BETTER FUTURE SERIES

SWAP DATA PAIN FOR DATA GAIN

Creating controls against unlicensed data extraction

White Paper v1.0
May 2022

THE OZONE PROJECT

FOREWORD

First-party data: We're not even halfway through 2022 and already this looks set to be one of the year's most talked about topics in the world of marketing and media. As the sector looks for viable, compliant and, most importantly, scalable solutions for future digital growth, the subject is likely to stay forefront in the minds of marketers, agencies and publishers alike.

Since the publication of their 2019 report – 'The Dividends of Digital Marketing Maturity' – Boston Consulting Group has been tracking how leading marketers use data and technology to stay close to customers. It will likely come as no surprise that BCG has found that one of the biggest drivers of digital maturity – and of marketing results – is the use of first-party data.

Sophisticated marketers understand that first-party data is differentiated (because it's theirs and no-one elses), relevant (it directly relates to the company and its customers), and consistent and high quality (as it comes from the source) – this exact same thinking can be applied to the first-party data belonging to premium publishers. Advertisers believe first-party data is critical to better understanding consumer

behaviour, segments, and trends; to delivering more tailored and meaningful messages to customers; and to measuring effectiveness at multiple touchpoints along the customer journey. Indeed last year, BCG released further research that clearly demonstrated the impact of data-driven marketing and its ability to double revenue and increase cost savings by 1.6 times.

With such powerful data assets at their disposal, it comes as no surprise that the modern marketer is looking to partner with organisations with equally insightful and actionable user data. The imminent cull of the third-party cookie has accelerated the opportunity for those with their own rich data sets, while arguably accelerating the pain for the have-nots. As the gatekeeper of unique and incredibly powerful reader consumption data, this changing landscape puts premium publishers in pole position to help brands thrive in this new era of digital and programmatic advertising.

Which brings us to the purpose of this paper. There is a definite sense of being at another critical juncture when it comes to the publisher's role within the digital advertising ecosystem – and this time first-party data is the desired asset in question. The challenge is that, for the most part, the horse has already bolted and publishers are only now beginning to see that data extraction represents a significant hidden business cost, which is starting to hit the P&L.

Across the next few pages, we will investigate the scale of this challenge while also identifying strategic and operational approaches to ensuring premium publishers remain in control of their prime assets. And while at times the reading may be sobering, or even fear-inducing, we truly believe that this shift represents a massive opportunity for premium publishers like those Ozone represents to reengineer their programmatic business to work in their favour..

It's time to cast aside data pain, take control and focus on the data gain.

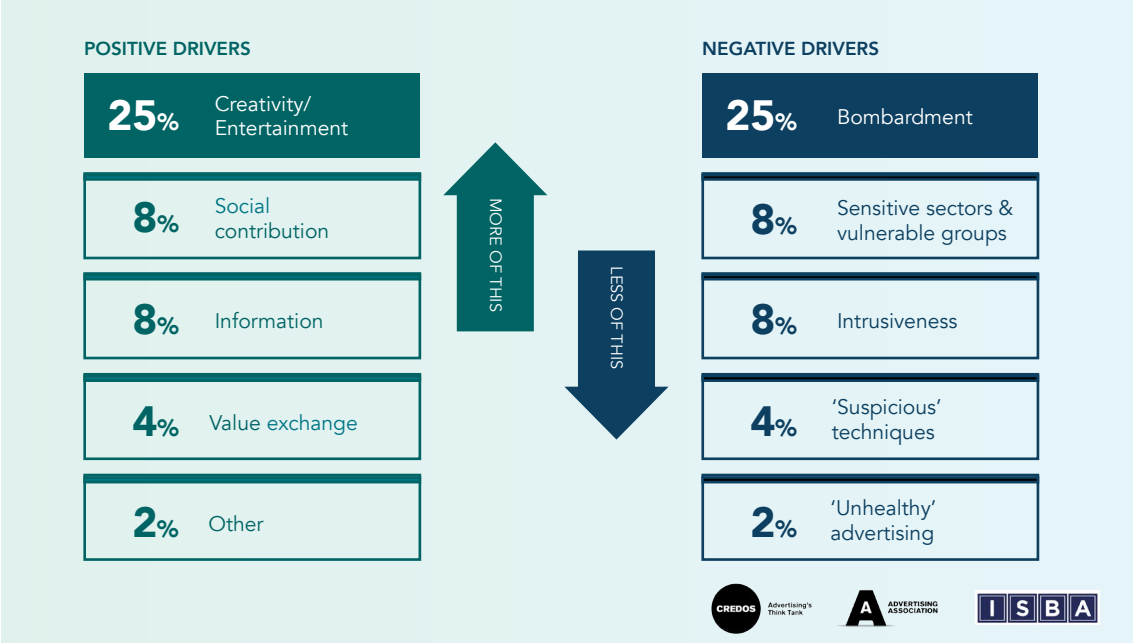
DANNY SPEARS
Chief Operating Officer
The Ozone Project

