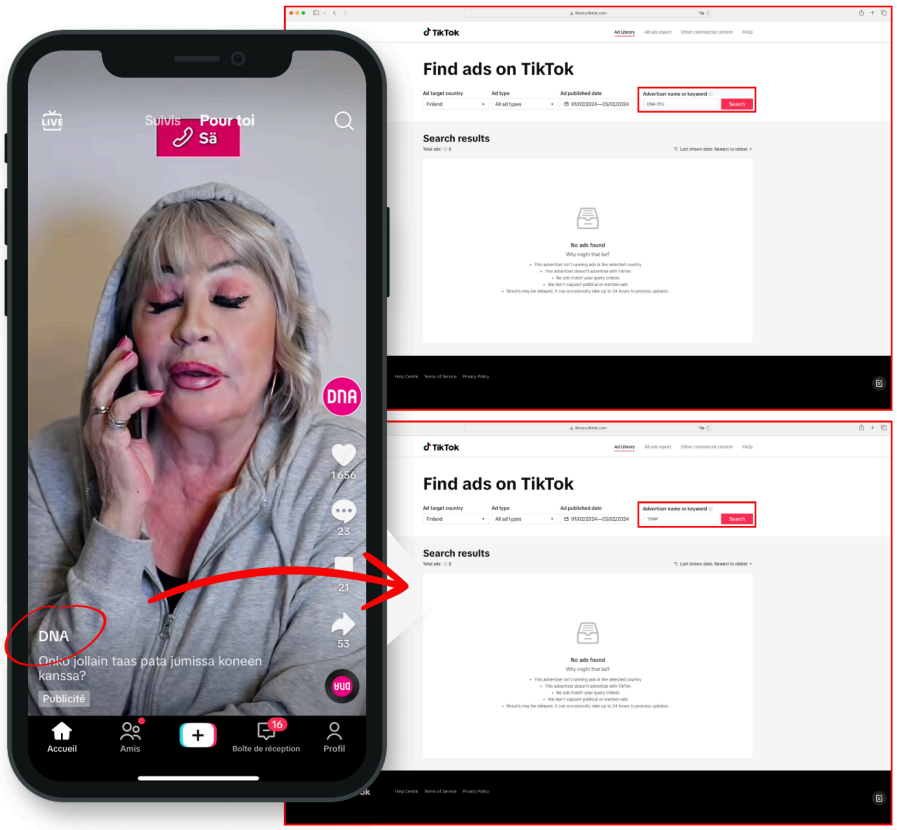
TikTok's repository has a sub-page with content declared by the creator as commercial (branded or influencer content). We do not assess this, but point to recent analysis by Stiftung Neue Verantwortung.[19]

Historical data is available from October 2023.

## 2. Web Repository Testing

When conducting our accuracy test, we found 83% of ads we saw in the "For You" feed captured in the repository.

[19] Tik-Tok, DSA O'Clock? (Auditing TikTok - Alexander Hohlfeld, Anna Semenova, Martin Degeling, Greta Hess, Kathy Meßmer, Feb 2024).
https://tiktok-audit.com/blog/2024/Tik-Tok-oclock/




The repository's reliability is strong, it offers various sorting options but lacks filtering capabilities.

TikTok provides a help page and tooltips. There is a limited feedback mechanism -a yellow flag icon for reporting specific ads - but no direct way for users to offer general feedback.

## 3. API Analysis

TikTok's API is accessible to researchers who register as developers and apply for access to the Commercial Content API. This application process involves detailing the intended research and providing personal information about the developer. The dataset includes a wealth of information, such as the content of advertisements (including URLs of videos and images), extended information on advertisement entities, ad duration, and targeting parameters. TikTok offers exceptional detail on targeting parameters. Data is available for up to one year.

Additionally, it provides an API for branded/influencer content declared by the creator as such.

## 4. API Testing

TikTok's repository is structured using the REST architecture, offering a flexible querying system. Users can search the dataset by various parameters such as keyword, country, date range, or a combination thereof. We found the API reliable but occasionally encountered “internal error” messages. The dataset was consistent, but we found the “ad.image_urls” field often turned out to be empty. This absence deprives researchers of the pre-loading images that are displayed before a video starts to play, which could reduce usefulness. TikTok provides comprehensive documentation in English, complete with code examples.

