GDPR and Google Analytics

Marissa Goldsmith and Heather Burns //
WordCamp Belfast // 26 May 2018

Who we are

Marissa

- Based in Washington DC
- Former web developer
- Now a full time Google Analytics specialist for charities and not-for-profits
- This is my second WordCamp
- Curator of an appalling collection of 1990s pop music
- Not a lawyer

Heather

- Based in Glasgow
- Former web designer
- Now a full time digital law and tech policy specialist for agencies and digital businesses
- This is my 18th WordCamp
- Curator of an exquisite collection of French pop music
- Not a lawyer

What we're going to cover

What you need to *know*:

- The European privacy framework
- Definitions
- The right to be informed
- Consent

What you need to *do*:

- Configuring Google Analytics for optimal user privacy
- Beyond GA: the risks of aggregation
- Other analytics issues
- Your privacy notices

What you need to know

The European privacy overhaul

GDPR: yesterday!

- Replaced the Data Protection Directive of 1995
- Maintains original principles, expands and modernises
- Data at rest: collection, usage, retention

ePrivacy Directive: TBD

- Will replace the ePrivacy Directive of 2002
- Data in transit: cookies, telemetry, advertising beacons, marketing, and analytics
- Late 2018 - early 2019?

Definitions: personal data

Personal data

Any information relating to an identified or identifiable natural person. This can be one piece of information *or multiple data points combined in a record.* New definitions under GDPR include genetic data, biometric data, location data, and **online identifiers.**

Sensitive personal data

Information about racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health data, sex life or sexual orientation, past or spent criminal convictions

Definitions: controllers and processors

Data controller

The data controller is a person or an entity, such as you or your business, which decides what data is processed, how it is processed, and whom it is shared with. (“Processed” simply means “used”.)

Data processor

The data processor is any person other than an employee of the data controller who processes the data on behalf of the data controller.

Individual rights

- The right to be informed
- The right of access
- The right to rectification
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- Rights related to profiling

For analytics, this means

- Informing your site visitors that you use analytics in your privacy notice, and whether you aggregate that with anything else
- Disclose analytics cookies
- Providing information on how to opt-out
- Removing any analytics records you've created or aggregated upon request

GA & GDPR: who is who?

You are the Data Controller

You are responsible for protecting your users' data, including analytics records and any information you have aggregated with them

Google Analytics is the Data Processor

- Google is the GA data processor for cookie identifiers, IP addresses, device identifiers, and client identifiers
- If you use GA in Europe, you must sign their new Data Processing Agreement - this is their legal backside covered
- You must give them contact information and your DPO, if applicable.

