

Sustainable without surveillance

ICCL review of sustainable publishing and tracking-based advertising

Foreword

Tracking-based online advertising imperils fundamental rights and publisher sustainability.

The tracking-based ad system causes a massive data breach (**page 4**) that undermines our fundamental rights to protection of personal data and privacy. Profiles built with these data may also threaten freedom of thought.

This report also reveals that publisher sustainability is imperilled by **four acute economic harms caused by tracking-based ads**:

First, leakage of publishers' audience data fuels Google & Facebook businesses at publishers' expense (**page 5**).

Second, tracking-based ads enable intermediary firms to charge large hidden fees, known in the industry as "adtech tax" (**page 8**).

Third, the tracking-based ad system leaks data about what everybody does online. So, while a publisher could normally sell their own audience's attention to advertisers exclusively, that same audience can now be found on other venues around the Internet. This "audience arbitrage" creates a business model for the bottom of the Web (**page 6**).

Fourth, tracking-based ads facilitate "adfraud bots" that divert billions of euro from legitimate publishers into the hands of criminals (**page 7**).

Together, these harms threaten the media pluralism guaranteed in Article 11 of the EU Charter. The massive data breach caused by tracking-based ads also allows voters to be profiled and then micro-targeted with disinformation (**page 6**).

But practical evidence from European publishers now shows that publishers' ad revenue can increase when tracking-based advertising is switched off (**page 3**).

A switch off across the entire market will amplify this effect, protecting fundamental rights and publisher sustainability. We urge lawmakers to play their part.

October 2021

Introduction

1. Foreword and table of contents
2. Key insights

Evidence

3. Removing tracking has increased publisher revenue

4. Tracking-based ads cause a massive data breach
5. Tracking-based ads siphon publishers' audience data to Google & Facebook
6. Tracking-based ads enable disinformation
7. € billions diverted from publishers in fraud

8. € billions diverted from publishers in "adtech tax"

Appendices

9. End notes and acknowledgements

Key insights

Tracking-based advertising diverts data and revenue from publishers, favouring Big Tech and the bottom of the web. It also enables massive fraud and micro-targeted disinformation.

- A Norwegian news publishing group earned an **average of 391% more for contextual ads than tracking based ads** over 12 months.
- Advertising **revenue increased by 149%** when Dutch publisher NPO Group replaced tracking-based ads with contextual-based ads.
- Tracking-based ads targeted using “Real-Time Bidding” (RTB) cause **the biggest data breach ever recorded**, jeopardising the right to privacy, right to protection of personal data, and right to freedom of thought.
- Audience data leakage prevents publishers from selling their own audience exclusively. And it exposes citizens to profiling and disinformation. **The “IAB Audience Taxonomy” standardises how people are micro-targeted (example: IAB code 600 denotes “Christian”).**
- Half of Google’s ad revenue once came from helping publishers show ads on publishers own properties. But **now nearly all (85%) of Google’s ad revenue comes from displaying ads on its own websites and apps, with the benefit of data siphoned from publishers’ websites & apps.**
- Fraud made possible by tracking-based ads **diverted an estimated €30.1B to €58.8B from legitimate publishers 2020.**