## 5. Recommandations:

| Web Interface | API |
| --- | --- |
| Ensure complete disclosure of ads, ad entities, including detailed information about micro-targeting strategies and funding sources. | Enhance ad data by providing extensive details on ad content, ad duration, targeting strategies, audience metrics, and advertiser disclosures. |
| Extend the availability of historical ad data to at least 10 years. | |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | |
| Establish systems for collecting and incorporating reports and user feedback. | |

# X

***With only a (slow to load) CSV file instead of a web interface, this was a major disappointment.***

## 1. Web Repository Analysis

X offers public access to its ad repository, but ads are only accessible through the export of a CSV file. X includes information about the funding entities of ads and the duration of ad display. However, there are gaps in targeting parameters and recipient data, limiting transparency into audience reach. The “creative information” (the content of the ad), is not disclosed - only a URL to the ad is available.

The platform's search functionality only offers the ability to search by advertiser, country, and date combined.

X offers the ability to download specific historical Twitter datasets for political advertisements covering the 2018-2019 period, but we were unable to assess the access to historical data. To test this, we would have needed to find an advertiser with historical data, to know already the country where its campaign was displayed, and to know already the time period when it was displayed. In our view this is nearly impossible (though perhaps feasible to scrutinize an ad we purchased ourselves).

## 2. Web Repository Testing

Our accuracy test found 100% of ads we saw also captured in the repository. However, our reliability test found that the CSV took between five and 10 minutes to load, which is dramatically slower than all of the other platforms studied (most took seconds to respond to a query).

The repository does not offer filtering and sorting capabilities immediately (though you can filter after converting the file). It also lacks in providing user assistance, offering no tooltips, guidance or feedback mechanism.

## 3. API Analysis

While X's web repository is publicly accessible, its API was paid access only during our testing period. For this reason we did not include it in our analysis. However, as of March 18, access to the API is now listed as possible with a free account.[20]

## 4. API Testing

The API has not been tested.

## 5. Recommandations:

| Web Interface | API (untested) |
|---|---|
| Ensure complete disclosure of ads, ad entities, including detailed information about micro-targeting strategies and funding sources. | |
| Extend the availability of historical ad data to at least 10 years. | |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | |
| Improve the user interface for better navigation and user experience | |

[20] Twitter API Documentation (X Developer Platform, 2024), https://developer.twitter.com/en/docs/twitter-api

| | |
|---|---|
| Establish systems for collecting and incorporating reports and user feedback. | |
| Ensure the advertisement data is regularly updated and maintained for relevance and accuracy. | |
| Incorporate tooltips and a help section for immediate assistance and improved user understanding and navigation. | |

# Zalando

***The web repository doesn't allow some basic capabilities like combined searches, and we also found only limited information about targeting.***