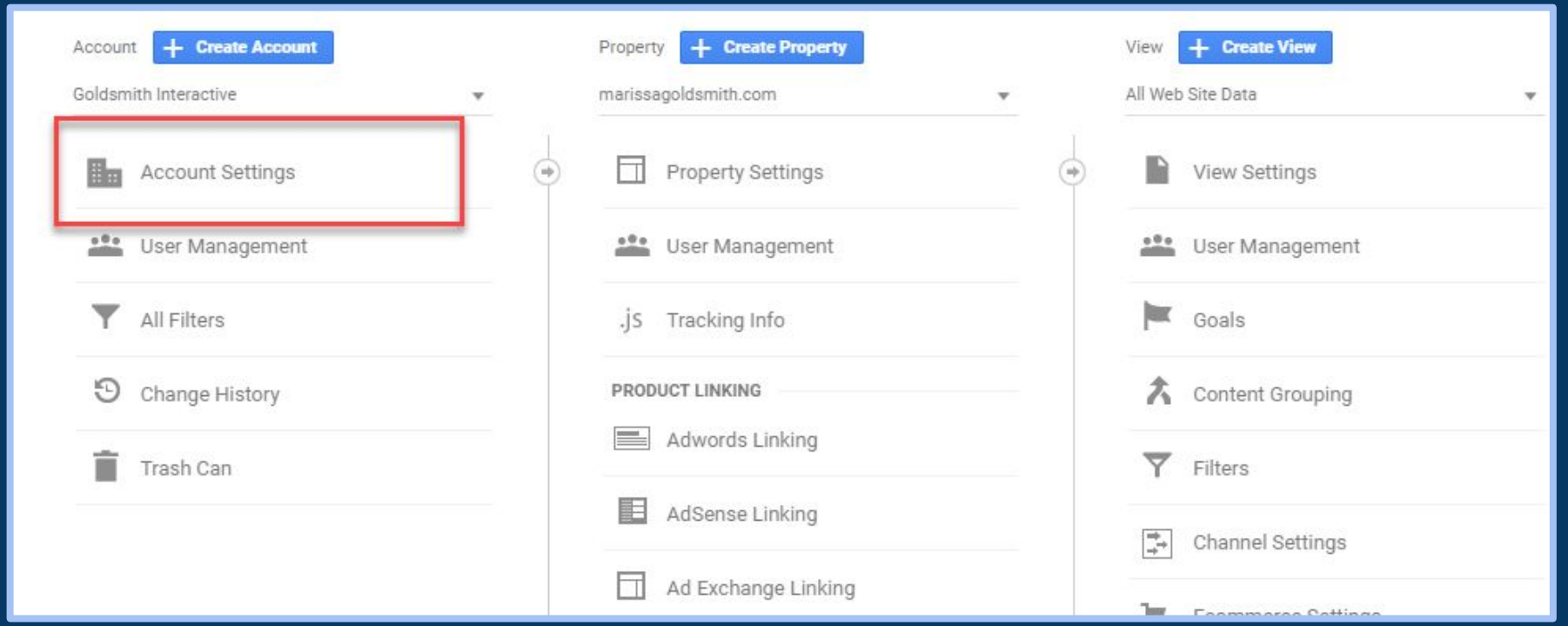


Image: Screen shot of admin screen of Google Analytics. First Column, Account Settings, is highlighted

