The Measure Pod: #42 Consent Mode and behavioral modelling in GA4

[00:00:00] **Dara:** Hello, and welcome to The Measure Pod episode number 42. This is a place for people in the analytics world to talk about all things data and analytics, especially Google Analytics. I'm Dara MD at Measurelab.

[00:00:28] **Daniel:** And I'm Dan, I'm an analytics consultant and trainer also at Measurelab. So Dara, we talked about it a bit last week when we were chatting with Ed, but actually what we are going to be talking about today is the most recent or biggest update to Google Analytics 4 probably thus far. And this is the rollout of behavioural modelling using consent mode. And we gave a little sneak peek last week, what it was about but we need to give it a bit more time and attention to actually explore what this feature is, how it works and more importantly what it means for us and everyone else using GA4.

[00:00:57] **Dara:** Yeah definitely a big piece of news and this is us following up on our promise. We committed to deep diving this, so this is exactly what we're going to do this week. We're going to talk about what it means and what we think of it and just see where we go with it really. So let's start at the top, let's just recap what the update is.

[00:01:15] **Daniel:** Yeah, for sure and just as a heads up to anyone listening, this is going to be quite a link heavy episode, and we're going to be putting a bunch of links in the show notes for a load of resources, guides, tips, and so pointing you in the right place. So in your app if you have a look in the show notes there, or just click the link to our website and all of the links and the details will be alongside the transcript there.

[00:01:34] **Daniel:** So the news Dara, so this was announced on the 2nd of June. So, as of the recording, it's about two weeks ago, but what they introduced is what they call behaviour modelling for consent mode. What this means is it's in a sense, one of the final pieces of the machine learning puzzle that they've been promising for GA4 for some time. So a couple of the other models they're already using. So data-driven attribution (DDA), their predictive modelling to do things like anomaly detections and the predictive conversions and churners, audiences that they can do for the ecom data. They're doing the conversion modelling, so that's something that they introduced a couple of months back, but finally, what they're looking at now is behaviour modelling.

[00:02:09] **Daniel:** So what this affects is all data basically so it's not a small rollout, what they've done. It uses a feature called consent mode, which we'll talk about in a moment, but basically what it does is it understands that there's going to be missing data for users that don't consent to be tracked right? So on website and apps, you need to give consent to be using things like Google Analytics and Google Analytics 4. So what this behaviour modelling does is it uses very clever models to fill in those gaps to estimate the actual total number of users and page views and sessions that you would attract if you didn't need consent.

[00:02:41] **Dara:** So on the surface, a really great bit of news. It's making all the numbers more accurate, filling in all the blanks. Everybody can have complete trust in what analytics tells them now.

[00:02:52] **Daniel:** Yeah on the surface it does. It feels like a really positive view, but then maybe it's the cynic in me thinking about it. But then I think about well why? What's the point? What's the angle? Why is Google doing this? Why is Google for a free version of a product, so let's not forget that we're using a free version here and we're getting access to all of these machine learning models that is on the surface there to help us have some good reports, and it doesn't seem to help Google in any way. Other than the fact that we would want to keep using Google and hopefully spend money through their other platforms.

[00:03:22] **Daniel:** So I think this is something that ultimately one of the reasons probably why they're deprecating Universal Analytics, and this is something, the way they explain this is that they have quote unquote, solved this problem with GA4. There's an issue that third party cookies, some issues are first party cookies and consent throw into the mix which makes tracking in the traditional context very difficult and impossible to, I suppose, from their perspective, make money on. So what we are looking at here is a way for Google to use estimations, very clever estimations, black box models, you know, we don't know how they work. So we've got a level of faith that we're putting into Google here to do the right thing and to model it correctly and effectively. But ultimately what we're looking at is a way for Google to continue reporting in Google Analytics at the same level than kind of like pre cookie banner era, and actually that all comes down to this new feature, which they've rolled out across gtag.js and Google Tag Manager. So we'll probably be talking about tag manager mostly. But this is something called consent mode and consent mode's been around for a little while, but ultimately, if you're not using consent mode in the implementation, you're probably not going to have access to this new feature, whether you like it or not, whether you are for it or against it.