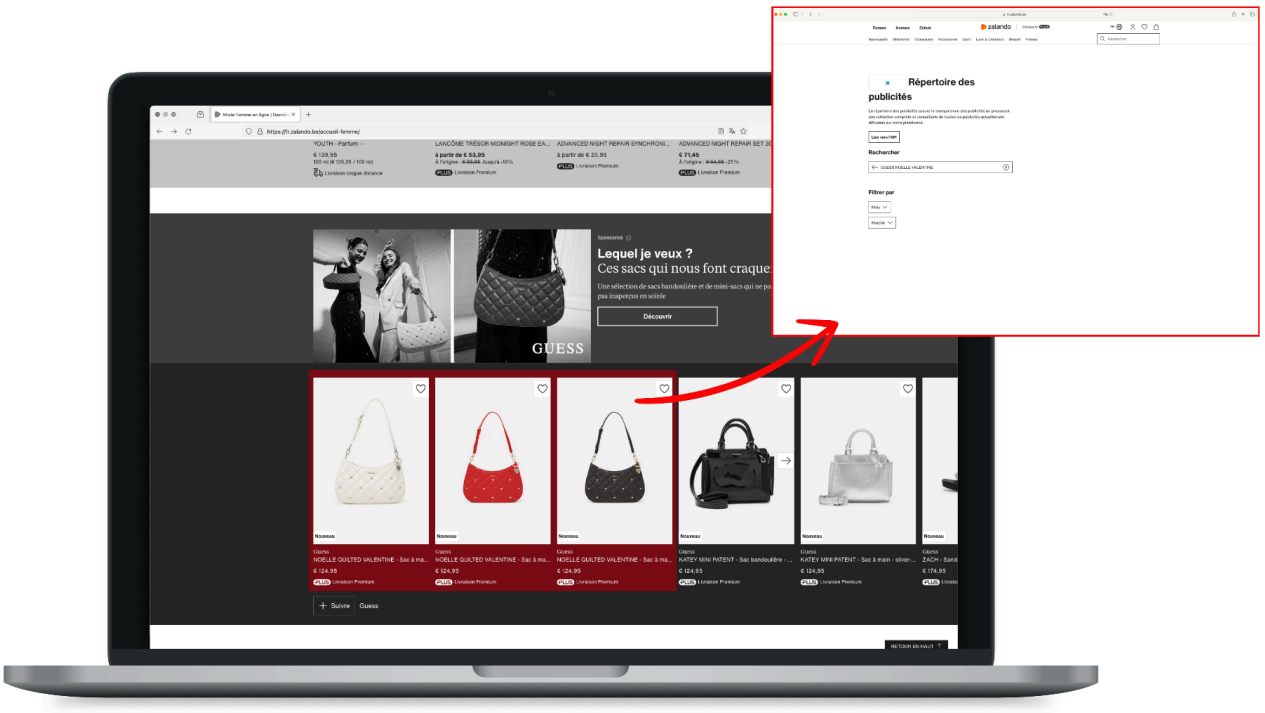
## 1. Web Repository Analysis

Zalando offers public access to its ad repository, which includes both images and text. The platform discloses advertisement sponsors and the duration of ad displays. It includes targeting parameters like market and total reach, but not a detailed breakdown of recipient numbers. Payment details are limited to a single "sponsor" field. Targeting information is a template text used for each ad.

Search functionality exists - users can search by keyword, but the repository does not support combined searches.

Historical data is available for up to one year.

## 2. Web Repository Testing




*Fashion women section with specific ads in the repository.*

We were unable to conduct our accuracy test on Zalando: We struggled to match with certainty any of the ads we saw on the interface to ads in the repository, since ads in the repository are not all presented in the same way.

We also found the repository's reliability weak, several times while conducting our stress test we found the platform was down/unaccessible. While it offers some filtering options like month and market but does not allow for sorting.

There are no support features like tooltips and guidance features and there is no feedback mechanism.

## 3. API Analysis

Zalando's API is open to the public without the need for registration. The dataset provided includes information about the advertisement's content, its duration, and a limited set of targeting parameters. Similar to its web version, the repository reveals the name of the sponsor in the advertisement entities. However, the disclosure of recipient data is somewhat constrained, as there is no breakdown available per member state. The dataset is accessible for up to one year from the current date.

## 4. API Testing

Zalando's repository utilizes GraphQL architecture, offering a querying system that is limited to keyword and time range searches. While the API does not require registration, we found it had blacklisted IP addresses from multiple data centers, requiring us to use a home connection thus sharing our IP address with Zalando. We found the dataset to be comprehensive, with no instances of missing or empty fields.

Zalando does not provide any formal documentation for its API usage. Although there is a link to what appears to be a document on the interface of the repository, it does not offer guidance or instructions.