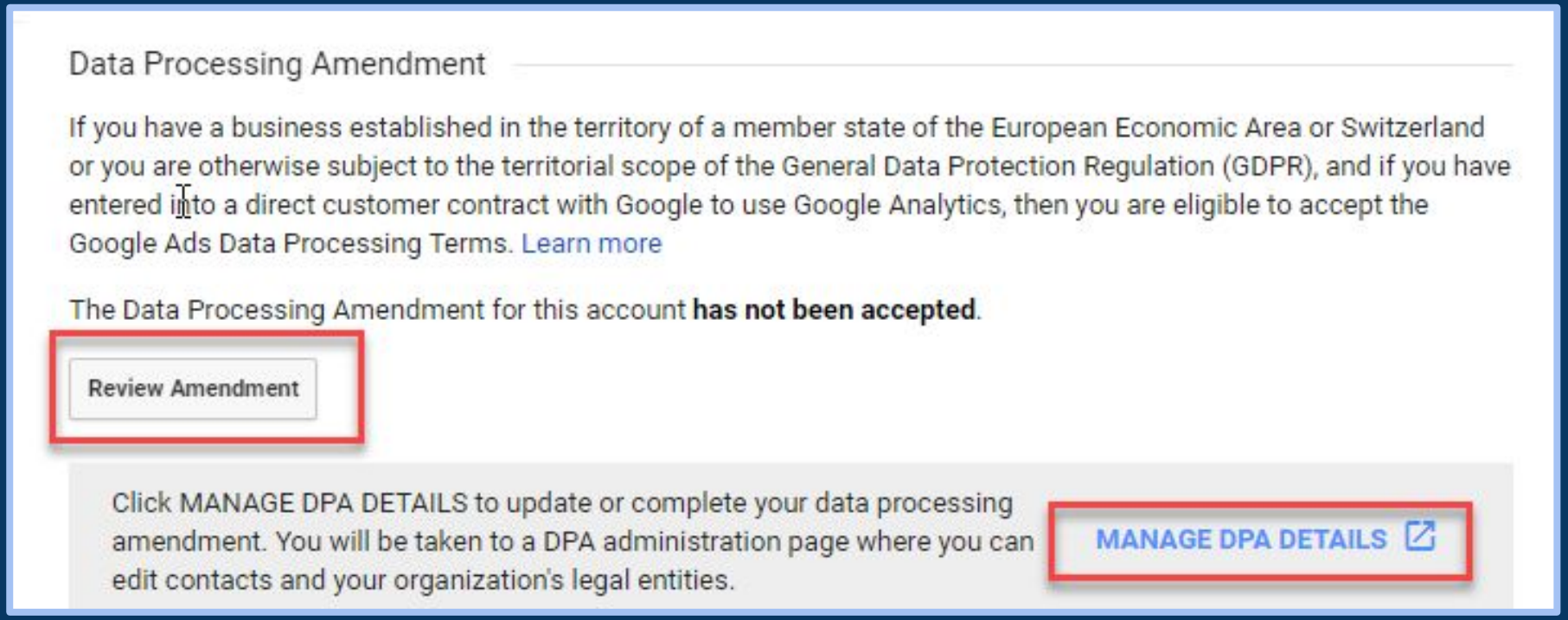


Image: Screen shot of Account Settings. Information on the Data Processing Amendment, with link to Review Amendment and Manage DPA Details highlighted

What does that mean for active consent?

(We're going to fall out here)

Image: Marissa and Heather go to lunch

We're both right, BTW

The contractual view

- You are already complying with GA's terms of service which forbid the collection of PII.*
- You're also compliant with the existing cookie law (of course).
- You will take several of the precautions we'll detail shortly.
- You've got a lovely privacy notice
- So active consent *is not* required.

The ideological view

- Aye, so maybe *you* are only using Google Analytics to collect anonymous web audience data.
- But what is Google doing with that data?
- There is a reason why Google Analytics is free...
- So active consent *is* required.

So what's the best thing to do here?

- Remember your Privacy by Design fundamentals: *collect as little data as is required, retain it for as little time as is required, and delete it when it is no longer needed.*
- Also remember PBD's rule against zero-sum scenarios: *there is a middle ground which respects both user privacy and site functionality;*
- So let's configure GA to achieve the middle ground;
- Let's be as transparent as we can be in our privacy notices;
- Let's look ahead to the future of the ePD.

What you need to do

GDPR and Google Analytics: The Technical Part

(Open up your laptops and log in to your GA accounts.)

GA: IP Address

- IP addresses are Personal Data.
- Google Analytics has a parameter called “anonymizeIP”.
- Set it to “true” and you won’t be recording IP address anymore.

Consequences for data:

- Some of your geolocation information may not be accurate (studies show anywhere from 5-20%);
- You will no longer be able to exclude your office’s IP address (but there are other, better ways to do this anyway).

GA: IP Address

Image: Screenshot of GTM Google Analytics Variable. ***More Settings > Fields to set.***
Second setting sets variable anonymizeip to true.

GA: In the Page Path

- With some tools, personal data can appear in the page URL.
  - *Example:* http://www.mysite.com/?email=marissa@marissagoldsmith.com
- This generally happens when you use “GET” instead of “POST” in your form submissions
- You are already violating GA’s Terms of Service if you do this

Image: Screenshot of a tweet showing a hotel's web site which used the guest's email address in the URL, revealing her interests, room preferences, and membership number.

GA: In the Page Path

- Just Use Post. Go on go on go on go on go on
- Check your WP form plugins to make sure they use Post.
- If, for whatever reason, you can't, you must programatically remove this information *before* it hits Google Analytics.
- Filters and Parameter Exclusions in the settings **are not enough**

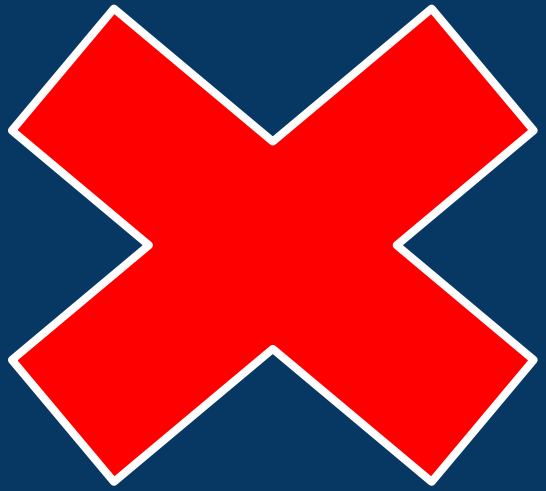
Tag manager users:

- Simo Avaha has a nifty redaction script

GA: In the Configuration

Let’s now review a variety of configurations that cause Google Analytics to collect more data than you need.

