Reset.
AUSTRALIA

Response to Digital Platform Services Inquiry – March 2024 report on data brokers

Reset.Tech Australia

August 2023

Contents

About Reset.Tech Australia & this submission

Reset.Tech Australia is an Australian policy development and research organisation. We specialise in independent and original research into the social impacts of tech companies, including social media companies. We are the Australian affiliate of Reset.Tech, a global initiative working to counter digital harms and threats. We welcome the opportunity to respond to the ACCC's report into data brokers.

We are thankful to Wolfie Christl (Cracked Labs) who contributed research towards this submission. Mr Christl is one of the foremost experts on the personal data industry and data protection issues. His research has contributed to the filing of legal complaints against consumer data brokers and led to fines against unlawful data practices. Mr Christl's research has been taken up by policymakers across the world, quoted in reports to the European Commission and the US Federal Trade Commission, and helped stimulate global debates about systemic data misuse for commercial purposes.

We have prepared our response around the following specific questions:

- *Question 1: Who are the data brokers operating in Australia that predominantly collect information from other sources (i.e., not directly from consumers)?*
- *Question 17: What consumer harms may arise from the collection, processing, analysis or storage of information by data brokers? Which consumers are most likely to be harmed and why?*
- *Question 18: What consumer harms may arise from the use of data products and services sold or provided by data brokers? Which consumers are most likely to be harmed and why?*
- *Question 20: To what extent are consumers aware that their data is being collected and used by data brokers? How are they made aware?*

Research background and methodologies

The findings in this submission originate from a previously publicly accessible spreadsheet on Xandr's website. Xandr is a large adtech firm, formerly known as AppNexus. Xandr was acquired by Microsoft in 2022.[1] **The "Xandr file" contains more than 650,463 rows that describe so-called "segments" offered on Xandr's "data marketplace" that advertisers can use to target people online via the Xandr platform**. Each row has four columns ("Data Provider Name", "Data Provider ID", "Segment ID", "Segment Name". Most segments come from a number of third-party data providers. Many of them can be considered third-party data brokers. Often, data brokers that are mentioned in the file by "Data Provider Name" sell segments from yet other data brokers that are mentioned in the "Segment Name". As such, the file provides insight into data practices of a large number of data brokers globally, on the whole supply chain, and on what kind of data is being traded. We confined our analysis of the file to data products we believed were on Australians.

Key Points on 'Segments'

Segments are lists of IDs that refer to persons with certain characteristics, for example, lists of IDs that refer to persons by reference to gender, demographic, search history, or location history. Many different kinds of IDs are used across the data broking industry to track, follow and profile people, and to recognize them again when these persons use websites, apps and platforms across the digital world.

The most important IDs include Google/Apple "advertising IDs", browser cookie IDs and hashed/obfuscated versions of email addresses. Segments are the most important format of how personal information is being packaged, commodified and traded by thousands of companies today, similar to lists of postal addresses referring to households with certain characteristics that have been traded for decades. While lists of postal addresses or phone numbers have been used for direct mail or robocalls, segments - their digital equivalent – are being used to target people who visit a website or use an app; or to "personalise" websites, apps, services etc.

Xandr explains their data marketplace contains two types of segments: audience or behavioural segments,[2] and contextual or real time segments.[3] We have high confidence that the data providers in this report provide audience rather than contextual segments.

Limitation 1 - Data Practices May Be Historic

The file is dated May 2021. The file was available online until a week before publication of the articles by The Markup[4] and Netzpolitik[5] in June 2023. Xandr/Microsoft did not respond to the media inquiries from The Markup and Netzpolitik, but removed the file from its website. As such, the file may describe "segments" that are not in use anymore. Some responses from German and European data brokers suggest that at least some segments in the file were in use only until April 2019. Yet, other data brokers confirmed that the segments listed in the file are still in use as of 2023. In any case, the file describes data that has been offered via the Xandr data marketplace over the past five years.

[1] See: https://about.att.com/story/2022/xandr-microsoft.html

[2] https://docs.xandr.com/bundle/data-providers/page/audience-data-integrations.html [Accessed 13.8.2023]

[3] https://docs.xandr.com/bundle/data-providers/page/real-time-data-integration-instructions.html [Accessed 13.8.2023]

[4] https://themarkup.org/privacy/2023/06/08/from-heavy-purchasers-of-pregnancy-tests-to-the-depression-prone-we-found-650000-ways-advertisers-label-you

[5] https://netzpolitik.org/2023/surveillance-advertising-in-europe-the-adtech-industry-tracks-most-of-what-you-do-on-the-internet-this-file-shows-just-how-much/

Limitation 2 - Segment Names Are Subject to Interpretation

Key insights rest on information in the 'Segment Name' column. As such, it is subject to interpretation. Information provided by data brokers at other places can help to verify these interpretations. Where we have drawn upon secondary sources, we have indicated accordingly.