CONTENTS

1: THE IMPACT OF THE EVOLVING DATA-DRIVEN LANDSCAPE

From 2020 onwards, we have seen consumer data and identifiers used for advertising purposes thrust further into the spotlight than at any time in the past. This has manifested itself in many ways, whether that's as a result of regulatory questions raised by the likes of the ICO following their report into adtech and real time bidding, Apple and Mozilla's removal of third-party identifiers in their browsers (alongside Google's intention to do the same) or indeed through increased consumer awareness as a result of blanket news on the introduction of GDPR or the arrival of consent banners on their favourite, most visited websites.

Back in 2019, Credos – the Advertising Association think tank – highlighted growing consumer distrust in advertising fuelled in large part by practices seen most prominently within the digital advertising ecosystem. Creepy ads following them around the web, combined with the sheer frequency of message bombardment, meant the average consumer has become increasingly wary of advertising that tracks and identifies them, while also becoming more attuned to their data being collected and monetised with little – if any – reward for them.

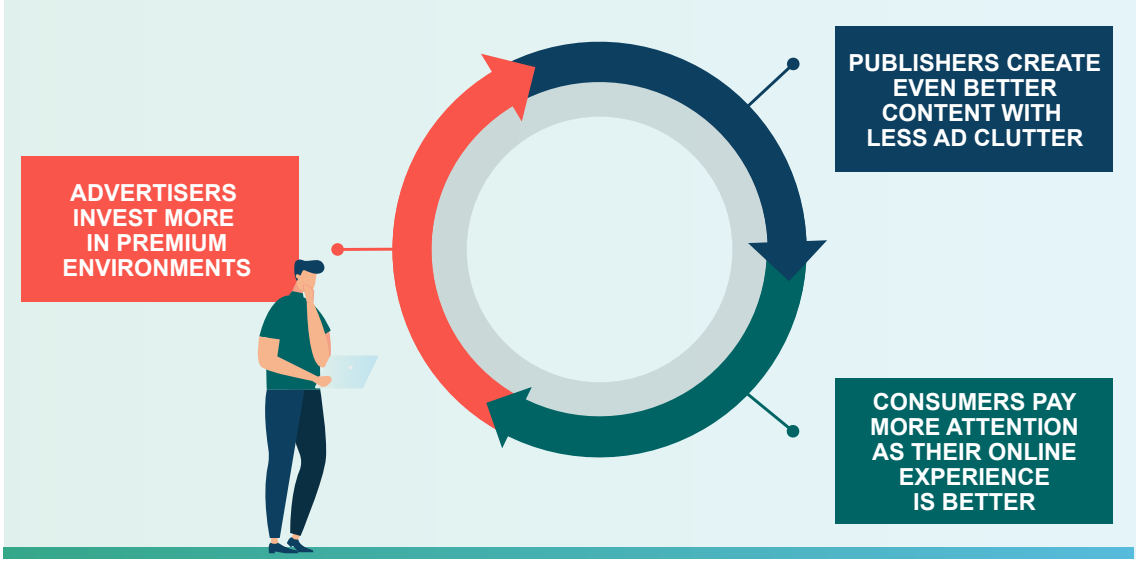


From an advertiser's perspective, the world of programmatic has also been seen to be guilty of failing brands. Phil Smith, director general of ISBA, called the current system 'broken' when their landmark study, conducted in conjunction with PwC and the AOP, highlighted only 51p in every £1 of advertiser spend reaching premium publishers – with a significant 15% disappearing into an 'unknown delta'. This lack of transparency at best frustrated, and at worst angered, advertisers and placed huge importance on actions to 'clean up' and avoid similar future scenarios.