Unless you get consent and/or **really know what you are doing**, turn these off.



GA: In the Configuration

- **Property Settings > Advertising Features.** Enable Demographic and Interest Reports
- Turn it off. I've yet to find it overly useful, anyway.

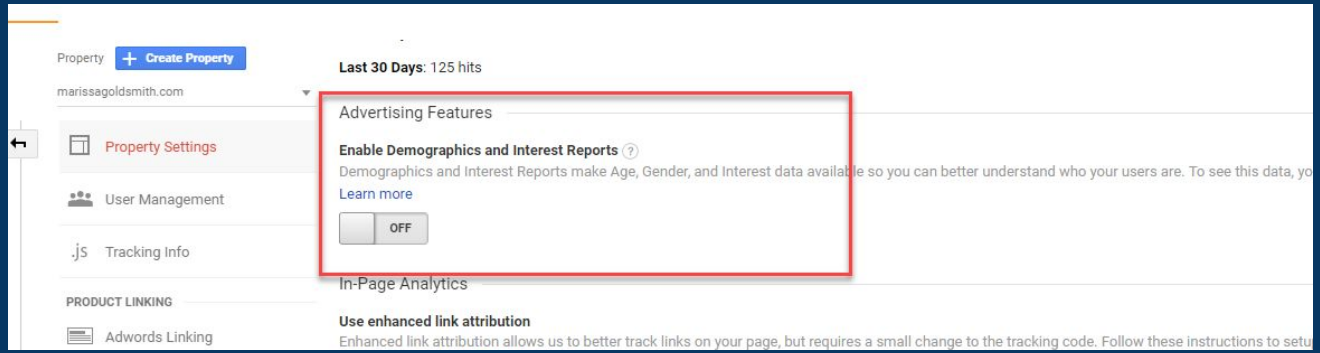


Image: Screenshot of Google Analytics Property Settings. Highlighted feature shows Enable Demographics and Interest Reports are turned off.

GA: In the Configuration

- **Tracking Info > Data Collection.** Remarketing and Advertising Report Features.
- If you don’t use these tools actively, shut it off.
- If you do use it, you’ll need consent.