[00:04:27] **Dara:** Do you want to tell us a bit more about what consent mode is?

[00:04:30] **Daniel:** Yeah cool. So consent mode is a way of managing consent within the Google and non-Google tag ecosystem. So most people by now will have some kind of like consent mechanism or consent management platform, so CMP for short on their websites and apps. And that's all well and good but actually that doesn't do what I call called the plumbing. It doesn't plum into things like Google Tag Manager very well, definitely not automatically. What consent mode is, is that plumbing, it's the glue between your consent management platform. So what levels of consent and what things the users have consented to. And then the actual application of the tags, the implementation.

[00:05:03] **Daniel:** So what it isn't, it isn't Google’s CMP, it isn't a way to collect consent using a Google technology. What it's saying is that, okay, well, let's manage all of my tags based on certain levels of consent. So what you can do in Tag Manager quite simply is you can just roll out and say, if someone's opted into analytics tracking, then deploy all these tags. If they haven't opted into this tracking, then don't fire these tags. So it's really quite a clever kind of like trigger system or trigger grouping system that's just a bit more easy to manage and streamlined. However, there is one benefit from the Google tags, so if you've opted into analytics tracking, great, you can do the regular GA (Google Analytics) tracking. If you've opted into advertising tracking, then great you've opted into, things like Google Signals and some of the additional things that Google Analytics can do.

[00:05:48] **Daniel:** But if you haven't, if you haven't accepted analytics tracking or ads track, what happens is Google Analytics can do something new, and what it does is it sends what it calls a cookieless ping. So it doesn't have access to the cookies and it doesn't contain the user ID or the client ID, and it doesn't contain a session ID, but it still sends that piece of data, that event, that page view that click event or whatever you're tracking, it still sends that into the Google Analytics servers. So in a sense now in the backend system, within Google Analytics 4, you have two types of data, you've got consented data and unconsented data. So the consented data have their user ID and the session ID, and you can tie all the things together and do attribution modelling and you can do your kind of total session reporting and user reporting, all the fun stuff that we're used to. But now we've got a bunch of data that's each one is completely different, right? We don't know who and what session has triggered each of these events.

[00:06:38] **Daniel:** So that's the kind of pool of data that now Google Analytics 4 is using to do the modelling, to do its behaviour modelling. By the way this data, whether it's consented or not is exported to BigQuery. So if you've got the BigQuery export, you can actually do your own behaviour modelling if you're really, if you're really keen as well. So what consent mode is just to recap, because I know I've talked a lot. But what consent mode is, is a way to manage consent levels within something like Google Tag Manager so that you can fire tags conditionally based on what consent you've got from your users. But there's one small caveat there, that the Google tags have a very special bonus case that they have this whole cookieless ping, cookieless tracking activity if you don't opt in. And again, without doing this, you won't have that set of data to do the modelling. So that's really the kind of nuts and bolts of how this behaviour modelling works, is you have to do this consent mode and you have to be collecting these cookieless pings, so that these models can start training.

[00:07:27] **Dara:** So it's interesting, it's called behavioural modelling. It's relying on consent mode, but it doesn't affect the existing conversion and attribution modelling that they have in GA4. Why does the conversion modelling not need to use consent mode, but this new behavioural modelling does? Do we know how those two sets of model data work together? So if this is trying to plug some of the gaps on users and you maybe user based metrics, how does that then fit in with the conversion modelling that it's using?

[00:07:58] **Daniel:** It's a really interesting question, I don't know if we're ever going to get an explicit answer, definitely not from Google. I think if the rest of us are just going to be guessing, maybe some educated guessing going on, but it's still a guess. From what I understand, the only way that you'd be able to use this kind of modelling within attribution modelling is if you're fingerprinting, if you're fingerprinting user behaviour and you're somehow stitching unconsented data back into users. So you have a user journey to do things like attribution modelling. Google have come down very, very heavily on the kind of, we will never do fingerprinting. We're never going to ever do fingerprinting in any of our technology. So what this modelling is doing is looking at aggregates, so it's going to be like total users will be modelled up to be the kind of blended total users, but it's not going to blend down or it's not going to model down to the individual users.