Throughout this time, premium publishers have also faced significant challenges. Nowhere is this more acute than seeing digital assets being monetised at a rate that undervalues their true market value – largely thanks to being benchmarked against low quality, unregulated, long-tail content.

2: THE VALUE OF PUBLISHERS TO SOCIETY AND OUR SECTOR

Collectively, premium publishers spend billions of dollars funding journalistic content and investing in new ways of engaging their audience – the antithesis of the short-termist 'made for advertising' click-bait content movement. The 'reader first' approach to creation and curation employed by premium publishers creates incredible value for society at large; from keeping citizens informed, scrutinising governments or by shining a light on the instances where an individual, or business, is using its power to influence our democracy.



Editorial quality extends beyond the hard stuff too – premium publishers play a critical role in keeping people entertained and inspired. This has never been more acute than during the global pandemic, when publishers around the world saw record audience growth as populations turned to them for lockdown recipes, fitness ideas and suggestions for what to watch on the box. As we've emerged from this unprecedented period, premium publishers are once again inspiring travel plans, reviewing the hottest theatre tickets or rounding up the latest live sporting action.

It's this ability to engage and hold people's attention that creates unique value for advertisers. Research from industry voices such as the IAB, MediaCom, Magnetic and Newsworks has long demonstrated better advertising results are delivered as a result of ads appearing in high attention digital environments. In fact our own research, conducted with leading eye-tracking and measurement company Lumen, highlights that display advertising on premium publisher websites – like those Ozone represents – receives on average 51% more attention than equivalents on the rest of the web, a figure that more than doubles when comparing online video formats.

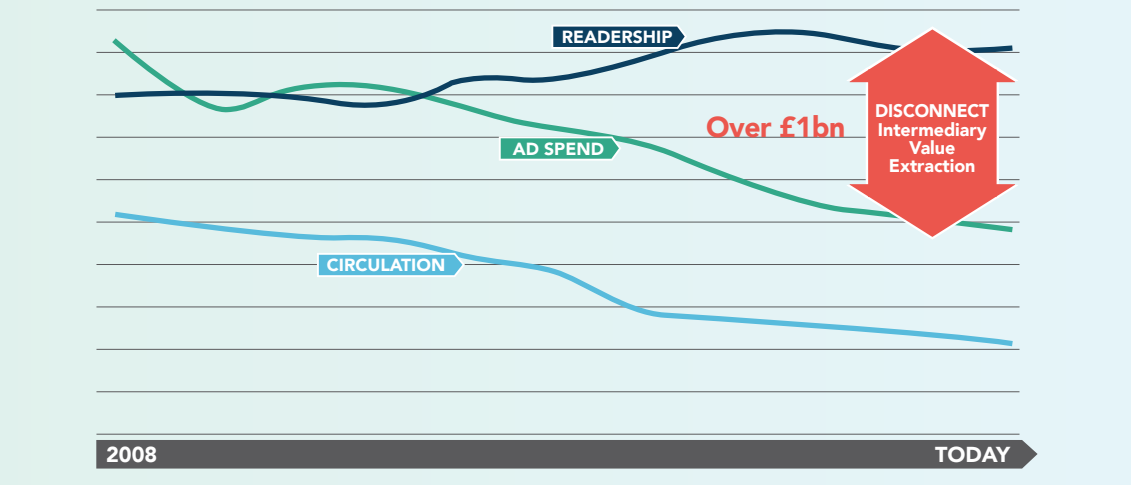
The attention paid to the content on premium publisher websites delivers incredible depth of data and insight that is a potent tool in shaping, activating and understanding the impact of advertisers' campaigns. And with all of this reader engagement able to drive better advertiser results, one might think premium publishers would be in an incredibly strong position within the advertising mix. Yet the stark reality is – that for the most part – they are far from it.

THE READER ATTENTION PAYBACK

In an aggregated cross-member body of work, The Attention Council interrogated the link between attention metrics and outcomes across fifty different cases. The study – which looked at the relationship between these measures and outcomes across the full funnel (such as recall and sales lift) – demonstrated significant correlation between advertising in attentive media and business outcomes.

