PIWIK PRO

How to Approach GDPR Consent and Data Requirements: A 4-Step Guide for Analysts and Marketers

Table of Contents

Introduction and summary

The future European regulatory environment for data privacy is shaping up. We have the final version of the GDPR and a draft copy of ePrivacy. We don't know exactly how they will be enforced, we'll need hands-on experience for that, but we know enough already to make and execute detailed analytics plans.

If you need a reminder of terminology and of what the regulations mean, check out these articles we've written recently about what GDPR and ePrivacy will mean for analytics and advertising:

- How Will GDPR Affect Your Web Analytics Tracking?
- Your Most Burning Questions About GDPR Answered.
- What Is PII, non-PII, and Personal Data?

Everybody doing web and mobile analytics is a little tense. Big changes, even if we agree with the direction, are hard in practice. What will it look like to define an analytics goal then collect and process data to achieve it? What different ways will organizations use to stay compliant throughout? What kind of software and processes will they need to make it happen?

Remember that this isn't a set of regulations from on high. There is abundant evidence that European internet users (among others) want privacy and data protections. A recent SAP survey showed that the top reason for rejecting a brand was "used consumers' data without their knowledge". It even ranked higher than bad customer service. Increased use of aggressive ad blockers indirectly says the same thing: internet users are fed up with opaque data collection and usage.

There is a good chance that analytics methods complying with GDPR and ePrivacy will also make our visitors happier (and customers, clients, patients, citizens...). That's sensible for the effectiveness of individual organizations and beneficial for the health of the whole analytics and marketing ecosystem.

We won't go into details for specific tools. There are too many possible variations of marketing and analytics stacks to do that. Instead, we've tried to focus on the functions and tasks that need to be accomplished by an analyst or an automated process.

Let's look at each step and explore related issues along the way.

Keeping both regulatory and technology considerations in mind, we've come up with four steps:

Define an analytics goal

Understand data requirements

Engage with visitors about data collection

(Link consent status to data collection)

Maintain the trust of your users

(Respond to changes and data subject requests)

Step 1: Define an analytics goal

In the pre-GDPR regulatory environment, having extra data on hand didn't seem so bad. Maybe it'll come in handy someday, right? Now that logic has been turned on its head. A Data Protection Officer will say "Delete it. Extra data without a specified purpose is too risky to have around".

Privacy by design and data minimization are founding principles of the new regulations and it shows.

Data minimization will most likely come out of a Data Processing Impact Assessment (DPIA), but it also helps to think about it at the level of a single goal.

Let's start with a specific goal for your data and then work backwards to understand the kind of data is needed

- Define an analytics goal such as: track conversions, analyze visitor engagement or record behavior to feed a personalization engine.
- What is the minimum dataset needed to achieve this goal? Also answer these questions:
    - Do I need personal data?
    - Do I need to share this data with a third party to accomplish this task?
    - Does this data contain sensitive information – religion, sexual orientation, etc.?

Each 'yes' to those final questions will complicate your analytics process a little more. So if you need a lot of data, loop back to the beginning and ask yourself how important that goal is. If the goal isn't high priority, can you reduce the scope and use less data? Don’t spend too much time on this yet, you’ll learn more about the overhead costs of data in the next section.

What do we mean by "personal data"?

For now, just remember that any **personal identifier (cookies, device ids, fingerprints) is considered personal data** under GDPR. If the word "tracking" appears in your goal, you probably need personal data. See this article for more info. We'll also discuss data requirements in detail below.

Can't I cite legitimate interests and continue as usual?

Is data processing based on legitimate interests a way around all these extra restrictions? Maybe we can keep the status quo and claim legitimate interests... Not so fast.

There is a possibility this could work for limited cases, but to use it broadly would be a bad idea. **If you base all your marketing activities on legitimate interests, your marketing strategy will be built on shaky ground**.

According to the GDPR, consent from a user is just one of many ways to process personal data; legitimate interest is another possible way. To use legitimate interests you have to show that those interests are "real and not too vague" and that the rights of individual data subjects don't outweigh the interests of the business. Data processing based on legitimate interests must also be stopped if any user raises an objection. If that sounds tenuous then you're starting to get the point.

A good example of legitimate interests might be keeping a users email address after opt-out to ensure they aren't sent any more emails. You keep a small piece of personal data to ensure that the users wishes are respected.