[00:08:39] **Daniel:** So that's why things like attribution modelling won't actually be affected by this, and I would imagine things like conversion modelling also. Again, because it uses that same kind of data. So that's my interpretation of why things like conversion modelling and attribution modelling, data-driven attribution, won't be affected by this behaviour modelling because they're actually not stitching together or fingerprinting user journeys, what they're doing is just taking a total number per day, which is total users or sessions or page views, or by whatever dimension you've got and then modelling that number, not the underlying data.

[00:09:08] **Dara:** But is the total number of conversions and the conversion rate metrics, do we know yet if they're going to be modelled based on the behavioural modelling or are they going to be using something else?

[00:09:21] **Daniel:** No, they are everything like that, so conversions, purchases, revenue, sessions, users, they're all going to be modelled. But as of the modelling effective date, so in some cases we've seen that as the 15th of May, some of the times, the 10th of May, but it seems to be around sort of middle of May from our experience, it could be different for you. “As of that date, analytics is estimating all possible data that's missing due to factors like cookie consent” and I'm quoting directly from their webpage there. So what they're saying is they're going to be modelling every single piece of data that they can find. And what I've done is I've pulled a, obviously I'm not going to tell you who it is. But I've pulled a couple of sets of data from some of our clients that are using this already that have been seeing the effect of modelled data, using things like consent mode, comparing that to the observed data, the real data, the consented data. And what we found is that yes, things like users and sessions have been modelled and we'll go through some of the numbers we've found in a moment.

[00:10:12] **Daniel:** But just for the sake of it, I just wanted to pull a couple of other stats. I pulled out pageviews, so that's the count of the page\_view event, purchases (an eCommerce website) and revenue, all five of those numbers, users sessions, page views, revenue, and transactions, all five of those have been modelled using this behaviour modelling. So for this one client I've pulled the last 28 days of data up to yesterday's date, and just to give you some context, I won't give you the actual numbers. But we've seen a 25% increase in the total count of users, so 25% of those users have been modelled as it's gone up 25%, sessions has gone up 24%, pageviews has gone up 16%, purchases has gone up 8.5% percent and revenue's gone up 9%.

[00:10:52] **Daniel:** So it looks like less severe of an increase using the model data than the observed data as you go down into the kind of purchase and revenue line, but one would imagine that if you are looking to purchase through this website, you're more likely to have consented for tracking right? I think that's the idea, at least. But yeah, so we are seeing everything from 25% difference in sessions and users down to sort of, 8, 9, 10, 16% difference to page views, purchases, and revenue. So yeah, quite a big difference. Again, this is just a one sample that we've pulled just for the sake of having this conversation, but already feels quite a substantial proportion of their traffic.

[00:11:26] **Dara:** It is and going back to what I said at the beginning, and even though I said it in a kind of half jokey way, but on the surface this seems like a good thing because it's not inflating the count it's trying it's best to close that gap that has been caused by users who obviously don't consent and are using privacy blockers or whatever. So this is Google's way of putting all their might behind trying to say we want to give you a better view of those numbers, bearing in mind some of these people aren't going to consent. So I'm seeing it as a good thing, even though we don't know how it works, but you know, with the Google brains on this, you'd have to think their modelling is certainly better than most companies could do on their own.

[00:12:04] **Daniel:** Yeah I mean, I'm not going to disagree. I think they've got the might and the power and the time behind them to be able to kind of perfect these models or at least make them as accurate as possible. I think like anything, I think I'm always going to be the glass half empty or the cynic on this, just because it's like, if I don't exactly how it works, how would I ever know that I can trust it? Or how do I know how I can use it? If it's doing something that I have no idea. But I think this something we were chatting up actually about off mic Dara the other day, which is actually, is this any different to what we had in Universal Analytics?