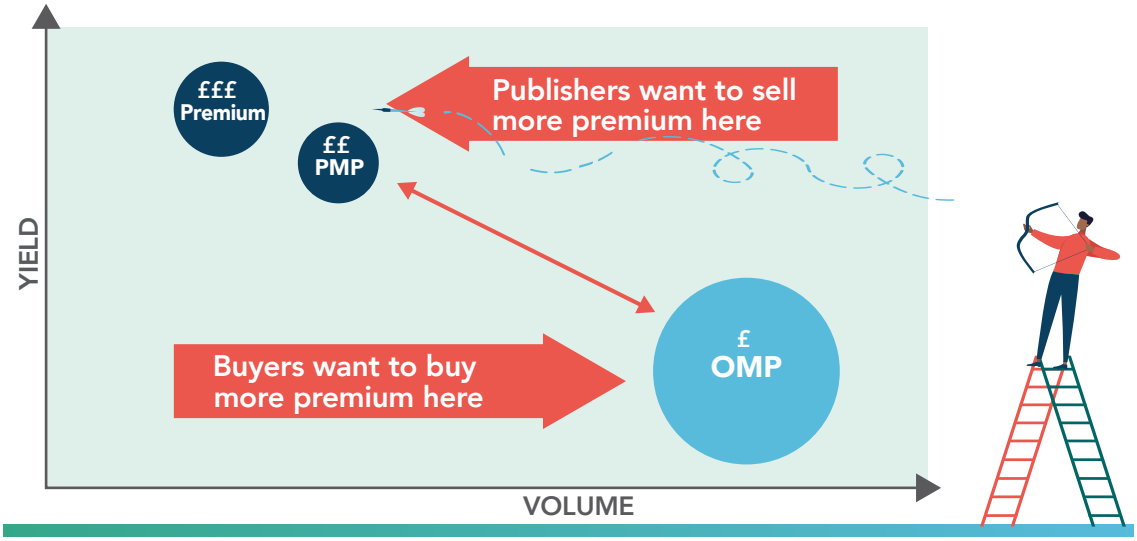
3: THE ECONOMIC DISCONNECT BETWEEN AUDIENCE AND AD INVESTMENT

The chart below highlights the disconnect between publisher audience growth and ad revenue received from brands. While digital eyeballs have helped boost and maintain market-wide readership levels, advertising investment making its way back to publishers has followed the downward decline of circulation revenues.



With this UK-led example, Ozone estimates this disconnect to be in the region of c.£1bn, a figure we believe to be driven by value – knowingly or unknowingly – being extracted by third-party intermediaries. Remembering the previously mentioned ISBA and PwC report into the programmatic supply chain, only 51p from every pound of advertising spend reached the end-publisher, with a further 15p totally unaccounted for.

Publishers are acutely aware of the challenges and conflicts they face as a result of their inventory being freely available through a number of different sales channels. Many have focused their attention on high yielding, direct-sold campaigns as a strategy for growth, with low yielding activity sold through the open marketplace giving additional, yet undervalued, income. Unfortunately for the publisher, what the buyer sees is an abundant opportunity to access audiences through discounted channels, completely diluting the premium publisher's core business.



Today, we stand at a crossroads where publishers are facing the very same dilemma with their data as they did with their inventory. It's time to act.