On the other hand, say a company tailors on-site content based on personal data, but without consent. They argue it's for the visitor's own good: users expect personalization from the service so they have a legitimate interest to offer it. We don't know if this argument will work, but let's admit it's twisted logic. If it's so obviously good for visitors, why not get their consent from the beginning? A privacy-conscious user will object before too long anyhow and the company will end up justifying their logic in court instead of in a friendly consent form.

Legitimate interests aren't an easy way out, they put more responsibility on the data collector (read ICO's wonderful guidelines and checklist to see why). It's better to ask for consent from the beginning. Users will feel respected and you won't have to worry about compliance troubles down the road.

Step 2: Understand data requirements

You have an analytics goal and you have the outlines of the data needed to meet it. Next you need to think about the details involved with collecting that data and what software requirements they imply.

Let's review each data type you might need:

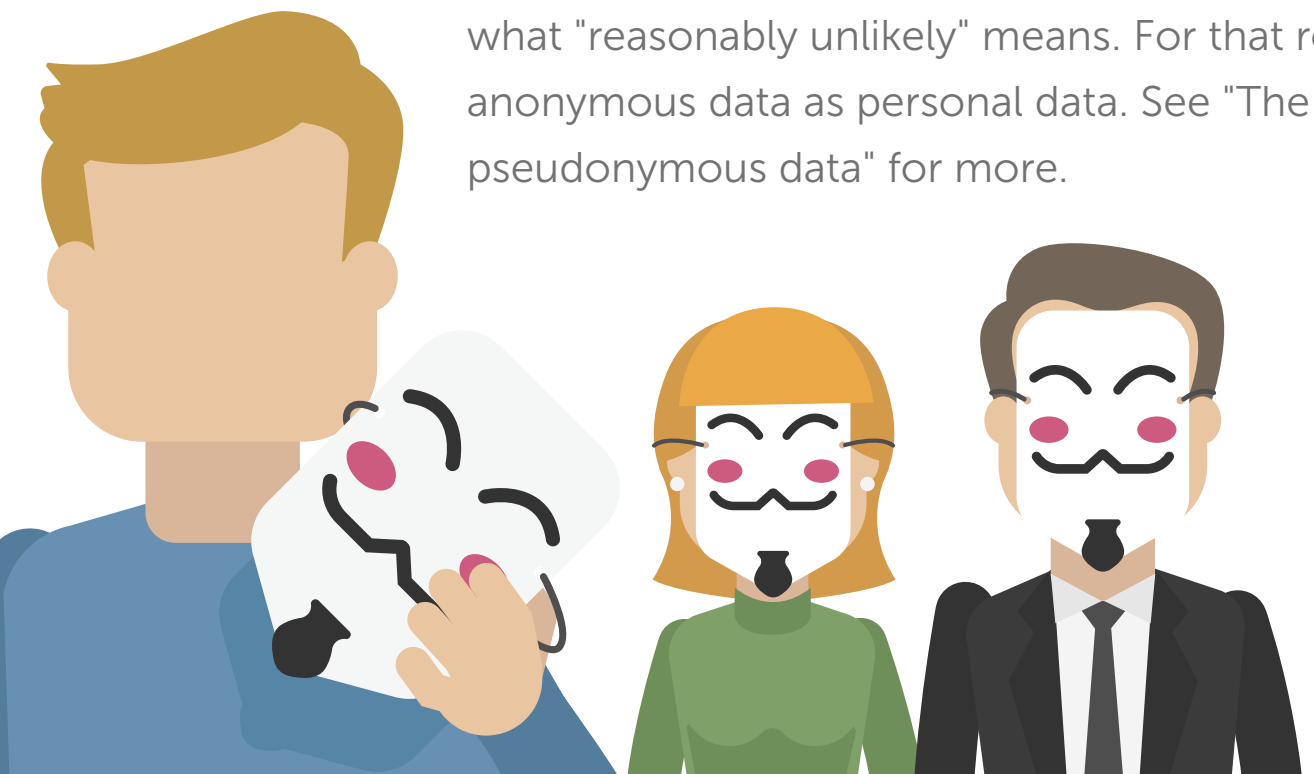
Anonymous data

You don't need consent, it's not considered personal data. But you do need to make sure the data is anonymous. Unfortunately that's not as easy as it might seem.