[00:12:33] **Daniel:** So take Universal Analytics sampling for example. If you run a report in Universal Analytics, and it turns out that they had a 10% sampling in the report. Would you go out of your way to kind of force kind of an unsampled report? Or would you go actually, if you're just looking at trends or looking at what I'm looking at, that's fine. I'll take that on the chin, and I was going to use the data in the same way I would've done, even if I knew it was exact data. But then I suppose pose that question to you Dara, if that was 10%, but then what if that was 25%? What if in some cases that's more, what if that's 50%. So in a sense, the sampling and this kind of behaviour modelling kind of works in a similar way where it takes a partial set of the data and models up the difference.

[00:13:11] **Daniel:** So in my head, it kind of works in a similar context, obviously different approach and different data sets behind it. I suppose that's the way to think about this is like what level would you be comfortable with? What's that tipping point where it becomes maybe a positive to see what they're doing, to becoming, I can't use this data, even if it's got good intentions.

[00:13:28] **Dara:** Maybe I'm not thinking this through enough, but I'm thinking it's different, even though I get the analogy with sampling. But with sampling, it was taking a subset of the actual data to speed up the reporting, the processing and it was modelling based on that. But some of the metrics, presumably like session-based metrics, if we're allowed to still talk about sessions. But the session-based metrics and the event data that is an accurate count isn't it? It has the cookieless pings, it's not having to guess what those actual numbers were, it would know what they are.

[00:14:02] **Daniel:** Well, you're right with the event data, but not the session. I think the session falls under the same umbrella as the user data.

[00:14:08] **Dara:** Because it doesn't have a session ID.

[00:14:09] **Daniel:** Yeah, exactly. So the session ID and the user ID is not provided. That's the kind of the principles of those cookieless pings. So anything at a raw kind of like data level. And I think that's why you get access to these in BigQuery, to be honest, that's why they feed them through into the BigQuery export. So there is a sense that you're quite right, I haven't validated this data set I've pulled out to know, for example, the purchases and the revenue should be exact, right? Because we have the purchases and the revenue event data. They just have cookieless pings, but I don't need to know the session and I don't need to know the user to get the totals of those.

[00:14:41] **Daniel:** So what I could do is use the BigQuery export to kind of validate that side of it. Obviously there's always going to be a level of like missing data in GA (Google Analytics) or kind of, you know, web based and there's issues inherently with that, but I think we can all accept. But actually that's a really interesting point, I’ve not thought about are they modelling the event level data because the numbers have gone up, we saw 8.5% and 9% difference for purchases and revenue for this ecom website. But is that different again to what they have in the BigQuery export, I don't know. So I need to kind of take that away and maybe update on a later episode of just what that difference is.

[00:15:12] **Dara:** Yeah see, that's surprising, isn't it? The purchases is different, because you would think that they would, if they have account of something, they'd use it. And then they'd only use the model data to fill in the blanks or to tie it back to a number of users.

[00:15:24] **Daniel:** Yeah and this is really interesting actually the way they word it. And I think this is what's going to, I mean, if it's throwing us, I think it's going to be throwing other people too right? And this is the along with this rollout of behaviour modelling, using consent mode, they've also updated a feature in Google Analytics 4, which is called the reporting identity. So in the GA4 property settings, one of the sections is called reporting identity and unlike Universal Analytics where it just used the cookie ID, there was some half arsed approach to using user ID features as well, which never truly worked the way we wanted them to. But the reality is Google Analytics 4 has a fresh approach to it and it uses like a hierarchical approach to user identification. So if it's got a user ID, it will use that if not, it defaults to a cookie ID.

[00:16:03] **Daniel:** But if only it was that simple, there's now multiple different types of identity you can use within GA4. And the default actually has four different layers of user identification. So at the very top, if you provide a user ID, it'll take that as the source of truth, kind of the golden number, the golden user ID, we won't kind of question you about it. If there's no user ID, it then uses Google Signals. So Google Signals is Google's sort of user graph that it identifies across people logged in across multiple different Google accounts. If that's not there, it'll use the device IDs, now we're back into cookies, right. So third layer down, we're using the cookies, and then there's a fourth layer beyond cookies. So where there's no cookie, it then uses the model data.