GA: In the Configuration

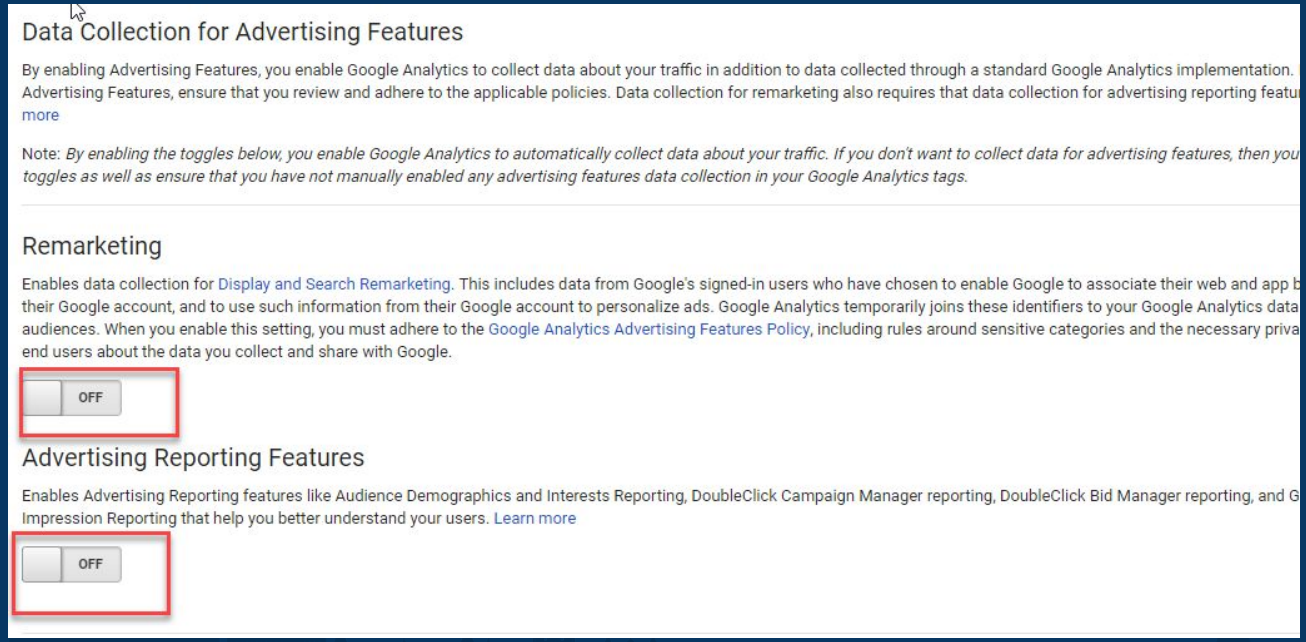


Image: Screenshot of Google Analytics Data Collection Settings. Highlighted feature shows Remarketing and Advertising Reporting Features in the "Off" position

GA: In the Configuration

- **Tracking Info > Data Retention (New For GDPR!)**
- Google Analytics' own default is 26 Months.
  - GDPR/PBD: do not retain data for longer than necessary
  - Your own data retention schedule may vary
  - 26 months is good: it gives you a two year comparison
  - If you look at 3 years worth of "non-aggregated" or event-based GA data at once, pint on me
- Does not affect the standard aggregate reporting.
- **THE SKY IS NOT FALLING!**

GA: In the Configuration

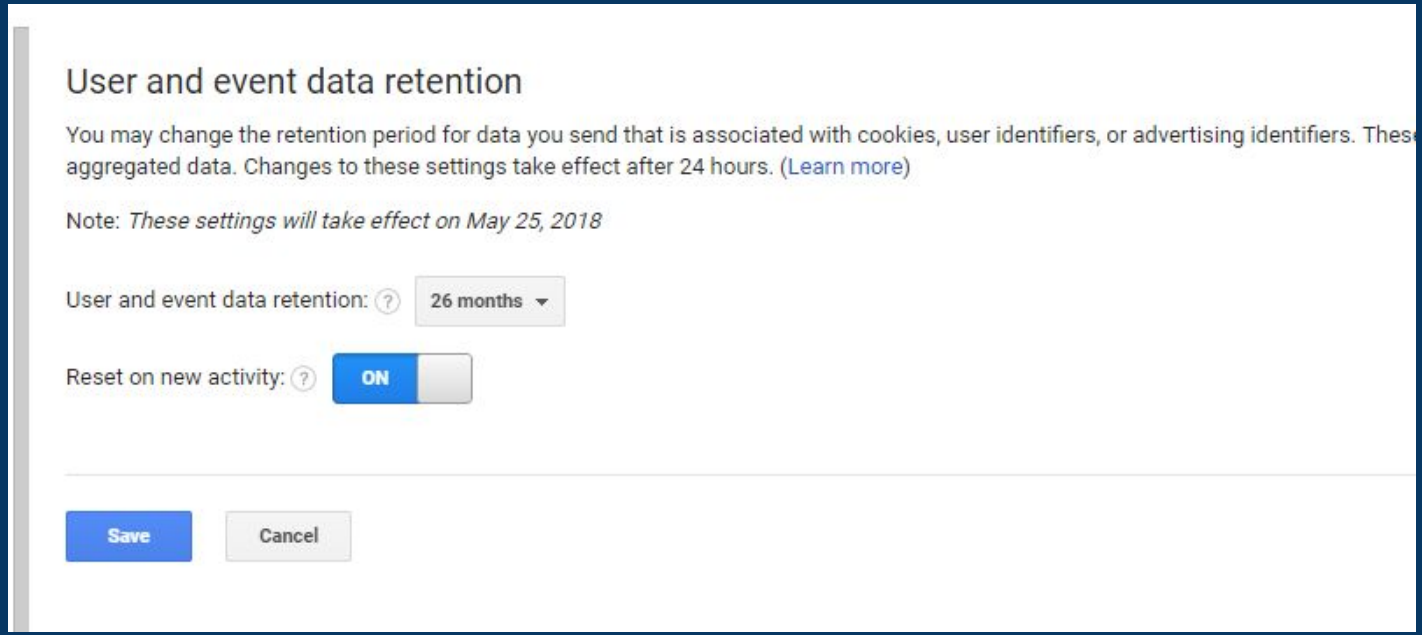


Image: Screenshot of Google Analytics Data Retention settings. It shows User and event data retention of 26 months, and reset on new activity set to On.