4: THE PUBLISHER IMPACT OF UNLICENSED DATA EXTRACTION

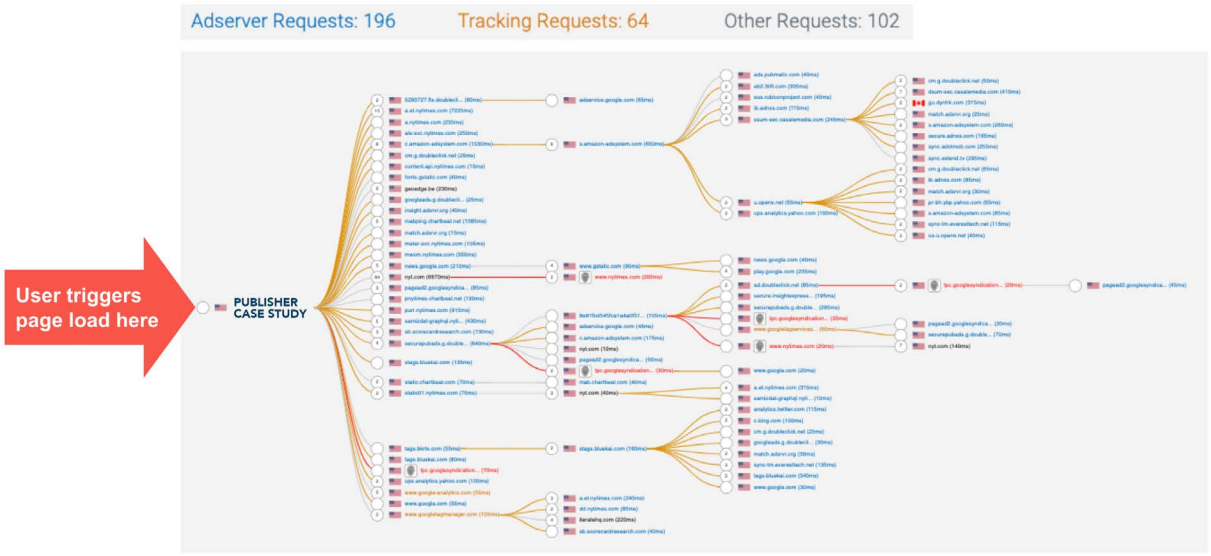
Publishers across the world are waking up to the need to protect their first-party data assets; taking control of them in a more focused way than they may have done in the past. The most significant challenge in this space is data extraction – often referred to as data leakage – an issue already impacting premium publishers in every global market.

Data extraction is the act of extracting and exporting data from publisher websites without the publisher's permission, and very often without their knowledge. Where this tends to play out is when third parties – for example, an ad exchange, DSP, SSP or data aggregator – takes this data from the publisher with no form of value-exchange.

This presents a significant issue for the future-facing, digital-focused publisher, as it devalues one of their most coveted assets, their first-party data, by making it available across the open web and sold through intermediaries. The net result – as we've already seen with inventory – is that this widespread availability directly sets a low-market price for publisher data, which in turn dilutes the same publisher's premium and direct-sold offering.

PAGEXRAY BY FOUANALYTICS OF US-BASED PREMIUM PUBLISHER

Actual data flow triggered by user visit



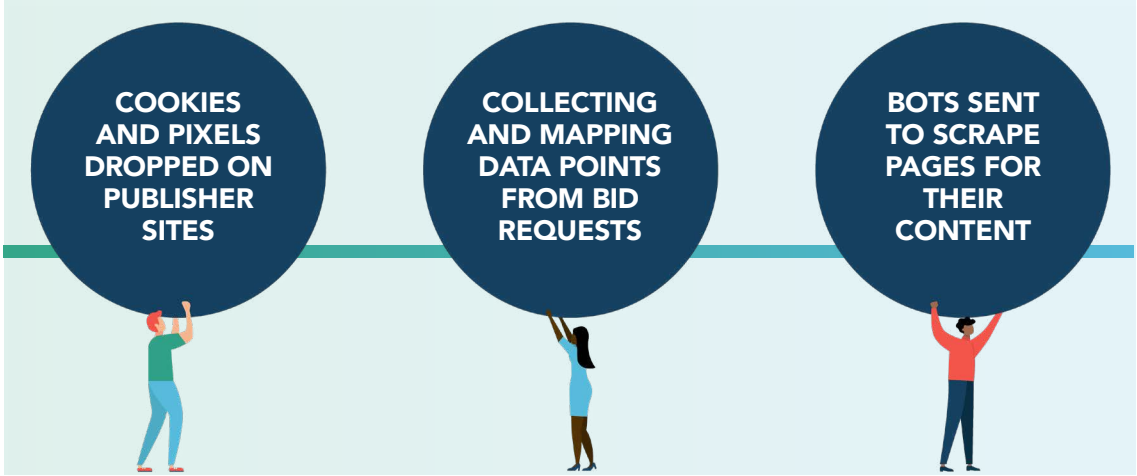
5: THE PROCESS OF DATA EXTRACTION WITHIN THE DIGITAL ECOSYSTEM

When investigating publisher data extraction, there are three primary areas of concern. The first concerns the use of **COOKIES & PIXELS** that advertisers and technologies drop on publisher sites when bidding for inventory. These pixels collect user data and transmit these profiles back to intermediary-owned servers. In this situation, the publisher receives no remuneration for the data received, which is then available for potential (mis)use by those extracting it.