[00:16:42] **Daniel:** So now we've got this sort of ecosystem of identification. Some of it, we don't have access to like the Google Signals data. But the rest of it we do like user ID, device ID and the modelling is the new feature they've just rolled out. I think where I'm going with this is that the default is what's called ‘blended’ as a reporting identity, which uses all four. You have the choice of turning them off, so you can say don't use model data, just use observed data, in which case, then it'll use all of them, but the model data, and there is a secret third option which is tucked away under the ‘show all’ button which is then to device based only. So you can if you want go back into just using cookies, just like Universal Analytics did. So if you wanted to do a side by side comparison, you can.

[00:17:20] **Daniel:** Probably one of the best things about this feature, so there's a lot of questions I have around it, but one of the best things is that this isn't a permanent change. So you can change this setting, click refresh in your report and it'll just re-aggregate the data based on the new identity. So you don't have to worry about changing this and then affecting data going sort of historically, or going into the future. So that's actually how I pulled that data is I used the blended data using the model, using the Signals and using everything else. And then I turned that off and just looked at the observed data and then just refreshed the report. I just clicked refresh in the browser and pulled out the new numbers. So I think that that for me is a really useful part of all this, that you don't have to use this data.

[00:17:59] **Daniel:** The issue, and I think this is something you raised to me the other day, which is that when you go into any sort of standard report now in the reports workspace in GA4, you get where the sampling indicator used to be in Universal Analytics. You'll get a very similar indicator in any report that you happen to go in and just as a heads up, there is no sampling in GA4. So the reports workspace, there is no sampling. So I don't know why, but it actually tells you this is an unsampled report, there is a hundred percent of available data. But it will also now tell you if you're using model data. So there's a secondary part to that that says this is using model data to account for things like missing cookies or not complying. But the issue for me is that it doesn't tell you the percentage.

[00:18:38] **Daniel:** So going back to my analogy of Universal Analytics having sampling and this being kind of analogous to that. In Universal Analytics, you could see quite easily, okay we are using 20% real data, 80% sampled. Whereas now we just know that some is modelled, but we don’t know how much, unless we do go through the motions like what I just said of changed the identity, refresh the report, change it back and see what the difference is. It feels quite long winded, it feels like they could surface that because they do know, I feel like it's just something they are deliberately decided not to show.

[00:19:08] **Dara:** It'll be interesting to see, it could be that it does roll out, but at least in the meantime, even though it is a pain, at least you can do it in the way that you've done, where you can switch between them and then export the data and compare. And I just wanted to say as well, I feel like I'm in a very favourable mood today to Google. I seem to be the playing the opposite to your cynic, but the flexibility it offers it's not bad is it in terms of those reporting identities? Because if you, for some reason, want to stick with good old fashioned device ID and just rely on cookies, you can do that. But you can also go for all the observed options if you have user ID, if you have Signals, or you can then choose to go the whole hog and have blended and include the model data. So you get kind of like multiple choices, do you want model data or not? And then do you want to have the Signals user graph data and the user ID options as well? So you've got a fair bit of control there in terms of what you can and can't do, and the benefit is you can switch between them.

[00:20:02] **Dara:** Maybe the reason I'm feeling so positive is I'm just comparing this to some of the more rigid features in Universal Analytics and the fact that you can actually choose something now just seems like a step forward.

[00:20:12] **Daniel:** For sure and the whole it back dates, which is something that Universal Analytics I think never had. If I had to pick a hole in this, and that would be the fact not to do the kind of model data, which is kind of the focus of this conversation, but as a slight side point is the Google Signals aspect of this because there is no option, which I would love to see, which is using device ID and user ID only. So there's no way to use your IDs and the cookie IDs without putting Signals in the middle. There's lots of issues with Signals, I think we can have another conversation entirely around what Google Signals is and some of the issues and the kind of challenges we found with it.

[00:20:46] **Daniel:** But that's a whole concept of thresholding, so thresholding comes in quite quickly when you're using Google Signals as a reporting identity. So there's layers, there's layers here, right? We're talking about one of the four layers, you know, that we could have many conversations around user ID and what cookies mean in the context of the today's landscape, modelling and also Google Signals. For me, it's like you blend them all together and it becomes so complicated. You know, even the best intentioned people, specialists in this will struggle to keep up and understand what is a user in GA4 anymore, because we don't know, it could be one of many things. Some of it is hidden from us, we don't know what Google Signals is and how it's working, and how it's joining users, that isn't exported to BigQuery by the way, it's not exported anywhere else. So again, validating between the UI and BigQuery is now never going to match, but how much is enough for us to be confident with it? I don't know.