## 5. Recommandations:

| **Web Interface** | **API** |
|---|---|
| Ensure complete disclosure of ads, ad entities, including detailed information about micro-targeting strategies and funding sources. | Enhance ad data by providing extensive details on ad content, ad duration, targeting strategies, audience metrics, and advertiser disclosures. |
| Extend the availability of historical ad data to at least 10 years. | Upgrade the API and repository functionalities to support advanced research, with simplified access and updated documentation for researchers. |
| Develop advanced search tools to allow for more efficient and comprehensive exploration of advertisements, like filters and sorting options. | Remove access restrictions, loosen the strict rate limits to facilitate more efficient data extraction |
| Improve the user interface for better navigation and user experience. | |
| Establish systems for collecting and incorporating reports and user feedback. | |
| Ensure the advertisement data is regularly updated and maintained for relevance and accuracy. | |
| Incorporate tooltips and a help section for immediate assistance and improved user understanding and navigation. | |

# 6. Key Takeaways From Comparative Analysis

**Approaches to ad transparency vary widely among the platforms studied. Our main takeaway is that even the best approaches don't meet our baseline.**

**We have several general observations following this stress test:**

### Platforms offer various degrees of "public accessibility"

There is a spectrum of accessibility for the web repositories and the APIs, considering factors like the need for a login, the need to apply formally, etc. On one end, platforms like Apple and Booking have free access to the web repository and API even without an account. On the opposite end, Snap and AliExpress had no APIs at all, as of March 18, 2024.

A few repositories are accessible through several clicks from an ad in the interface, but many could only be found by digging into the platforms' terms and conditions.

### Platforms show various ways of disclosing who is behind the ad (entities)

There are many approaches to disclosing the entity behind the ad and it's difficult to assess what is most useful or meaningfully transparent. Meta discloses the beneficiary and the payer, while most platforms disclose the "Advertiser" or "Sponsor" without further context. TikTok, Bing, and Google, for instance, include the registered location of the entity paying for the ad.

### Varying degrees of targeting granularity

There is a great variance in how platforms handle targeting parameters and recipient data. Pinterest, for example, provides extensive targeting parameters, including age, gender, interests, and keywords, and Meta and TikTok offer detailed information on targeting and microtargeting. Meanwhile, others offer simple country and language-based targeting. Zalando and X do not provide a breakdown of recipient data (only totals).

### Variation in search functionalities

A robust search functionality allowing for multiple criteria and combined searches, as seen in Meta's ad repository, is very helpful since it allows the user to locate specific ads efficiently. In contrast, platforms like AliExpress and X's repo's have limited search capabilities.

### Consistency in ad accuracy is a common challenge

Our accuracy testing found many cases where ads in the user interface were not located in the ad repository. This can limit the usefulness and trustworthiness of the repositories as a transparency tool.

### Reliability varies across platforms

While some platforms like Google and LinkedIn show strong reliability, ensuring consistent and up-to-date ad information, others like X in particular were not always consistent, reducing their usefulness dramatically.

### Sometimes filtering, sometimes sorting, but never both

We found that effective navigation through ad repositories was hampered by the lack of filtering and sorting options. While TikTok offers useful sorting options and Meta and Bing useful filtering options, none of the examined VLOs have both sorting and filtering options.

### Tooltips, Help Features, and Feedback Mechanisms are not universal

Many platforms neglect tooltips and help features, which are essential for user navigation and understanding, especially for new users. Many also lacked mechanisms for users to report content or provide general feedback to improve the tool.

### Historical data seems to meet the regulatory requirement, but not necessarily best practice

Most platforms offer historical data up to one year, or, in the case of AliExpress, LinkedIn, and Booking.com, seem to date back to around the time of the DSA obligation coming into force for them. Google and Meta provide data going back six years. But none provides data going back as far as Mozilla's recommends, which is 10 years, to allow for the analysis of long term trends (i.e. over multiple election cycles).

### Paid influencer content or 'branded content' remains elusive

We note that Meta, LinkedIn, YouTube, Snapchat, TikTok, and Pinterest include branded/influencer content in their repositories, though it is not certain that they enforce its inclusion in practice. Previous Mozilla research has found this kind of content is underreported and this problem appears to persist, according to recent research into TikTok by Stiftung Neue Verantwortung[21]. While some platforms may forgo this because they do not allow influencer content on their platforms (likely the case for Apple App Store, Booking.com, Bing, Google Search), we would expect Zalando, X, and AliExpress to have this content.[22]

### A note on data retrieval speed

[21] Tik-Tok, DSA O'Clock? (Auditing TikTok - Alexander Hohlfeld, Anna Semenova, Martin Degeling, Greta Hess, Kathy Meßmer, Feb 2024) https://tiktok-audit.com/blog/2024/Tik-Tok-oclock/
[22] It is our understanding that AliExpress, X, and Zalando have influencer programs or some kind.