149%

Actual increase in publisher ad revenue when ad tracking removed

€30B - €59B

Cost of tracking-based ad fraud estimated in 2020

35% - 70%

Hidden “adtech tax” fees charged to publishers by tracking-based ad industry

Removing tracking has increased publisher revenue

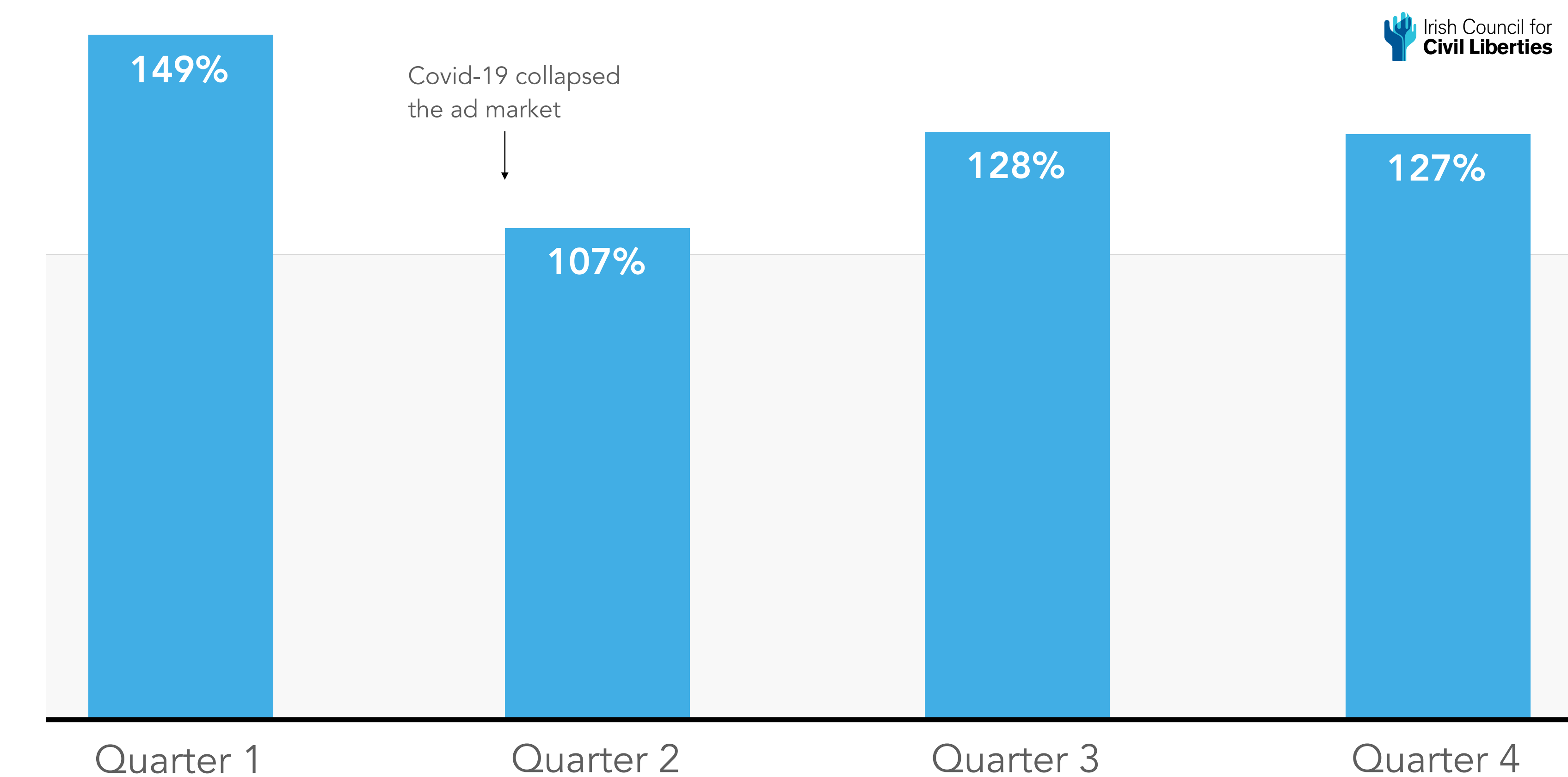
Practical evidence from European publishers shows that revenue increases when tracking is removed.

Evidence:

- **Actual advertising revenue increased by 149%** when Dutch publisher NPO Group replaced tracking-based ads with contextual-based ads. Revenue continued to increase dramatically **despite Covid-19's severe impact on the advertising spending.**
- Websites operated by a Norwegian news publishing group earned an **average of 391% more for contextual ads than tracking based ads** over 12 months.
- TV2, a major Norwegian news website, reports that ads sold through Kobler's contextual targeting return a **210% higher average price** than competing tracking-based ad targeting.

Revenue increase from removing tracking

Increase in publisher (NPO Group) revenue in 2020 using context-based ads, versus previous year using tracking-based ads.



Contextual v tracking-based ads

Quarterly average ad price (CPM) on websites in a Norwegian news publishing group, July 2020 - May 2021



Tracking-based ads cause a massive data breach

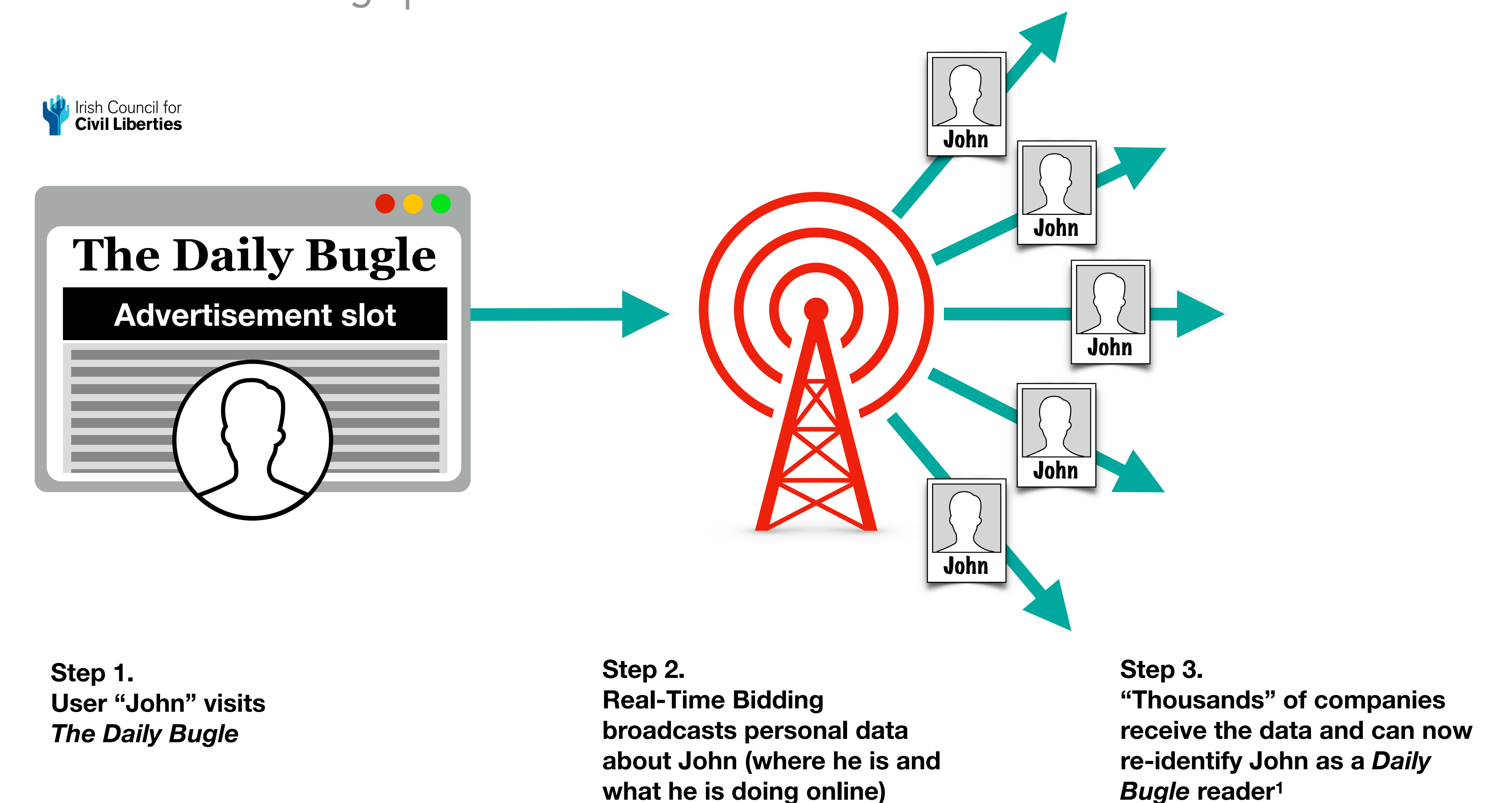
Tracking-based advertising turns legitimate publishers' audiences in to commodities that can be bought cheaply on low value & disinformation websites.

