



July 9, 2021

Australian Competition & Consumer Commission
23 Marcus Clarke Street
Canberra, Australian Capital Territory, 2601
Australia

platforminquiry@accc.gov.au

Re: Digital Platform Services Inquiry

To Whom It May Concern:

Considering the Australian Competition and Consumer Commission’s forthcoming September 2021 interim report on the provision of web browsers and general search services to Australian consumers and the effectiveness of choice screens (which we refer to as *preference menus*), we write to share details about two recently published search engine market share reports from Cloudflare and Wikipedia.

As we have stated in past submissions to the Australian Competition and Consumer Commission (ACCC), confusion exists around search engine market share because common data sources vary significantly. The oft-cited source of comScore and, to a lesser but still significant extent, Statcounter have significant deficiencies.¹

By contrast, we believe that the recently released market share report from Cloudflare is more accurate both because of superior methodology and a significantly larger sample size. Therefore, we recommend the ACCC use this source for baseline assessments of search engine market dominance and for assessments of the efficacy of competition interventions like requiring a preference menu.

Wikipedia has also recently released a market share report with a methodology like Cloudflare’s.

¹ For more on the issues with Statcounter’s and comScore’s search engine market share reports, see *White Paper on the Search Engine Market*, DUCKDUCKGO (Mar. 2021), https://staticcdn.duckduckgo.com/press/DuckDuckGo-White-Paper-on-search_March-2021.pdf.

However, as Wikipedia is only one site, we recommend using Cloudflare's data as it incorporates data from millions of sites. Nevertheless, the Wikipedia report is more granular and can provide useful insights.

In this letter we outline these two reports, their methodologies, the reports' caveats and drawbacks, and how they compare to Statcounter's and comScore's reports.

Cloudflare's search engine market share report.

Cloudflare accelerates and protects traffic for large parts of the Internet with a global network that spans more than 100 countries. Approximately 25 million Internet properties rely on Cloudflare's network to reach their users.² Cloudflare services include content delivery network services, distributed denial of service (DDoS) mitigation, Internet security, and domain name system (DNS).

Cloudflare's services operate in the space between a website's visitor and the Cloudflare user's hosting provider, acting as a reverse proxy³ for websites. More than 10% of all websites connect through Cloudflare's reverse proxy, including 17% of the Fortune 1000 companies.⁴ Cloudflare recently published a search engine market share report⁵ based on the raw server logs that show referral data (i.e., the referrer header⁶) with bots removed⁷ from the report. Cloudflare's report is updated monthly. Whenever a user clicks a link from a given search engine's search engine results page (SERP), that search engine is identified by the referrer information in the header.⁸ Cloudflare can then count how

² *Cloudflare Radar: Search Engine Referral Report*, CLOUDFLARE, <https://radar.cloudflare.com/notebooks/search-engines> (last visited July 7th, 2021).

³ *What is a reverse proxy? | Proxy servers explained*, CLOUDFLARE, <https://www.cloudflare.com/learning/cdn/glossary/reverse-proxy/> (last visited July 7th, 2021).

⁴ *Search Engine Referral Report*, *supra* note 2.

⁵ *Cloudflare Radar: Search Engine Referral Report*, CLOUDFLARE, <https://radar.cloudflare.com/notebooks/search-engines> (last visited July 7th, 2021).

⁶ *HTTP referer*, WIKIPEDIA, https://en.wikipedia.org/wiki/HTTP_referer (last visited July 8th, 2021).

⁷ Bots generally result in spikes in traffic in market share reports. To learn more about how Cloudflare removes bots, see Alex Bocharov, *Cloudflare Bot Management: machine learning and more*, CLOUDFLARE BLOG (May 6, 2020), <https://blog.cloudflare.com/cloudflare-bot-management-machine-learning-and-more/>.

⁸ Inherent in how the Internet works and online traffic is routed, when a user visits any website, that site will know the user's immediately preceding website (this is called referrer information). For example, if



many clicked links are coming from search engines like DuckDuckGo, Google, or Bing. Cloudflare’s report is based solely on the referrer origin information and not referrer paths or query strings, which may reveal sensitive information. Cloudflare’s large network allows the company to measure search engine market share data more accurately. Cloudflare stores traffic metadata to Cloudflare sites (i.e., websites that proxy their request through Cloudflare’s platform), with every 1/100th request to Cloudflare sites being logged in their databases.

Report methodology.

The Cloudflare report looks for referrer data from the following general search domains: 360 (so.com), Baidu, Bing, Cốc Cốc, Daum, DuckDuckGo, Ecosia, Gmx, Google, Info.com, Lilo, Mail.ru, Mojeek, Naver, Qwant, Seznam, Sogou, Yahoo, and Yandex. The report separates out irrelevant subdomain referrer data—such as translate.google.com as opposed vs. google.com.

This methodology assumes that the average amount of clicks conducted in a search engine visit is the same across search engines, as Cloudflare is measuring link clicks and not searches. The report assumes that the percentage of the Internet that Cloudflare sees is representatives of the Internet as a whole in terms of link clicks from general search engines. The report also assumes that the referrer header remains intact.⁹

Market share data is split by OS, device (mobile vs. desktop), country, and month. The market share report includes the following operating systems: iOS, Android, Desktop (ChromeOS, MacOSX, Linux, and Windows), Mobile (Android, Blackberry, iOS, Kindle, Nintendo, Playstation, SmartTV, WebOS, Windows Phone, and Xbox). Anything that is not identifiable as being into one of those operating system categories is excluded.