[00:21:35] **Daniel:** So there's lots of ambiguity I think that things like behaviour modelling, this kind of machine learning modelling throws up on top of all the other ambiguity that GA4 had that Universal Analytics didn't. So I'm not saying that GA4 is better or worse, and I'm not saying that it's because of GA4. It's actually just a symptom of the times, you know, consent management, cookie deprecation of third party cookies, first party cookies being sort of ring fenced into sort of one week or 24 hours. I think there's just a natural evolution of what we used Universal Analytics for. I think GA4 is as good as any replacement for that.

[00:22:05] **Daniel:** Maybe I need some positivity from you Dara on this. But I think maybe this is in some way, deliberate, maybe from Google to complicate things and to make things a bit harder to unpick. So that again, all of this data is being curated by Google. You know, it's not just a data aggregator and processor anymore. This is a curator, it's creating the numbers, even if they didn't exist, which again is great on paper because it's helping us work around a lot of these sort of issues that we are coming up against nowadays. But again, I have to go back to one of the first things I said, it's like, what's the point? What's in it for Google. Why is Google providing this as part of a free service? And that's the thing that I still kind of come back to.

[00:22:40] **Dara:** It's got be the obvious hasn't it? It's the same reason that they've always done it, to keep people in their ecosphere and get people to spend more money on their media. Something I was surprised about yesterday when we were talking about this in prep for the podcast Dan, was the fact that the default for these reporting identities has actually changed as a result of these. Well I guess as a result of the introduction of this behavioural modelling with consent mode.

[00:23:04] **Daniel:** Yeah so the default now is ‘blended’, which uses that for reporting identity hierarchy. You can, again, change that without affecting the data too much, historically or going forward. So you have the choice, but the defaults changed. Much like when they rolled out data-driven attribution (DDA), they just rolled it out as the default, switched everyone over. And I think in a sense, maybe they can get away with it because you can just change it back but again, who does that? We've been working in this space long enough Dara to know who changes the Session Timeout, or Campaign Timeout in UA (Universal Analytics) or the engagement timer in GA4, just people just don't bother really, so I think it's quite sneaky of Google doing this. So the default is ‘blended’, so the next time you go into GA4 it is worth checking to see what that reporting identity is, so that you are aware of what you're looking at.

[00:23:44] **Dara:** And this is potentially a biggie because just to be clear so the difference with this change now is it's forced the introduction of the model data to everybody. So your numbers will change from when this change kicked in, and I think I'm right in saying that this is maybe rolling out on a timeline so it'll be available in different properties at different times, and you need to meet the prerequisites as well, which maybe will come onto next. But if this is available in your property, then you're suddenly going to see more users from the day this change kicked in. Again, going back to saying that this is not necessarily a bad thing, but the fact that people might not know about it is worth calling out because we might have a lot of companies saying brilliant, we've got loads of new users today we must be doing something right. What campaign did we launch that's driven this increase in traffic to the site, and actually it's just the kind of sneaky introduction of this model data behind the scenes.

[00:24:38] **Daniel:** I think that's my reticence for this, and it's not the fact that it's impressive feet of engineering or anything like that. I think it's more the fact that they are changing the default and I don't think people are well versed or educated or aware enough of what that means fully. I mean, we're struggling with this, and this is our day job. So, I think the reality is when it comes to any change like this yes, it's great they can put a banner in GA (Google Analytics) at the top with a ribbon that says, this is now happening. But again, we're the minority really, you're logging into GA4 and going to get the numbers. A lot of people will be sitting in getting the data from a warehouse or pulling it into Data Studio or Tableau using the API. So I think this is the key thing, and the reality is the kind of days of validating data is kind of, I think finally ended with this. Yes, I can validate revenue that's one number. I can't validate users, right? That's the whole point is I can't check the baseline to see if this is accurate or not.