Fast data retrieval speeds enhance user experience. This was a more subjective test given the many factors that contribute to speed. However, in our experience, we found most of the repositories loaded data quickly, but that AliExpress was relatively slow and X very slow. This might suggest a need for backend optimization.

# 7. Conclusion

It is absolutely critical that researchers studying online platforms' impact on society have effective tools and access to key data. We conducted these independent stress tests to understand how VLOs could improve these important transparency tools for the research community and the wider public. Rather than assess whether or not these platforms meet the letter of the DSA, we tried to assess how practically useful these tools are for researchers in this year of elections which includes the European elections in mid-June.

This research follows up on Mozilla's research and advocacy from 2019 before the last European elections, when we co-created a set of guidelines for ad archive APIs for Google and Facebook with other experts, bolstered by these platforms' voluntary commitments under the Code of Practice on Disinformation. Now, following their designation under the DSA, 11 of the world's largest tech companies also have ad repositories. This in itself is progress.

None of these ad repositories are fully fit for the purpose of researchers, in our view. We find that AliExpress's repository provides the absolute bare minimum information and the bare minimum interface. X only provides a CSV file for (curiously slow) download, and historical access that's only useful if you already know everything about the ad you're searching for. While we see significant improvements by Google and Facebook from our 2019 assessments, like the critical inclusion of targeting and engagement data (as they are required to do under the DSA), even these don't serve as a perfect example of what makes a good ad repository for researchers or the wider public, given limitations in functionality and access.

The repositories are far from compatible with each other, which makes it challenging for researchers to systematically assess paid influence across multiple platforms. Finally, while we did not include this in our stress test for methodological reasons, we have reason to doubt the usefulness of the branded content repositories for understanding self-promotion and influencer content.

Are platforms taking one step forward, two steps back? We are perhaps most disappointed to see X (formerly Twitter) make such a minimal effort, considering that it remains a central space for civic discourse. This may be why the European

Commission has included X's ad repository in its formal proceedings against the platform under the DSA.[23]

But we'd also like to consider the glass half full, and look forward to further improvements. Each of these platforms play a critical role in our information system, and this year will truly test their usefulness for researchers and watchdogs working to safeguard elections from disinformation and other types of interference.

# 8. Recommendations

We have the following recommendations for providers of online platforms and search engines, as well as for regulators.

## a. Online Platforms

### 1. Recommendations for Web Repositories

The design and functionality of Web Repositories are crucial to usability. Our recommendations for these repositories are centered on creating a user-centric experience. These recommendations are crafted to ensure that Web Repositories are not only accessible and navigable for all users, regardless of their technical expertise, but also rich in features that cater to diverse needs. From unrestricted access to advanced search and filtering capabilities, and from inclusive design to responsive feedback mechanisms, these recommendations aim to transform Web Repositories into intuitive, inclusive, and informative platforms.

1. **Remove access barriers.**
    a. Ensure the repositories are **easily accessible from the main user interface** of the platform and from any ad.
    b. Provide access **without any account or login requirements**, allowing visibility to anyone.
    c. **Allow unrestricted browsing** without rate limits, employing other protection mechanisms if needed.
    d. **Ensure compliance with web accessibility standards** to accommodate users with disabilities (like WCAG).
    e. **Incorporate tools for visual representation** of data to enhance user understanding.
2. **Provide more comprehensive data.**
    a. Make sure that the repository **contains all the advertisements on the platform**.