Caveats and drawbacks of the Cloudflare market share report.

The report is based on a sample of Cloudflare’s referral data. Thus, the accuracy depends, in part, on the

a user searches for “durable umbrella” on Google.com and clicks on UmbrellaStore.com on the results page, then the UmbrellaStore will know that user came from Google.

⁹ In some browsers, it is possible for a user to disable the referrer field, but this seems very rare. In such cases, that field would come up empty.



amount of data available (i.e., having more customers that make use of their reverse proxy server). In some countries and operating system combinations, the data could be thin but still much greater than the available alternatives (e.g., Statcounter).

Why the Cloudflare report is superior to a report from Statcounter or comScore.

Cloudflare is seeing a much greater sample of the Internet than Statcounter or comScore.

Unlike the Statcounter market share report, Cloudflare's is not affected by advertising- or tracker-blocking tools or other browser extensions that block analytic requests from third parties. Cloudflare operates as a first-party and sees the direct requests to a website in their role as the reverse proxy server.

Unlike the comScore market share report, the Cloudflare report does not suffer from panel selection bias issues, and it includes mobile searches.¹⁰

Wikipedia's search engine market share report.

Wikipedia operates an open-source online encyclopedia. Wikipedia's site is used to populate knowledge panels located near the top of the SERP for many popular search engines, including DuckDuckGo, Google, and Bing. Wikipedia's recently published search engine market share report¹¹ is based on data showing how many Wikipedia page views were referred from common general search engines on any given day. In other words, the report is based on the HTTP referrer header data that Wikipedia sees in their own logs when users click a Wikipedia link on a SERP. In contrast to the Cloudflare report, Wikipedia's report is based on the total number of referrals and not a sample of the referral total. Like the Cloudflare report, Wikipedia's report excludes bots.¹² The Wikipedia search market share report is updated daily.

¹⁰ *White Paper on the Search Engine Market*, *supra* note 1.

¹¹ *Search Engine Referrals*, WIKIPEDIA, <https://wiki-search-referrals.wmcloud.org/#referrals/> (last visited July 7, 2021).

¹² *Analytics/Data Lake/Traffic/BotDetection*, WIKIPEDIA WIKITECH, https://wikitech.wikimedia.org/wiki/Analytics/Data_Lake/Traffic/BotDetection (last visited July 8th, 2021).

Report methodology.

Wikipedia’s report includes domains from the following general search engines: AOL, Ask, AU, Baidu, Bing, Cốc Cốc, Daum, Docomo, DuckDuckGo, Ecosia, Google, Lilo, MyWay, Naver, Qwant, Rakuten, Seznam, Startpage, Yahoo, and Yandex. The report is interactive and allows users to split market share data by country, browser, operating system, and language.¹³

To avoid exposing sensitive content, Wikipedia set the report threshold at 500 so that an attribute (search engine, country, language, operating system family, browser family, etc.) that did not meet the 500 threshold is represented by “other.” Wikipedia believes that this threshold allows the report to retain accurate counts of search engine referrals while also minimizing privacy risks.¹⁴

Caveats and drawbacks of the Wikipedia market share report.

The Wikipedia report is based on referrer traffic to one site and is therefore dependent on how much search engines promote Wikipedia. Generally, search engines promote Wikipedia frequently. However, DuckDuckGo believes Bing, for example, under-promotes Wikipedia relative to DuckDuckGo and therefore skews lower relative to our search engine. By contrast, the Cloudflare report does not have this issue as Cloudflare is aggregating data from millions of sites and sees traffic from a large percentage of the entire Internet.

Another bias in the Wikipedia report relates to the query mix of the user base. Because Wikipedia also skews towards English language sites, this report may be less helpful for measuring search engine market share in countries where English is not the dominant language.

Compared to the Cloudflare report, Wikipedia’s report is more granular, allowing users to see daily statistics. Users can also pull the data from the report.

¹³ See, e.g., [Search engine market share by country for June 2021](#), WIKIPEDIA, (last visited July 8th, 2021); [Search engine market share by country and OS family for June 2021](#), WIKIPEDIA (last visited July 8th, 2021); and [Search engine market share by country and browser family for June 2021](#), WIKIPEDIA, (last visited July 8th, 2021). Wikipedia (last visited July 8th, 2021).

¹⁴ [Analytics/Data Lake/Traffic/referrer daily](#), WIKIPEDIA WIKITECH, https://wikitech.wikimedia.org/wiki/Analytics/Data_Lake/Traffic/referrer_daily (last visited July 7th, 2021).



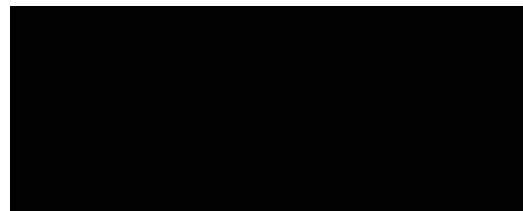
Why the Wikipedia report is superior to a report from Statcounter or comScore.

Unlike the Statcounter market share report, Wikipedia's is not affected by advertising- or tracker-blocking tools or other browser extensions that block analytic requests from third parties. Wikipedia operates as a first-party and sees the direct requests to its website from SERPs.

Unlike the comScore market share report, the Wikipedia report does not suffer from panel selection bias issues, and it includes mobile searches.¹⁵

DuckDuckGo thanks the ACCC for the opportunity to explain the benefits and drawbacks of two new general search engine market share reports. We are available to answer questions about these reports.

Sincerely,



¹⁵ *White Paper on the Search Engine Market*, *supra* note 1.