[00:25:28] **Daniel:** So I think with all this, with the kind of the post validation era and all that stuff I feel, and I think I can see a uptake from this account that I'm looking at, it was introduced on the 15th, I can see a definite uptick in total users on the 15th. But the reality is that could be fluctuations, that could be seasonality, that could be a hundred other factors, but again, I've got no way of validating this. If during this role out specifically, but when you're doing year on year, or if you're doing your month or month, even over this period of time, you're going to see a jump because, you know, in the introduction of this model data has been introduced on a date. It hasn't back dated and rolled back over time, it is introduced on a date. So there's definitely like an onboarding period with this kind of stuff as well. And again, the less we know about it, the more you feel like, oh, something interesting's happened. It has, but it's not in your performance or marketing or the website or the apps, this is a modelling thing, this is a backend thing for Google.

[00:26:20] **Daniel:** So I think, again, that's my reticence around this feature. Not because of what it is necessarily I think it's just the approach to rolling this out and they are very aware that it's very confusing and complicated and a lot of people won't realise this is happening. I feel like they're even banking on that.

[00:26:35] **Dara:** I can't help but think and this is probably me just oversimplifying as I tend to do, but I can't help but think that this is fighting a bit of a losing battle that we should give up on trying to estimate or model what we don't know and almost split analytics into kind of observed data, even if it's with cookieless pings. In other words, maybe know for certain you have X number of users because they've identified themselves and you have the full end to end journey for those users. And then you have this separate category of data, which is the anonymous users, and that's accurate to an event level and maybe to a session level, but at the user level you just don't bother.

[00:27:14] **Daniel:** I'm with you, just because again, what's the point if it's never going to be a hundred percent. I play devil's advocate and say, what about my year-on-year reporting Dara? What about that? How would that look if the numbers go down? Yeah, just don't do it. I mean, that's easy enough for us to say, right? Just don't do year on year reporting, don't compare to a previous time.

[00:27:31] **Dara:** But can you anyway, with all of the changes that are happening, is it fighting a losing battle trying to compare year on year, unless you did it in the way that I'm suggesting, which is you kind of break it down into what you know and what you don't know, and you only compare on a more granular level, so you compare conversions but you don't compare users.

[00:27:46] **Daniel:** Yeah it is a losing battle for sure. But then this is the thing that I've always struggled with. And I think eventually you just kind of cave and do it, but it's like, when you come to compare any period on period, you have to be aware that the circumstances of the previous period is not the identical to the current circumstances, whether that's law, whether that's the rollout of GDPR or any other privacy update or adjustment local, national, international. Whether that's technological, where the latest version of Safari rolls out with a different approach to collection and cookies and whatever than the previous version or, or even whether it's your website. This is something we've always found, which is websites hopefully become better and more efficient, they evolve and change. So the function and what people do there and how they use them changes. So the reality is you take a point in time, any point you like and then you compare it to right now, everything's different. I feel very little value in comparing what was to what is now, because it won't help you with what is next necessarily. We can infer some trends and some changes, but it's so difficult to know what you're seeing isn't because of, you know, Chrome's latest update.

[00:28:47] **Daniel:** Introducing the cookie banner, this is the biggest one we've seen so far, which is fundamentally all of a sudden, you're saying no to collecting a bunch of data that you might have done beforehand. So, yeah you're right I think it's a losing battle. It's an uphill race, it's never going to be there, but it won't stop people doing it. And I don't think that's because they don't care. I think it's actually because they don't know. And I think this is the hardest thing it's almost on, you know, and like maybe everyone listening can relate to this, but it's almost like on us to explain.

[00:29:14] **Dara:** Just as a, well last point for me, at least just to kind of wrap up, I did mention the prerequisites a minute ago, so this isn't going to be available for everyone that's the reality of it. So first and foremost, you need consent mode turned on, which is something that's available to everyone that's an easy tick in the box. But you need to have a thousand daily events, and tell me if I'm wrong here, you need a thousand daily events for at least seven days where they've denied consent and you need at least a thousand daily events for seven days where they've granted consent for this modelling to actually work. And I think I read that even if you have that, there's still no guarantee that the model is going to be trainable in your particular case. So this isn't going to be a feature that's available for every website.