The bottom line:

- The dominant tracking based ads system is "Real Time Bidding" (RTB). It broadcasts Internet users' behaviour and real-world locations to many companies to run tracking-based ad auctions. Industry documents acknowledge that **"thousands" of companies can receive profile data about a person from a single ad.**¹
- This broadcasts **what everyone on the Internet is reading and watching, and where they are, many hundreds of billions of times a day.**
- This is **the biggest data breach ever recorded**, jeopardising the right to privacy, right to protection of personal data, and right to freedom of thought.

Broadcast of data by tracking-based ads

The "Real-Time Bidding" process



The biggest data breach ever recorded

Tracking-based ads (RTB) broadcasts by biggest "ad exchanges"

Index Exchange	120 billion daily broadcasts to unknown number of companies	Verizon Media	? (600 billion requests) daily broadcasts to unknown number of companies
OpenX	100 billion daily broadcasts to unknown number of companies	AT&T/Xandr	131 billion daily broadcasts to up to 1,647 companies
PubMatic	100 billion daily broadcasts to unknown number of companies	Google	? (Active on 9.8 million websites) daily broadcasts to up to 1,042 companies

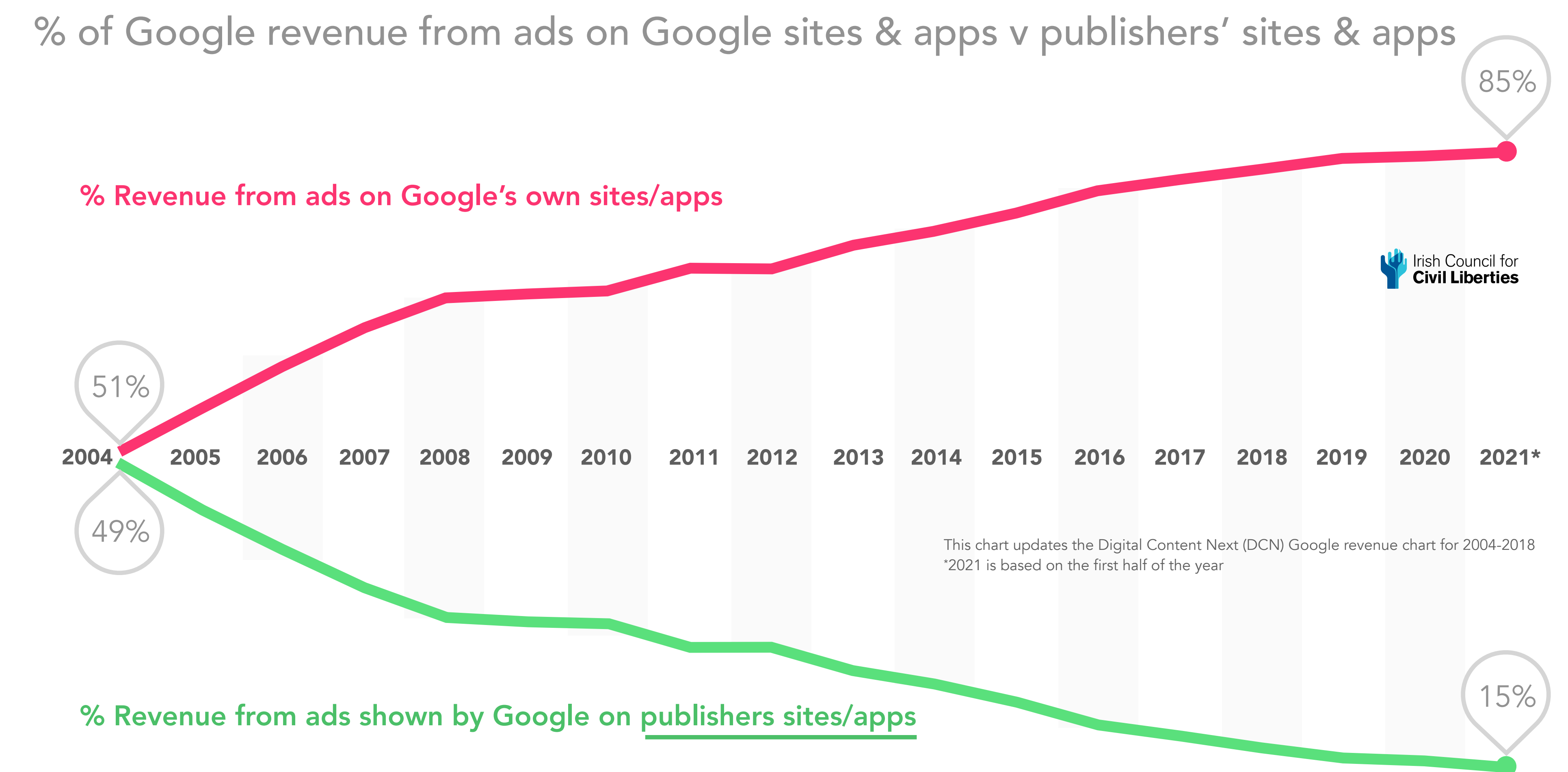
Tracking-based ads siphon publishers' audience data to Google & Facebook