[23] Commission opens formal proceedings against X under the Digital Services Act (Europa Press Corner, Dev 2023), https://ec.europa.eu/commission/presscorner/detail/en/IP_23_6709

    b. **Ensure data granularity** at least matches the API repository and vice-versa.
    c. **Allow unrestricted browsing** without rate limits, employing other protection mechanisms if needed.
    d. **Provide access to historical data**, with a suggested range of up to 10 years.
3. **Enhance search functionalities.**
    a. Offer enhanced **search capabilities** by, at least, keyword, advertiser, country, and date range. **Include filter and ordering options** for search results.
    b. **Provide options to export data** in common formats like CSV or JSON, mirroring the API's data structure.
4. **Provide better documentation and user support.**
    a. **Offer documentation** in all platform languages, using non-technical, human readable language.
    b. **Implement tooltips** in the interface to help user understanding.
    c. **Include a report feature** for each ad **and feedback options** at both ad and repository levels.
    d. **Establish a system for ongoing user feedback** and regularly update the interface based on this feedback.
5. **Increase harmonization.**
    a. **Aim for a harmonized design** across repositories for easier navigation and searchability.

## 2. Recommendations for Ad Repository API

A well-designed Ad Repository API is a crucial complement to the web repository. To make these APIs useful in practice, we recommend the following:

1. **Remove access barriers.**
    a. **Simplify registration**, and only require an email address if absolutely necessary. Guarantee free access for researchers. Do not include IP restrictions.
    b. **Establish fair use policies** to balance server load and user needs.
    c. **Implement clear rate limiting guidelines** to ensure fair access.
2. **Provide more comprehensive data**
    a. **Ensure data granularity** at least matches the web repository and vice-versa.
    b. Make sure that the repository **contains all the advertisements on the platform**.
    c. **Provide access to historical data**, with a suggested range of 10 years to facilitate long term trend analysis.
3. **Enhance search functionalities.**

    a. **Expand filtering options** and enable more targeted queries.
4. **Provide better documentation and user support.**
    a. **Offer documentation** in all EU languages.
    b. **Include code examples and libraries** with detailed explanations.
    c. **Provide clear versioning information** for the API.
    d. **Regularly update API features and documentation**. Maintain a status page for API availability updates.
5. **Increase harmonization**
    a. **Implement a RESTful API design** for ease of use and broad documentation availability. Prefer universally adopted standards over niche solutions like GraphSQL or BigSQL.
    b. **Standardize API endpoints** and data structure formats, in order to facilitate cross-platform research.

## b. Regulators

The European Commission may issue guidelines on "the structure, organization and functionalities" of the DSA ad repositories. It may also encourage standardization with a view to making them more interoperable. The new EU regulation on the transparency of political advertising may also refine the expectations for platform ad repositories.

1. **The European Commission and the Board for Digital Services Coordinators should develop guidelines on ad repositories, in consultation with the research community**, **in particular civil society researchers and platform integrity experts**. These guidelines should be the minimum standards for API and Web repository access, functionalities, data accessibility, and user interfaces. Specifically, this guidance should require sufficient documentation and support both for the web repository (non-technical user) and the API (researchers). Guidance should also clarify fair use (rate limits), impose user-centric interface design, and call for regular monitoring and compliance checks, as well as reporting systems and updates based on user feedback.

2. In addition, **the European Commission should improve the standardization of APIs across the designated VLOs to increase usability and facilitate cross-platform research** (see our proposal below).

3. Despite DSA Article 26(2), branded or paid influencer content remains difficult to study. Leaving this to users to self-declare does not address the problem of undisclosed influencer content. **We suggest augmented disclosure requirements for this kind of content that would require the platforms themselves to conduct cross checks and ensure the presence of this content in public ad repositories.**

# 9. Proposed Standardized Ads Repository Format:

The SDAF (Standardized Digital Advertising Format)[24] format integrates detailed advertiser information, creative content specifics, precise targeting data, and financial disclosures, offering a holistic view of digital ad campaigns. This initiative is proposed by CheckFirst as a first step to improve harmonization and interoperability of the repositories while maintaining a focus on accountability and clarity. This builds on the best practices of each platform and compiles them into a comprehensive format, laying a foundation for further work, for instance work in the framework of the Digital Services Act Article 44.1(d).