Limitation 3 - Size

It is unclear how many identifiers or people the Xandr file refers to. A segment list could contain anything from a few thousand to a few million identifiers. We encourage the ACCC to investigate where possible to clarify the scale of Australian data available within each segment.

Question 1: Data brokers operating in Australia that predominantly collect information from other sources

The Xandr file contains segment lists whose names contains a reference to "Australia" from several data brokers:

- **Oracle, Lotame, Nielsen, Factual, AlikeAudience, HYP**
- **DBM Atlas, Defind, Dynata, Equifax, Experian, GfK, Greater Data, List Factory, Mastercard, RDA Research, Roy Morgan, smrtr, The New Daily, YouGov** (*via Eyeota*)
- **Mobilewalla, Lifesight, Affinity Answers, eGentic** (*via Oracle*)
- **Roy Morgan** (*via Oracle and Mobilewalla*)
- **Mobilewalla** (*via Lotame*)

In addition, the Xandr file contains segments whose names contain a reference to "APAC" (**Lifesight, Selling Simplified, Skimlinks)** and "ANZ" (**Affinity Answers**), both via **Eyeota**.

According to our research, other notable third-party data brokers include:

Quantium
In 2018, Quantium claimed to exploit data on 2.5 billion transactions from 10 million Woolworths customers and 3 million NAB customers.[6] The company provides "over 1000" audience profiles.[7] In 2021, Quantium announced it had stopped working with NAB and was shifting to Commonwealth Bank of Australia[8] to "unlock the power of Australia's largest aggregated and de-identified transaction banking dataset".[9] The retail data on 10 million Australians appears to be still available.[10] From 2013, Woolworths had a 50% stake in Quantium, which it increased to 75% in 2021, making Quantium a majority-owned Woolworths subsidiary.[11] Quantium's practices have been featured in the Consumer Policy Research Centre's reporting.[12]

Audience 360
Audience360 claims to sell so-called "first-party data" from different data providers, but emphasizes its "granular audiences" can be used "across social & entire programmatic ecosystems".[13] Note, while they

[6] See, for instance, pp. 10 on this slide deck from 2018: https://www.worldooh.org/members/docs/congress/174/06.oOh_media_Brendon%20Cook%20FEPE%20Presentation%20June%202018_1.pdf [Accessed 8.8.2023]

[7] See, for example, Quantium's brochure from 2020: https://landing.quantium.com/hubfs/Brochures%20-%20Website/Quantium_Q.Audience_Brochure_AUS_V6.pdf [Accessed 8.8.2023]

[8] https://www.afr.com/companies/financial-services/cba-forms-data-joint-venture-with-quantium-20210507-p57ps3

[9] https://quantium.com/commbank-iq/ [Accessed 13.8.2023]

[10] https://quantium.com/q-checkout/ [Accessed 13.8.2023]

[11] https://www.zdnet.com/article/telstra-and-woolworths-quantium-to-form-new-data-and-ai-joint-venture/

[12] https://cprc.org.au/wp-content/uploads/2021/12/CPRC-Research-Report_A-Day-in-the-Life-of-Data_final-full-report.pdf

[13] https://www.audience360.com.au/our-audience/ [Accessed 8.8.2023]

claim the data to be “first party data”, they make it available across the digital world. On this distributional basis, we consider Audience360 a third-party data broker. They sell segments on:[14]

- **17 million devices a month** actively researching cars, bikes, boats and car finance (from “carsales”)
- **1.5 million devices** researching and buying flights and associated travel services (from “webjet.com.au”)
- **11 million devices** using AFL
- **640,000 devices** searching for rental properties (from “rent” and “realestate view”

DataRepublic
Data Republic is an Australian data analytics firm and data marketplace operator.[15] It is not entirely clear whether it can be considered a “data broker” or merely a data intermediary that organises and facilitates data selling and buying. DataRepublic had a long and close relationship with Qantas Loyalty (now also known as Red Planet[16]), NAB Ventures and Westpac’s ReInventure Fund, who became major investors and shareholders in Data Republic in 2016.[17] In 2018, ANZ also took a stake.[18] In 2021, the company went almost bankrupt and was acquired by IXUP,[19] another firm that organises data sharing between companies.[20]