GA: In the Configuration

- **Tracking Info > User ID.**
- Feature of GA that lets you stitch user sessions together based on a system-generated user ID (instead of a cookie).
- Most common on sites that require logins.
- Unless you've got consent, turn it off.
- If you're not getting consent, start getting it.
- You'll need to exclude that stitching if the user doesn't get consent.

GA: In the Configuration

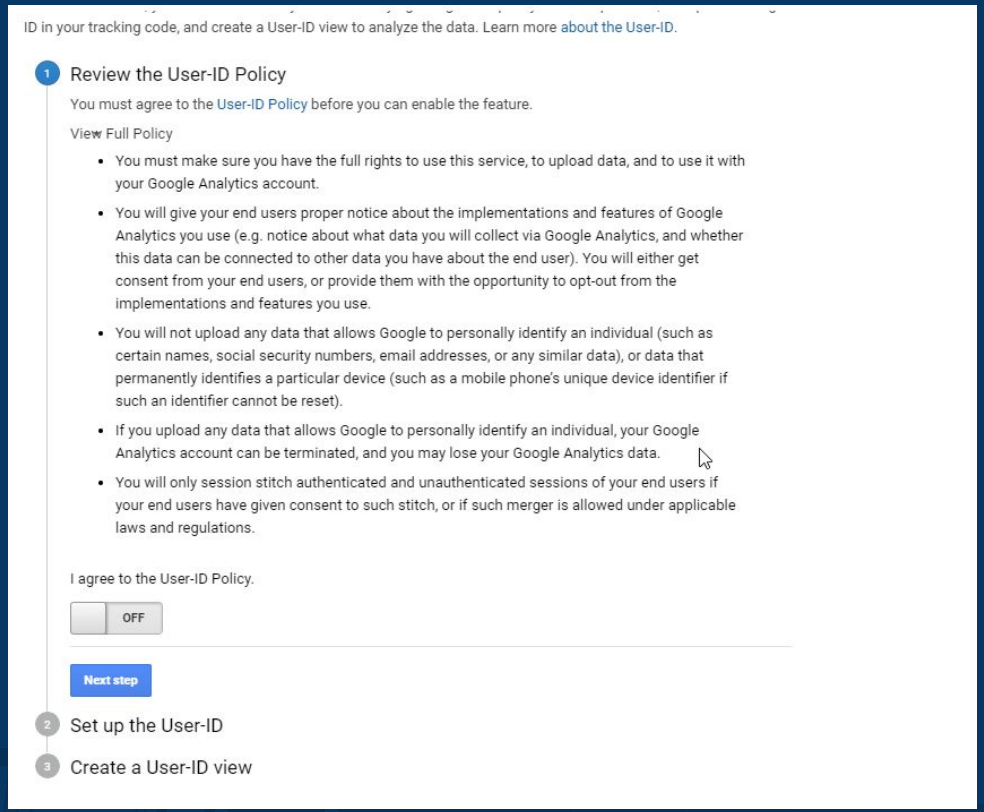


Image: Screenshot of Google Analytics User-ID Policy with a long and lengthy explanation of how you are going to use this information. That doesn't matter though, because unless you've gotten consent, this should be in the Off position.

GA: In the Configuration

- **Product Linking.** Don’t link any product (especially AdWords), if you’re not actively using it.
- **Postbacks.** Don’t have any.
- **Audience Definitions.** Nope
- **Custom Definitions.** We’ll talk about this later.
- **Data Import.** Just be careful (we’ll talk about it later).