1. **Campaign Identification:**
    - **Field**: campaign_id
    - **Purpose**: Serves as the unique identifier for the entire advertising campaign. This is crucial for tracking and analyzing the campaign.
2. **Campaign Timing:**
    - **Fields**: start_date, end_date
    - **Purpose**: These fields specify the active period of the ad campaign, marking the beginning and end dates. Understanding the campaign's duration is important for analyzing its effectiveness and temporal reach.

3. **Advertiser Information**:
    - **Section**: advertiser
    - **Purpose**: Provides detailed information about the advertiser or the entity responsible for the ad. This section should include fields like advertiser_id, advertiser_name, advertiser_location, funder_legal_name, and funder_vat. These details offer transparency about the advertiser's identity and legal status, which is essential for regulatory compliance and audience trust.

4. **Creative Content:**
    - **Section**: creative
    - **Purpose**: This array should include detailed descriptions of each creative element used in the campaign. Each item in the array has fields like creative_id, creative_type, creative_title, creative_description, creative_media_urls, and creative_action. This section is vital for understanding the content and format of the ads, as well as the specific call-to-action intended by each creative.

5. **Targeting:**
    - **Section**: targeting

[24] Standardised Digital Advertising Format (SDAF) Project (CheckFirst, Feb 2024), https://github.com/CheckFirstHQ/DSA-Standardised-Digital-Advertising-Format

- **Purpose**: This section should provide a comprehensive breakdown of how the ad targets different audience segments. Organized by country and further into regions and demographic groups, it includes details like languages targeted, age groups, genders, estimated reach, and impressions. This granular view allows for an in-depth understanding of the ad's intended audience and targeting strategy.

6. **Payments:**
   - **Section**: payments
   - **Purpose**: Outlines the financial aspects of the campaign, including the actual and expected costs, as well as the currency used. This transparency in campaign financing is crucial for accountability and budget analysis.

7. **Moderation:**
   - **Section**: moderation
   - **Purpose**: Reflects the ad's compliance status with the platform's policies. It should include the moderation status and a history of moderation decisions. This section is important for understanding the ad's adherence to platform guidelines and regulatory requirements.

8. **Additional Parameters:**
   - **Section**: parameters
   - **Purpose**: Reserved for additional, platform-specific parameters like keywords or other targeting criteria. This flexible section allows for customization according to the unique requirements of different advertising platforms.

The SDAF format represents an initial step towards creating a standardized disclosure template for digital advertisements. It's crafted to initiate discussions about the need for such a standard, acknowledging that while it may not be perfect, it serves as a constructive starting point. This format is conceptualized as a 'Belgian compromise,' striking a balance between the requirements set forth by the Digital Services Act (DSA) for platforms and the informational needs of the research community. It's important to recognize that this proposal may not be universally accepted as is, but it offers a foundation for negotiations and further enhancements. The intention here is not to set this format in stone but to spark a dialogue on the necessity of standardization in ad disclosures, catering to diverse perspectives and requirements in the digital advertising ecosystem.

# Acknowledgements

We'd like to thank the many people who made this report possible, in particular our external reviewers Martin Degeling (SNV), Aleksandra Atanasova (Reset.tech).

We also acknowledge the work done in 2019 by Mozilla and many other researchers and experts.

# 10. Annex

## a. Ad Repositories links

| Platform | Ad Repository | Commercial Content | Ads Repository API |
|---|---|---|---|
| AliExpress | Link here | - | - |
| Amazon | - | - | Link here |
| Alphabet | Link here | Link here | Link here |
| Apple App Store | Link here | - | Link here |
| Bing | Link here | - | Link here |
| Booking.com | Link here | | Link here |
| LinkedIn | Link here | Link here | Link here |
| Meta | Link here | Link here | Link here |
| Pinterest | Link here | | Link here |
| Snapchat | Link here | Link here | - |
| TikTok | Link here | Link here | Link here |
| Zalando | Link here | - | Link here |
| X (Twitter) | Link here | - | Link here |