Tracking-based advertising diverts publishers' audience data, and thus billions of Euro, to Google & Facebook.

Evidence:

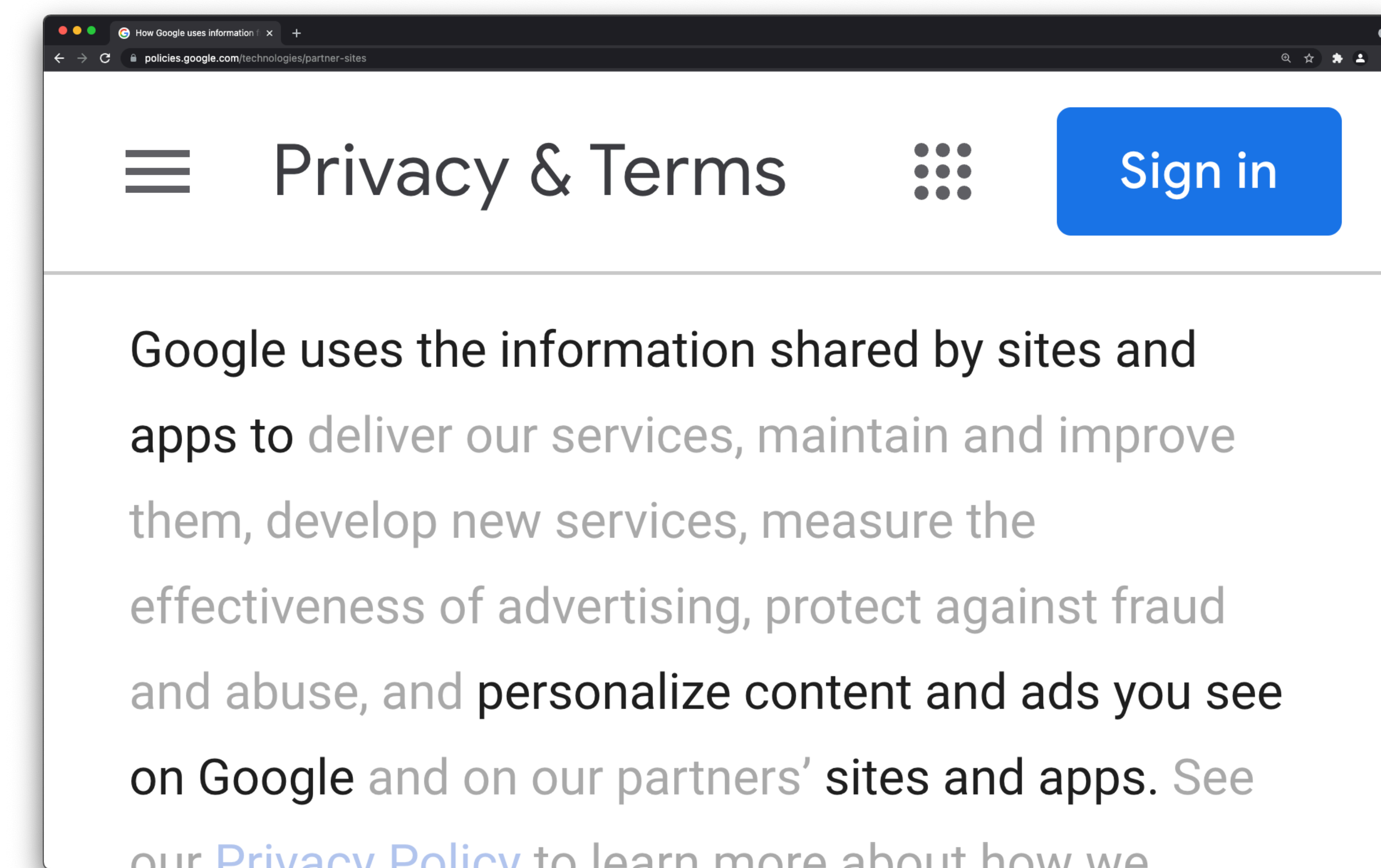
- Tracking-based advertising turns publishers' own commercialisable audiences into a Google/Facebook commodity. Google's terms say: "Google uses the information shared by sites and apps to ... personalise content and ads you see on Google ... sites and apps".²
- Google has diverted data and revenue from publishers to itself. In 2004, half (51%) of Google's ad revenue came from displaying ads on publishers' properties. Today, nearly all (85%) of Google's ad revenue comes from displaying ads on its own websites and apps, with the benefit of data taken from publishers' properties.