GA: With Custom Dimensions & Events

- Custom Dimensions and Event Tracking are the most awesome things you can do get data that is really relevant about YOUR website.
- Just be cautious.
- Do your custom dimensions and events track information about the website (Scroll Tracking, Clicks to Outbound Links, Blog Post Author), or your user (Zip Code, Eye Color, Income).
- If it's the user, you need to make sure that the combination of that data cannot point to an identifiable natural person.
- Also consider how it can mix with data GA already collects (like location).
- Note -these will likely disappear when your retention period is hit.

GA: With Custom Dimensions & Events

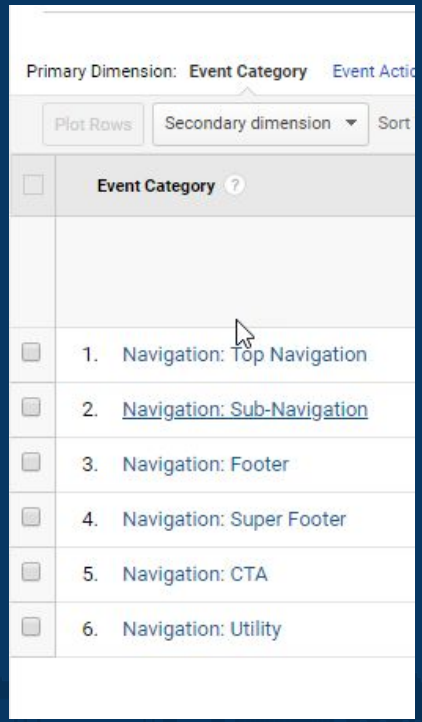
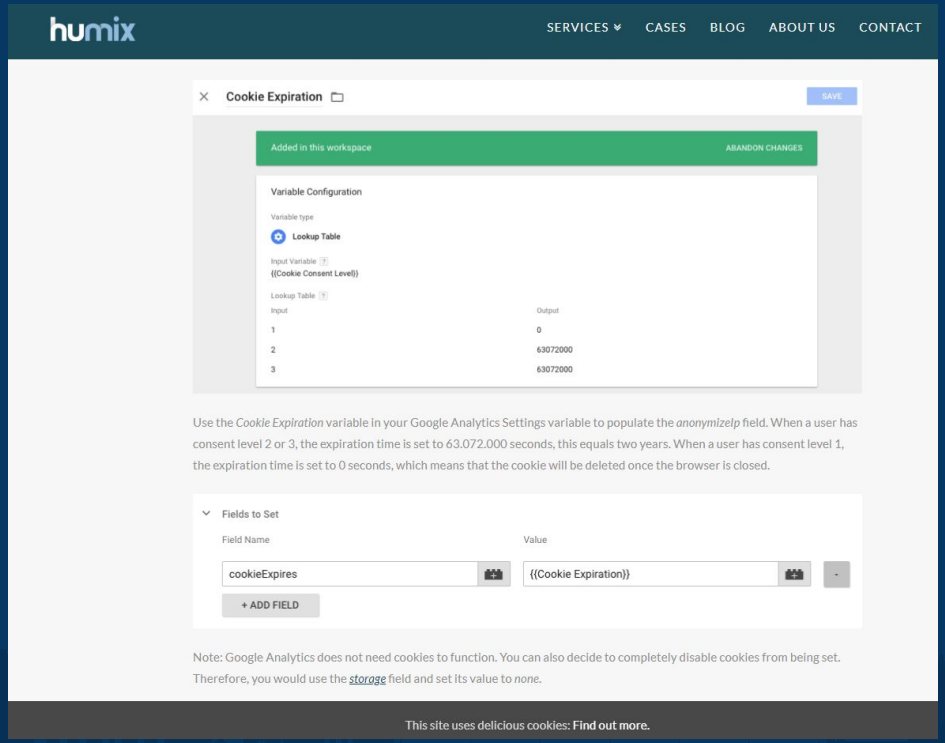


Image: Screenshot of Google Analytics Event tracking, showing clicks on various Navigation Sections (meant for improving the UX experience). GOOD!