In 2016, Data Republic helped Nine to exploit “in-store transactional data from Australia’s largest independent grocers’ loyalty program” to “improve its digital audience targeting capabilities”, including data on “five years of itemised basket data across 1600 stores nationally”, according to a trade press article.[21] The data came from the POS system vendor Worldsmart, which “covers a vast network of stores that nearly number as many as Coles and Woolworths combined”.[22] It is not clear whether the Worldsmart data came from an official loyalty programme, or was sold as payments data. Data Republic helped to create “aggregate grocery segments, which Nine will open up to advertisers in order to improve targeting across the Nine digital network”. A Nine representative said that by “being able to marry up users’ offline habits with our existing database, [Nine]will be able to allow advertisers to better target their campaigns”.[23] Note also, Roy Morgan has listed Data Republic as a “solution partner”.[24]

GourmetAds/HealthyAds
We have limited information on Australian data broker GourmetAds/HealthyAds,[25] though note it appears to provide sensitive[26] and super-sensitive[27] health data. Potentially, the most sensitive segments are not available in Australia, but this is not entirely clear. We would encourage further investigation into this data broker.

[14] Ibid.
[15] https://www.cio.com/article/207520/data-republic-launches-worlds-first-open-data-marketplace.html
[16] https://www.qantas.com/au/en/about-us/our-company/subsidiary-companies.html [Accessed 9.8.2023]
[17] https://www.zdnet.com/article/qantas-nab-and-westpac-behind-data-republics-au10-5m-funding-round/
[18] https://www.arnnet.com.au/article/633875/anz-takes-stake-data-republic/
[19] https://www.arnnet.com.au/article/688913/ixup-acquires-collapsed-data-republic-ip-3m, https://company-announcements.afr.com/asx/ixu/42ea5d1b-c725-11eb-af22-3acf87677c6f.pdf
[20] https://ixup.com/about/ [Accessed 9.8.2023]
[21] https://www.cmo.com.au/article/609836/nine-strikes-deal-data-republic-bolster-audience-targeting/
[22] https://www.adnews.com.au/news/nine-ramps-up-grocery-buyer-targeting-with-data-republic-deal
[23] https://www.cmo.com.au/article/609836/nine-strikes-deal-data-republic-bolster-audience-targeting/
[24] https://roymorgan-cms-dev.s3.ap-southeast-2.amazonaws.com/wp-content/uploads/2022/08/30042139/Helix-Personas-Booklet.pdf [Accessed 9.8.2023]
[25] https://www.gourmetads.com/articles/healthy-ads-acquired-by-gourmet-ads/ [Accessed 13.8.2023]
[26] https://www.healthyads.com/audience-segments/ [Accessed 13.8.2023]
[27] https://www.healthyads.com/targeting/, https://www.healthyads.com/targeting/cancer-targeting/ [Accessed 13.8.2023]

Question 17: What consumer harms may arise from the collection, processing, analysis or storage of information by data brokers? Which consumers are most likely to be harmed and why?

Question 18: What consumer harms may arise from the use of data products and services sold or provided by data brokers? Which consumers are most likely to be harmed and why?

We have combined our responses to these questions of **data collection** and **data use**. The data from the Xandr file reveals to us that companies are able to collect information on a seemingly limitless set of attributes. The file reveals that dozens of companies have a granular insight into the spending habits, financial circumstances, health information (including mental health), eating and exercise habits, and physical locations visited, for millions of Australians. While numerous companies use privacy-adjacent language to defend the practice,[28] this is partially undermined by the obvious attraction of segment data being the ability to link attributes to real people via identification numbers or devices in order to target them with advertisements in the future.

We observed numerous companies collect and trade datasets on Australians exhibiting signs of vulnerabilities, such as financial distress. Combined with other available data, such as age, gender, and family dynamics, we have grave concerns that third-party data brokers are able to identify society's most vulnerable and exploit those vulnerabilities. We have compiled a table of the most concerning segments in the Appendix. Where a segment is repeated across multiple providers, such as income, rent, mortgage repayments, we have only included it once. Our summary of especially concerning themes are:

- Lists on children
- Lists on teenage girls and teenage boys
- Lists on Indigenous Australians
- Lists on religious minorities
- Lists on unemployed people
- Lists on elderly people living alone
- Lists on people experiencing financial difficulties and distress
- Lists on people exhibiting gambling tendencies
- Lists on people experiencing pain
- Lists on new migrants
- Lists on people using browsers in other languages
- Lists on people deemed 'financially unsavvy'
- Lists on people with low education
- Lists of people who have visited certain medical facilities

[28] RDA Research, for example, claimed their dataset contained 'zero personal information'. However, this is something of a distraction, given that any attribute is intended to be linked back to a person via a device ID eventually. https://rdawebstatic.s3-ap-southeast-2.amazonaws.com/www.geotribes.com/factsheets/2023%20FACT%20SHEET%20Living%20Insights.pdf [Acccessed 8.8.2023]

There are two key features to how these datasets work that are relevant, which can be described non-technically as *identifiability* and *portability*. Buyers of these datasets are generally eager to use the segment data to sell ads back to the device from where the data originates. Brokers are also incentivised to make it possible to combine segments, so that multiple attributes can be combined to build a fuller profile on a person.