Google's ad revenue sources, 2004-2021



Google terms (emphasis added)

Google's terms provide for the exfiltration of data from publishers websites



Tracking-based ads enable disinformation

Tracking-based advertising allows legitimate publishers' audiences to be micro-profiled and micro-targeted cheaply on low value & disinformation websites.

Evidence:

- The IAB (the ad/tracking-industry standards body) **"IAB Audience Taxonomy"** standardises how people are micro-targeted.³
For example:
IAB code 193 denotes a person with "very low net worth"
IAB code 600 denotes "Christian"
IAB code 147 denotes "rural"
IAB code 199 denotes a conservative voter
- "Real-Time Bidding" tracking-based advertising lets hundreds of tech companies re-engage with publishers' high value audiences on low value websites. This arbitrage enables a **business model for the bottom of the Web**, and **deprives worthy publishers of the opportunity to exclusively sell their own audience's attention.**

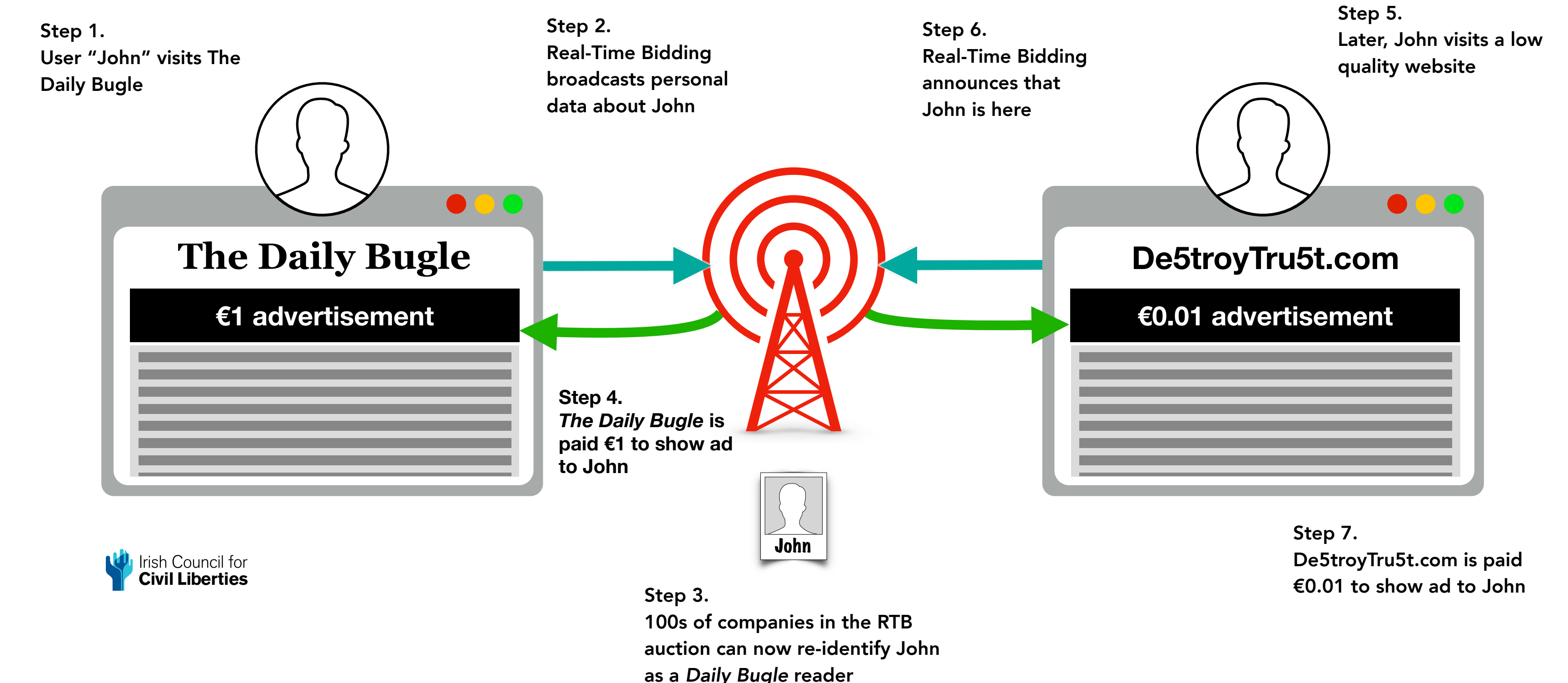
Tracking-based ads enable micro profiling