GA: With Cookie IDs

- So whether you do this has to do with the whole “should I be worried about what Google is doing.”
- Some folks recommend setting GA cookies to session cookies.
- In GTM, you do this by setting cookieExpires variable.
- Humix has a good article on how to do this (along with logging consent levels).
- Consequence: No more new/return visitor metrics.

GA: With Cookie IDs



GA: Google Tag manager

- So Google Tag Manager is a tool that we use to implement Google Analytics.
- But it implements lots of other things too.
- It’s use falls under the “What Does Google Do With it” argument.
- If you use GTM, go through it. Audit your tags. Delete ones you don’t use. I promise, it will feel good.
- Determine if you need and/or have consent for the ones you do use.

GA: Deletion API

- Was (finally) made official when Marissa was over the ocean getting to this event.
- Allows you to delete data based on GA's client_id.
- Looks promising, but it's too early to tell how it will work in practice.

Don’t forget
to bring it all in the open

Privacy notices

In your notices, you should document

- *Which* analytics package you use;
- *What* data is collected;
- *Why* you collect it;
- *Who* has access to it;
- *How* you audit it;
- *Whether* data is used for profiling;
- *What*, if any, cookies are used;
- *How* a user can opt-out through settings, a GA opt-out, etc

To make that notice, use

- The Privacy Notice tool in your WordPress dashboard

Or

- Any privacy notice page…
  As long as it's not a template
- WordPress.tv: WC London 2018 talk on GDPR privacy notices

Other analytics, not Google!

Analytics ≠ Google Analytics

Advertising and social media tracking

- Twitter: add this to your page header
  <meta name="twitter:dnt" content="on">
- Facebook
- Go to SuperMetrics - don't buy it, but look at their list of providers. If you use them, they collect data.
- Aggregators (BrandWatch).

Other analytics packages

- Matomo (formerly Piwik) is a privacy-specific analytics package
- Statcounter (Irish!) - you can use cookie-less analytics and obscure the last half of the IP address
- Jetpack stats do not collect personal data and the only cookie set is admin-only

Here be dragons: Analytics + other tracking

When are anonymous analytics not anonymous?

When you can take lots of data points and stitch them together.

- We already know NOT to do this in GA by collecting separate events and custom dimensions. But it's important not to do it across tools, such as:
  - Screen capture utilities like Hotjar
  - Facebook pixels
  - Advertising beacons
  - GA Custom Dimension Identifiers
- If you go into your eCommerce tool and see that Marissa Goldsmith, from West Springfield, Virginia, ordered something from my site at 8:33am, and I go into Google Analytics to track the patterns of someone from West Springfield at 8:33am, the data in Google Analytics is no longer anonymous.
- Set strict rules about cross-referencing data sources.

Don’t be a grass

- Marissa’s Law:
  The more data you collect across analytics tools and pixels, the more tempted someone in marketing will be to stitch them together.
- Don’t put a pixel on your site because someone tells you to.
- Develop a governance process around any and all tracking.
- Create a master field map, or conduct a Privacy Impact Assessment, so you know all the ways you may be collecting too much data.

What’s ahead for analytics?

The ePrivacy Directive revamp

May 2018 draft:
“Cookies can also be a legitimate and useful tool, for example in… measuring web traffic to the numbers of end-users visiting a website, certain pages of a website or the number of end-users of an application. This is not the case, however, regarding cookies and similar identifiers used to determine the nature of who is using the site…

Therefore, this Regulation should provide for the possibility to express consent by using the appropriate settings of a browser or other application…

Information provided…should not dissuade end-users from selecting higher privacy settings…including the compilation of long-term records of individuals' browsing histories and the use of such records to send targeted advertising.

Resources

Things we want you to *know*:

- UK ICO: https://ico.org.uk
- ROI DPC: https://dataprotection.ie
- WordPress's GDPR compliance project
- https://afterbrexit.tech

Things we want you to *do*:

- OptimizeSmart Checklist (I don't agree with all of the interpretations, but the tech is good).
- Humix Blog Post on restricting analytics by consent levels.
- Google Privacy Compliance Site.
- Brian Clifton Blog: Google Analytics, GDPR & Consent.
- GA Deletion API Overview.

Thank you!

No sleep ‘til Belgrade