Secondly, extraction also occurs through the collection of data-points from **BID REQUESTS** that come from a publisher and are subsequently mapped to an intermediary user ID for use in targeting. This means an ad partner connected via prebid can profile all of the users on the publisher website without having to bid on a single ad impression –

once again, the third-party is able to gain huge amounts of publisher data, without having to pay for the data, or even inventory.

Finally, and arguably the most blatant method, involves organisations sending **BOTS TO SCRAPE** publisher web pages for their rich, longer-form text content. The data extracted is then categorised for third-party contextual targeting solutions to help brands reach broader audiences – in this situation, the third-party may also map this contextual data back to a user ID in order to build user profiles for targeting across other websites. Again, the publisher receives no recompense for the value received from this data, value that has a significant impact on the business models of these third-parties.



6: THE TYPES OF DATA BEING EXTRACTED FROM PUBLISHERS

When a reader clicks 'Accept' on a website's consent banner, they are making a trade-off that gives them access to a raft of quality content in exchange for sharing their behavioural and reading behaviour with the publisher and their advertising partners.

While questions may arise as to whether consumers fully understand what they're agreeing to when they press 'Accept', it is a fair assumption that they would at the very least expect the publisher site they are visiting to be in control of their data and ensuring it is used properly. However, a report published in May 2022 by the Irish Council for Civil Liberties noted that in Europe, the real time bidding process exposes people's data 376 times a day, a figure that almost doubles in the US.

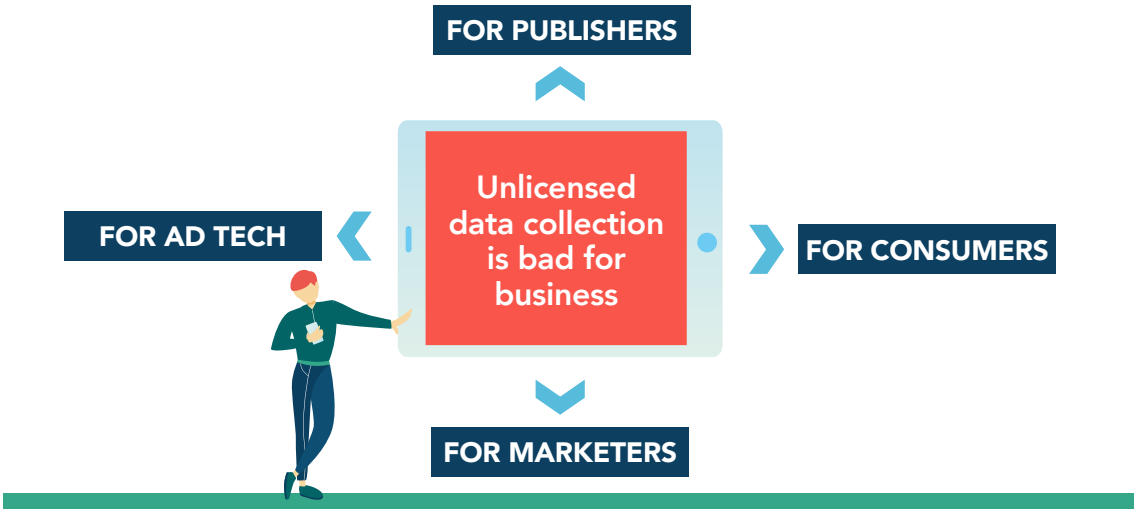
The vast amount of publisher controlled data that is being unknowingly extracted or indeed extracted without permission, falls into four broad categories:

CONSUMER IDENTIFIERS	CONSUMER PROFILE INFORMATION	BUSINESS INTELLIGENCE	PUBLISHER IP
Most likely to be user IDs that can be used to better understand and more accurately target consumers	Further depth about user sessions, specifically relating to technology, device, browser and location data	Advertising transaction data showing different advertisers and their bid prices for publisher inventory	Publisher content metadata that fuels contextual and semantic capabilities

Once extracted from the publisher, this information is under the control of intermediaries. When the publisher's data reaches this point, there is a greater risk of it being used for unlicensed commercial gain, or even worse, being offered as a competitive, lower-cost targeting alternative to working with publishers directly.