Screenshot from IAB "Audience Taxonomy", showing political profiling codes

Unique ID	Parent ID	Condensed Name (1st, 2nd, Last Tier)
193	191	Demographic Personal Finance Very Low Net Worth
194	191	Demographic Personal Finance Low Net Worth
195	191	Demographic Personal Finance Mid Net Worth
196	191	Demographic Personal Finance High Net Worth
197	191	Demographic Personal Finance Super High Net Worth
198	1	Demographic Politics
199	198	Demographic Politics Conservative
200	198	Demographic Politics Green
201	198	Demographic Politics Independant
202	198	Demographic Politics Liberal

How audience arbitrage enables the bottom of the Web

The "Real-Time Bidding" process across multiple websites



€ billions diverted from publishers in fraud

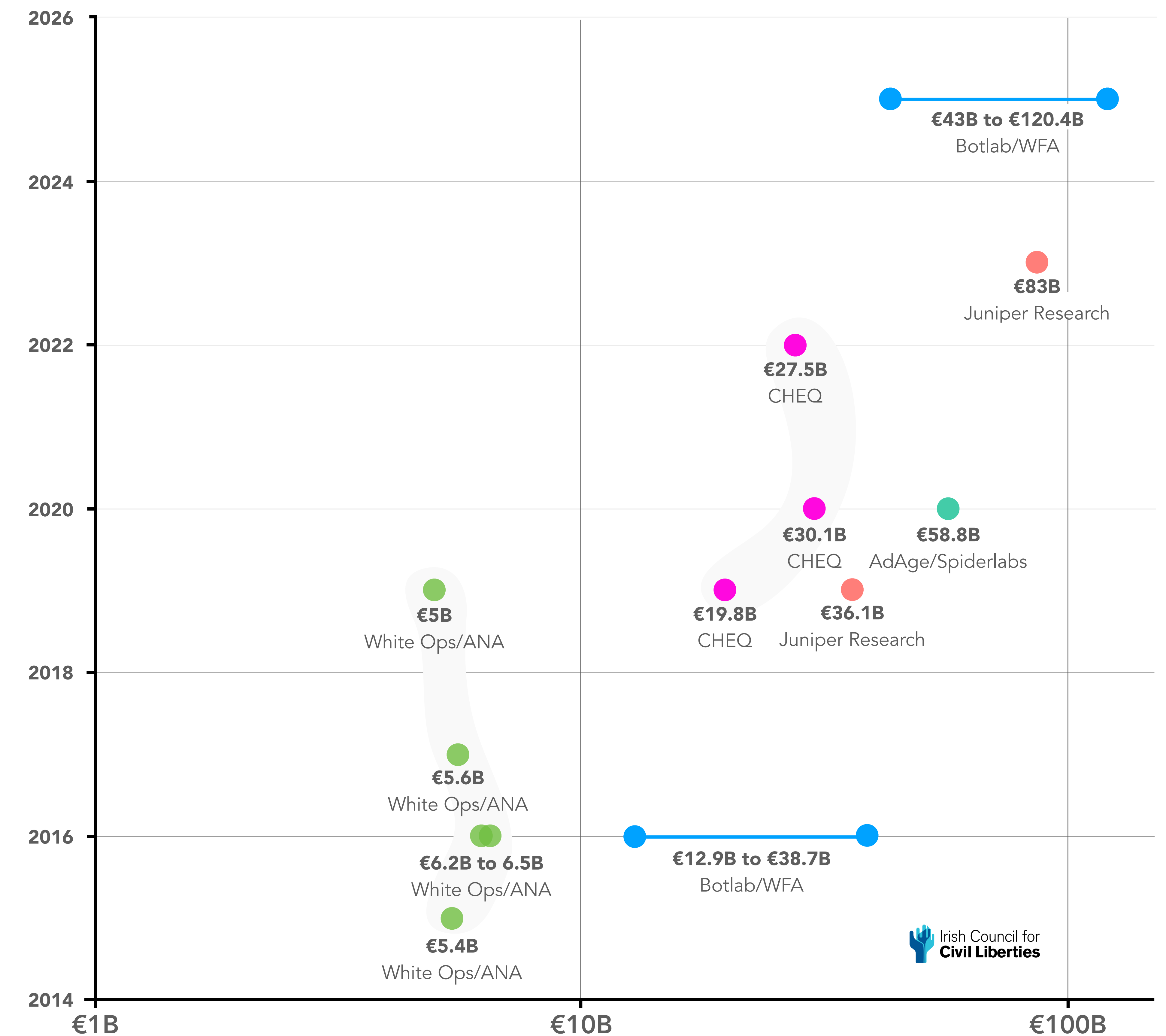
Tracking-based ads facilitate fraudulent “bots” that behave like humans on publishers’ websites & apps in order to draw tracking-based ad spending back to fraudulent websites.

Evidence:

- Estimates for 2020 indicate that tracking-based ad fraud **diverted €30.1B to €58.8B from legitimate publishers 2020**.
- The industry has **no authoritative estimate of the cost of ad fraud**.
- “Fake human” bots are very common online: **Facebook removed 5.6 billion fake accounts in the last 12 months - almost twice Facebook’s monthly active users (2.9 billion)**.