[00:29:57] **Daniel:** No, you're quite right. You need a thousand consents and a thousand unconsents per day for an extended period of time so that it can start modelling the difference. You need a period of time where that's consistently true so that it can start modelling. And if you've got a 90% opt-out rate, then you're probably never going to have this modelling feature turned on. It's never going to be modelled sufficiently enough to be surfaced to you.

[00:30:17] **Dara:** Famous last words, might be a topic we come back to with an updated point of view, once this has been out in the wild for a bit longer. Okay, switching gears, it's becoming my trademark line saying apart from worrying about these GA4 features, what have you been doing to wind down?

[00:30:32] **Daniel:** Well, a couple of weeks ago, maybe even months ago now I might have mentioned on one of these wind downs that I helped out in a very small way contribute and build a indoor skate park locally to me and it's finally open. So yesterday, as of the time of recording, yesterday was the opening night, and I went down and had a skate and it was really good fun and I'm really looking forward to spending a bit more time down there and skating something new, something different, something big, it's really big which is awesome, a different type of skating altogether. So yeah, new skate park, a friend of mine opened it up and just going to enjoy skating something new.

[00:31:04] **Dara:** You've got the scratches to prove it. Obviously our listeners can't see this, but we're on video and you've got some scratches, war wounds on your elbows.

[00:31:12] **Daniel:** Yeah as much as I try, I don't end up standing on top of my skateboard as much as I'd like, so, yeah, it doesn't feel like summer unless I have something hurting, you know, it's very much ever since I started skating, it's almost like a badge of honour, a little bit. You got to have some wound that you're nursing. But yeah, looking forward to skating there more, it was only a quick session yesterday. So, we'll go back later in the week and see how many more scrapes and bruises I get. Anyway Dara how about you? What have you been doing to wind down and escape from consent management and consent mode and other GA4 features?

[00:31:39] **Dara:** I feel like more often than not I talk about a film I've seen, which is what I'm going to do now, but I I'm going to just state, I do also get outside. So I've been doing all my usual dog walking, running out enjoying the sunshine. But we did also go and see Top Gun Maverick a couple of weeks ago, which was great it was better even. I thought I'd enjoy it, but I thought I'd enjoy it as almost like a bit of a guilty pleasure but it was actually really, really good. So I would recommend it to anyone if you haven't already seen it. And I wouldn't even say you necessarily need to have seen the original film. It would help, but it's not essential so yeah, really good.

[00:32:11] **Dara:** All right, where can people find out more about you Dan, apart from like the skateparks in Lewes. How can people find out a bit more about you?

[00:32:18] **Daniel:** So [LinkedIn](https://www.linkedin.com/in/danielperryreed/), search my name, Daniel Perry-Reed, and there is also my website [danalytics.co.uk](https://danalytics.co.uk/?utm_medium=podcast&utm_source=the_measure_pod&utm_campaign=42).

[00:32:25] **Dara:** And for me, it's on [LinkedIn](https://www.linkedin.com/in/darafitzgerald/) if you want to reach out to me. Okay that's it from us for this week to hear more from me and Dan and around all things GA4 and analytics related, all of our previous episodes are in the archive at [measurelab.co.uk/podcast](https://www.measurelab.co.uk/podcast/?utm_medium=podcast&utm_source=transcript&utm_campaign=42). Or you can obviously find them all in the app that you're using to listen to this.

[00:32:44] **Daniel:** And if you want to suggest a topic or ask us any question around consent mode, behaviour modelling, or any of the other fun stuff we mentioned today, there is a form, another link in the show notes, but we've got a Google Form in there you can submit and get through to us that way. Or you can just email us at podcast@measurelab.co.uk and me and Dara are on the other side of that.

[00:33:02] **Dara:** Our theme music is from [Confidential](https://open.spotify.com/artist/4erl5USs7RuiUU4QeiEQFO?si=xyGxD84PQHO78TKlvmZKLw), you can find links to their Spotify and Instagram in our show notes. I've been Dara joined by Dan. So on behalf of both of us thanks for listening, see you next time.