7: THE FULL IMPACT OF DATA EXTRACTION

Put bluntly, the unlicensed collection and exploitation of premium publisher data is bad for business. Not only will it further exacerbate the disparity between digital audience growth and advertising investment, but will also serve to set a lower market-rate for publisher media, as a result of intermediaries offering a lower-yielding alternative to marketers. However the ramifications of this data leakage are not solely restricted to the publisher realm, with wide-reaching consequences across the entire digital advertising ecosystem.



Firstly, for **CONSUMERS** who are already wary of digital advertising thanks to years of message bombardment and being followed around the web by creepy ads. These consumers are the same people publishers spend a huge amount of effort building relationships with, in order to make their website the readers' url of choice. This relationship is hugely valued by publishers and is often – quite rightly – prioritised over commercial opportunity. The danger from data extraction is that most consumers are unaware of the extent to which their profile information, consumption patterns and interest data are being leaked from the publisher. While the consumer has an agreement with the publisher, they are unlikely to know that their data is being collected and exploited by companies they don't know or trust. Should this become known, the consumer is more likely to blame the publisher for any breach, than an unknown ad tech company.

Secondly, there is a significant implication for **BRANDS**. Purposeful marketing, ESG initiatives and CSR programmes have become a huge part of a brand's armoury, a move that has seen millions of dollars invested in delivering positive contributions to society. Most of these brands are completely unaware that their digital advertising supply-chain may have become compromised by third-parties who have sold them unlicensed, and even unknowingly captured, intellectual property belonging to a publisher. Even if a brand looked beyond the implications on publisher pockets, few marketers would be able to do the same when considering the privacy implications for existing and future customers.

And finally, there's the impact on **ADTECH** itself, a sector still wrestling with the transparency challenges outlined in the ISBA and PwC supply chain report:

i	The data extraction issue is exacerbated by some DSPs and SSPs aiding and abetting by acting as a storefront for unlicensed publisher IP
ii	There is an inherent and undisclosed risk contained in some adtech business models that drives unsustainable valuations and exposes investors and shareholders to potential loss
iii	Lastly, there is an additional compliance risk due to some analytics and verification intermediaries providing an ID to buyers, without any form of consent as a means to build user profiles

For an industry challenged with cleaning up its act, not one of these actions present a positive step forward.

8: THE WAY FORWARD FOR PREMIUM PUBLISHERS

For publishers, the challenges and implications of data extraction are explicit and clear, so what measures should be put in place to safeguard their future?

STEP 1: **MAKE CONTROL A STRATEGIC IMPERATIVE**

The most significant impact to this challenge will require behavioural change and breaking free from the shackles that fuel the status quo – in short this requires a significant reset in how programmatic works within each publisher's business. Future thinking should be organised around the very simple operating principle of the publisher being in unilateral control of their supply chain, meaning intermediaries and third-parties will not be allowed to continue to dictate terms of trade.

This root and branch approach to future capitalisation of data assets would focus on restructuring the publisher's programmatic business – an approach that would require Board-level commitment and investment. In order to succeed, publishers would remove all adtech code from their pages and turn off open market programmatic, meaning inventory and data assets would only be available to buyers directly from publishers or accredited partners. This is a strategic move that represents short-term revenue risk, but also the quickest route to programmatic transformation.

STEP 2: **INVEST IN PUBLISHER TECHNOLOGY**

If the strategic imperative is to take back control, then publishers should adopt trusted publisher technology that has been specifically designed to let them do that. It should also be acknowledged that this technology is a vehicle for change; it is not an alternative to the strategy itself. As a business built for publishers by publishers themselves, Ozone has focused technology investment in building products that allow publishers to fully-manage and take control of their data – and indeed inventory – assets.

STEP 3 : **ACTIVATE PRACTICAL MEASURES**

There are a range of tactics that premium publishers can deploy to regain an upper hand in data conversations with partners before putting in place longer term protections. A number of potential steps are listed on the following page:

MINIMISE DATA EXTRACTION FROM THIRD-PARTY SCRIPT ON YOUR PAGES	Audit and remove unnecessary tags from your pages. Watch out for analytics partners who also operate in the media business (i.e selling your data).
MINIMISE DATA EXTRACTION FROM YOUR BIDSTREAM	Remove data-points from bid requests that don't return value (revenue) while retaining those that do. This varies by bid request and by ad partner.
MINIMISE DATA EXTRACTION PERMISSIONED VIA TCF (INCLUDING VENDOR PURPOSES)	Audit your provision of consent and purposes to vendors in your TCF vendor list. Remove permissions for all parties you don't have a direct relationship with. This should be taken as a message of your permission for any form of data collection to have been withdrawn; should a party continue irrespective they are in contravention of GDPR regulation and face legal risk.
BLOCK UNSOLICITED DATA COLLECTION VIA ADS	Reverse-proxy ad slots to strip data collection pixels from ads. Note that this will also block verification vendors' ability to classify your pages for brand safety. This means the page will be registered 'unclassified' and default to your domain-level score. As such, this may impact revenue.
BLOCK REMOTE EXTRACTION OF PUBLISHER IP AND METADATA VIA PAGE SCRAPING	Technical measures include blocking user-agents (crawlers) from accessing your pages. This can be supported by Commercial / Legal / Policy with communication to non-permissioned vendors. This could go as far as cease & desist, or lead to legal pursuit of those who are lifting your IP without licence.
REMOVE YOUR SUPPORT FOR UNLICENSED DATA ACTIVATION	Opt-out of intermediaries' multi-publisher packages - these act as an activation channel forunlicensed data. In addition, they often deliver unclear value in terms of revenue and as such are likely to have little to no business impact.
SET STANDARDS AND GET VOCAL	Be clear with your expectations of ad partners. Understand their policy for managing the sale of unlicensed data and the measures they have in place to ensure data that might be taken from your site without permission isn't then made available via their own platform.

OZONE'S SOLUTIONS TO PROTECT AGAINST DATA EXTRACTION

Rewind back to 2018 and the formation of The Ozone Project by our founder publishers. At the heart of our creation was the idea of creating significant upside for advertisers and premium publishers as the custodians of the consumer relationship within digital advertising.

For advertisers this meant creating a really easy way to buy audiences across premium, trusted, brand safe environments at scale, in a way that delivered premium campaign outcomes – offering a true alternative to the major platforms and a significant upgrade to buying through open market programmatic.

For our publisher partners, the mission was grounded in creating a more sustainable future for quality digital content, and ensuring the value of premium publisher assets - both inventory and data - was maximised. The net result would be greater ad investment reaching publishers themselves, rather than being extracted by third-party intermediaries.

Four years on, we are as focused on this mission as ever. We believe our commercial standards set a healthy bar for other publisher partners, in no small part down to the fact our publisher-built business has the best interests of quality content creators in our DNA.

Ozone's approach to publisher data

Working in publishers' best interest is reflected in our own approach to working with their data. Any audience data Ozone collects from our publishers is with the explicit consent of both that publisher and the user, with all contextual data collected with the explicit consent of the publisher. This is fully enshrined in contract, and our model ensures that our publishers are remunerated for their contribution to campaign delivery as a result of their data, in the same way they are remunerated for inventory use.

Contrast this approach to the wider market. We are very concerned that the publisher data ecosystem operates without standards, leaving publishers – and indeed Ozone – competing for advertiser investment alongside other businesses whose product is built on unauthorised or unlicensed publisher data. The challenge this raises is only set to grow, particularly with contextual data being hailed as one of the saviours post the third-party cookie demise. This should put premium publishers in a strong position to capitalise, but that is endangered if others are using their data without payment.

Putting publisher technology into action

Ozone's focus on creating our proprietary publisher technology has been driven by the desire to give publishers greater control over their 'last mile'. In practice this involves creating a control layer between themselves and adtech, a move that in turn allows the removal of adtech code from the publisher's pages.

An example of this in practice is our virtual, publisher-centric Ozone ID that has increased addressable audience (and monetisation potential) from c.50% to c.85%. Another is our proprietary 'smart bidstream' technology that gives publishers greater control over the data they make available to ad partners - reducing data extraction - while at the same time creating significant uplift through optimisation.

As mentioned earlier in this paper, this publisher technology is a means to execute a programmatic strategy shift, and has already been proven to deliver results for our partners. For example, by controlling data distribution in the bidstream and by optimising presentation of bid requests, our publishers have seen:

- 60-70% reduction in the volume of user profiles being transmitted to ad partners
- c.80% decrease in no-bid auctions
- +70% increase in addressability

The result of these actions to create value in scarcity has driven a positive response from buyers through an increase in bid frequency and bid valuation. To date, this has generated increases in total publisher digital revenues in the region of 35-40%, with publisher yields increasing by approximately one-third.