Multi billion euro tracking-based ad fraud

Estimates and forecasts of tracking-based ad fraud, in billions of euro



€ billions diverted from publishers in “adtech tax”

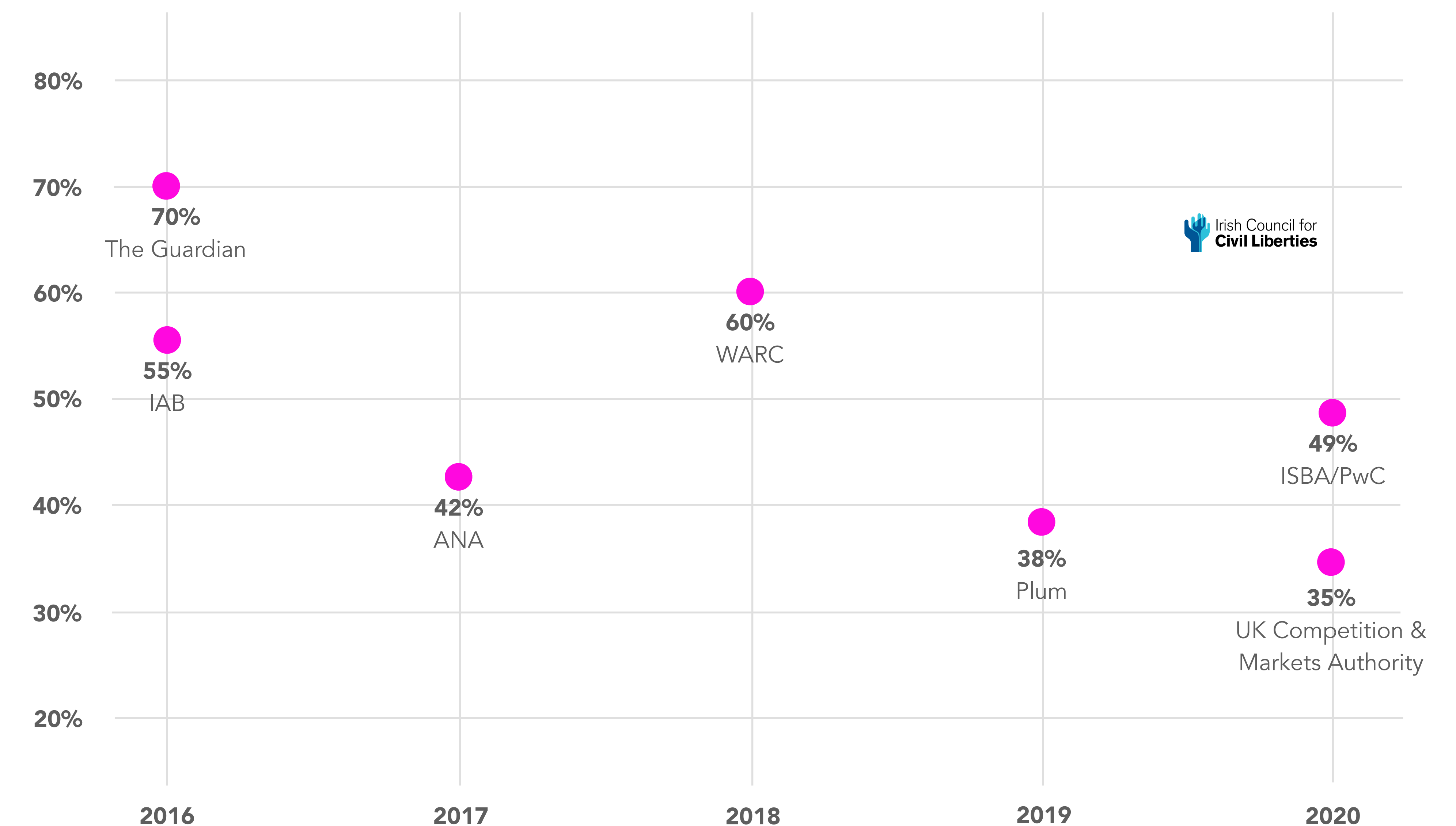
The tracking-based advertising industry is dominated by intermediary technology firms, not publishers. These “adtech” firms connect advertisers to publishers, and extract large hidden fees.

Evidence:

- In 2016, The Guardian newspaper ran a sting operation and used tracking-based ads to purchase ads on its own website. For every £1 that The Guardian spent as an advertiser, it received only 30p as a publisher. **70% was siphoned away by the adtech firms.**
- Estimates for **the cost of opaque fees charged on every tracking-based ad range between 35%-70%**. This threatens media pluralism (Article 11 of the EU Charter).