There is rich secondary literature on the consumer harms from the third-party data broking industry. For instance, Wolfie Christl's paper *'How Companies Use Data Against People*' makes a number of arguments for how commercial misuse of personal information can affect individuals, groups of people, and society at large.[29] Christl identifies two key zones where consumer harms arise from commercially-traded troves of personal data: automated decisions based on personal data, and data-driven persuasion. Frank Pasquale argues in *The Black Box Society*, a sweeping overview of the information markets over consumers' reputation, search, and finance, "regulators need to start forcing[data-collecting] firms to give consumers a sense of the sheer size of the data trove gathered about them—and its content."[30]

Based on the segments in the Xandr file, we suggest the presence of the following consumer harms arising from the Australian third-party data broking industry, noting this list is far from exhaustive. We have consulted the taxonomy of digital harms provided by London Economics in their 2023 report for the UK Department of Digital, Culture, Media and Sport.[31] We also note the three key consumer harms that typically flow from exploitative behaviour specifically: excessive data collection, excessive advertising, and excessive prices.

Consumer Harm	***Explanation***
Distorted consumer choices	Known also as 'dark patterns', the design and logic behind how firms present choices to consumers may lead to making decisions not in their best interest, including by creating a sense of urgency and disguising ads as editorial content.[32]
Algorithmic discrimination	Consumers may have prices offered to them for a product or service purchased online that has been calibrated in reference to various known or inferred demographic and behavioural factors. This dynamic creates a considerable information asymmetry which can unfairly prejudice the consumer.
Misinformed consent	Related to algorithmic discrimination, consumers may not be aware how their current or historic data is used in a transaction.
Algorithmic targeting	Combining algorithmic systems with targeting practices can lead to assigning consumers with 'group-like' attributes and targeting them with content that is inappropriate, irrelevant, or intrusive.[33]
Loss of control of personal data	The widespread use of consumers' personal data across a vast array of transactions generates legitimate concerns which can easily enter the domain of health and wellbeing harms. There is cause to consider cybersecurity risk, particularly given the use of location tracking on significant numbers of Australian mobile devices, including children.[34]

[29] Wolfie Christl, *How Companies Use Personal Data Against People*, 2017. Available: https://crackedlabs.org/en/data-against-people

[30] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, 2015, 145.

[31] London Economics, *Digital consumer harms - A taxonomy, root cause analysis and methodologies for measurement*, January 2023. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1129431/DCMS_consumer_harms_research_01-Jan-22.pdf

[32] We note and commend the Federal Trade Commission (US)' ongoing action against dark pattern tactics.

[33] Reset.Tech Australia is releasing research shortly on the Australian public's attitudes towards targeted ads. See also, our response to Question 20.

[34] One of the data brokers featured in the Appendix claimed in 2017 to make 6 billion "observations" from 15 million devices in Australia, i.e. they collected data from about 15 million Australian devices. A copy of the deck is available if required.

Question 20: To what extent are consumers aware that their data is being collected and used by data brokers? How are they made aware?

In July 2023, Reset.Tech Australia worked with YouGov to poll Australians on their views about targeted advertising.[35] While we did not ask Australians directly about data-broking practices, we noted that Australians routinely found targeted advertising to be an ***'intrusive'*** exercise. We found that around three quarters of respondents found targeted advertising very or somewhat intrusive, almost the same as the proportion who found targeted advertising not at all or only a little helpful. Further, 73 percent of respondents strongly agreed or agreed with the statement that they often receive targeted ads for things they found themselves "just thinking about".

We also asked respondents about their preferences around the ***collection*** of their personal data to drive advertising. Overwhelmingly, people wanted less data collection; 90 percent of respondents suggested that they would prefer less information collected about them online for advertising purposes. We also asked about the ***use*** of personal data to target advertising. Many of the current data uses were highly unpopular; 84 percent of respondents suggested that they would prefer that digital platforms stop targeting ads to them based on their online browsing history, which is the key source of data currently driving targeted advertising – and represented in the Xandr file. Further, 87 percent of respondents said they would prefer if platforms stop targeting ads to them based on sensitive personal information, about, for example, their political views, sexuality, or health, all of which is currently integrated into most targeted advertising mechanisms, and also present in the Xandr file.

[35] Polling data will be released shortly, in a public Reset.Tech Australia report.