Estimates of opaque tracking based “adtech tax” %

% of money spent by an advertiser to show an ad on a publishers’ website/app that is not earned by the publisher, but goes instead to tracking-based ad tech companies



End notes and acknowledgements

Methodology:

1. The chart "Revenue increase from removing tracking" uses data provided by NPO, Ster, and Opt-Out, and compares advertising revenue quarter-over-quarter.
3. The chart "Contextual v tracking-based ads" uses data provided by Kobler, comparing the average clearing price for all auctions won by Kobler for a group of websites to the average price received by the same websites through Open RTB, on the same day.
4. The table "The biggest data breach ever recorded" uses the following data:
 - Google: Builtwith's statistics on Google DoubleClick/Authorized Buyers installations on websites <https://trends.builtwith.com/ads/DoubleClick.Net>. Does not include additional 4 million 'redirected' websites, so the figure may be considerably larger. Google publishes a list of data recipients at <https://support.google.com/admanager/answer/9012903>.
 - AT&T/Xandr: This is likely to be a low estimate. Source: the company's previous web page said in 2019 that it had "a peak of 11.4 billion daily impressions". Separately, a partner company revealed in 2017 that it took AppNexus roughly 11.5 RTB auctions (which entail unlawfully broadcasting personal data) to deliver a single advertisement (an "impression"). Therefore, the 11.4 daily impressions reported in 2018 equated to roughly 131 billion auctions per day. Xandr lists the companies it passes data to at <https://docs.xandr.com/bundle/service-policies/page/third-party-providers.html#ThirdPartyProviders-Ad-serverPartners>.
 - Index Exchange: 120 billion "auctions" on a "pretty typical day" at <https://www.indexexchange.com/ix-traffic-filter-meeting-2020s-business-challenges-with-machine/>.
 - PubMatic: figure from <https://pubmatic.com/blog/optimizing-data-processing-at-scale/>.
 - OpenX: figure from <https://cloud.google.com/customers/openx>.
 - Verizon Media: figure from <https://www.verizonmedia.com/insights/reach-your-roas-goals-with-verizon-media>.
5. The chart Google's ad revenue sources, 2004-2021 is based on Digital Content Next (DCN) 2004-2019 chart, and uses Google's 10-K filings with the SEC to split Google advertising revenue between revenue from Google owned properties and "Google Network" properties that are not owned by Google.
6. The chart "Multi billion euro tracking-based ad fraud" plots estimates from World Federation of Advertisers and botlab.io (<https://wfanet.org/knowledge/item/2016/06/03/Compendium-of-ad-fraud-knowledge-for-media-investors>), CHEQ/Roberto Cavazos (<https://info.cheq.ai/hubfs/Research/Economic-Cost-BAD-ACTORS-ON-THE-INTERNET-Ad-Fraud-2020.pdf>), Juniper Research (<https://www.juniperresearch.com/press/advertising-fraud-losses-to-reach-42-bn-2019>), Association of National Advertisers (<http://www.ana.net/getfile/21853> and <https://www.ana.net/getfile/23332> and <https://www.ana.net/getfile/25093>).
7. The chart "Estimates of opaque tracking based "adtech tax" %" plots estimates from The Guardian (<https://mediatel.co.uk/news/2016/10/04/where-did-the-money-go-guardian-buys-its-own-ad-inventory/>), the IAB (<https://www.iab.com/wp-content/uploads/2016/03/Programmatic-Value-Layers-March-2016-FINALv2.pdf>), the Association of National Advertisers (<https://www.ana.net/miccontent/show/id/44602>), WARC (https://www.warc.com/content/article/Global_Ad_Trends,_March_2018_Threats_to_digital_advertising/121186), Plum Consulting (<https://plumconsulting.co.uk/wpdm-package/jan-2019-online-advertising-in-the-uk-final-report-pdf/>), the UK Competition & Markets Authority (<https://www.gov.uk/cma-cases/online-platforms-and-digital-advertising-market-study#final-report>), and PwC and ISBA (<https://www.isba.org.uk/knowledge/executive-summary-programmatic-supply-chain-transparency-study>).
8. Facebook's bot account numbers are taken from Facebook's quarterly reporting (<https://transparency.fb.com/data/community-standards-enforcement/fake-accounts/facebook/>). Facebook's monthly active users are taken from Facebook's Q1 2021 financial results (https://s21.q4cdn.com/399680738/files/doc_news/Facebook-Reports-First-Quarter-2021-Results-2021.pdf).

End notes:

1. This is acknowledged in "Pubvendors.json", IAB TechLab (<https://github.com/InteractiveAdvertisingBureau/GDPR-Transparency-and-Consent-Framework/blob/master/pubvendors.json%20v1.0%20Draft%20for%20Public%20Comment.md>).
2. Google's Privacy & Terms (<https://policies.google.com/technologies/partner-sites>)
3. IAB Audience Taxonomy v1 and v1.1. See archived version of website (<https://web.archive.org/web/20201101045842/https://iabtechlab.com/standards/audience-taxonomy/>)

Irish Council for Civil Liberties:

ICCL has been at the forefront of every major rights advance in Irish society for over 40 years. We helped legalise homosexuality, divorce, and contraception. We drove police reform, defending suspects' rights during dark times. ICCL is a membership organisation and is independent of government. More at [ICCL.ie](https://iccl.ie).

Acknowledgements:

This report was written by Dr Johnny Ryan FRHistS, Senior Fellow of ICCL. Dr Ryan previously worked in the tech, online advertising, and publishing industries.

Alan Toner contributed to this report.

The chart "Google's ad revenue sources, 2004-2021" is based on, and updates, Digital Content Next (DCN)'s chart on Google's 2004-2019 ad revenue sources.

Cover photograph by Anthony Shkraba.