There are competing definitions of what makes data anonymous. A more conservative definition can be found in the Electronic Frontiers Foundation Do Not Track Policy. In short, nobody should be able to crunch your data and see the identities or activities for groups of less than 5000 individuals. However in GDPR Recital 26, the definition is “data rendered anonymous in such a way that the data subject is not or no longer identifiable”. They don't mention groups, just individuals.

To be compliant you'll have to show authorities that your anonymization method makes it "reasonably unlikely" that anyone could re-identify individuals from the data set.

The problem is that we don't yet know what definition will be enforced and what "reasonably unlikely" means. For that reason it's probably best to treat anonymous data as personal data. See "The dangers of anonymous and pseudonymous data" for more.



What is anonymous data?

Anonymous data (more specifically **de-identified data**) is rare in the world of online analytics and marketing. Most analysts want to tell a story about an internet visitor, which de-identified data can't provide.

For example, person X is from Paris and looked at winter boots, expensive scarves and wool pea coats. They can imagine a story to advertise or speak to person X more effectively. To do that though, they need an identifier such as a **cookie, device ID or fingerprint: all are personal data under GDPR**. For security and privacy the data may be made "pseudonymous", scrambled and protected by a separately stored key, but the individual stories stay encrypted in the data. Even in the pseudonymous form, it's still personal data.

Anonymous data, however, leaves no trace of those stories and no way to decipher them. It's less useful and an unnecessary extra step under more lax data regulations. Why de-identify data when something better is allowed? But post-GDPR **anonymous data is increasing in value because it doesn't require consent.**

You might rightfully ask: *How do I use it at all*? Say you're a government ministry hosting thousands of pages to keep the public informed about your services. You may not care about habits of individuals because you're not trying to sell or communicate to individuals on your website. You want to answer questions like: How many people visit each page? Do they read to the end? How long do they stay? What are popular search terms? The ministry can use answers to these questions, based only on anonymous data, to improve existing content and plan new content.

Certain industries like healthcare know these issues well and have developed effective methods for anonymizing data. There have been many books written on the subject like the well-written gem *Anonymizing Health Data*. Inevitably we'll borrow methods and lessons learned from the healthcare and medical fields for the purposes of online marketing and analytics.

Once we have a solid definition and enforcement standards under GDPR and ePrivacy, anonymous data will be much more useful: it won't require consent and won't be considered personal data.

Pseudonymous data

You need consent, it's still personal data, but the data can be used more flexibly than other forms of personal data. Pseudonymous data is different from anonymous data because there is a key to unlock the data, permitting the re-identification of individuals. It's like reversible anonymous data. But as long as that "data key" is stored separately from the pseudonymous data, it benefits from a special GDPR status allowing for purposes beyond those specified in the consent.

Why the special treatment?

Let's go through an example of a user unsubscribing from a mailing list. The data processor keeps the email address not for any marketing purpose, but to keep their promise of truly unsubscribing the user. The data processor then goes one step further and hashes the email address, turning it into pseudonymous data. Now they don't have the plain text email address anywhere in their system.

Later the marketing team uploads a list of email addresses from the CRM platform to your mailing software. The mailing software will check against unsubscribed addresses to make sure not to send to them. In this case, they can do so without unscrambling the list of hashed email addresses. They know that the hashing function will produce the same output when given the same input as before, that is the email address. They send the new email address through the hashing function to check if it's been unsubscribed. If that email address is in the list, they return to the visitor with a message saying: "Sorry, we can't sign up this email address, it was unsubscribed already". They can do so without ever storing the unsubscribed email address again in a human readable form.

So pseudonymous data is special for two reasons: the added security of the data being stored in an unreadable form, and the ability to make use of it without making it readable again. It's still personal data, we can't emphasize that enough, but it's secure enough to qualify for special status.

Should I use anonymous and pseudonymous data?

Anonymous and pseudonymous data seem great, right? The first doesn't require consent at all and the second gives possibilities beyond the original consent. But before we get too carried away we need to address a big unknown in the GDPR.

The EU Article 29 working group has stated that these data types lose their special status if they are "reasonably likely to be re-identified". In other words, if it is "reasonably likely" that a third-party can take your data and pick out individuals and their personal data, then you're in real trouble. You will have processed or used personal data without specific consent. It comes down to what "reasonably likely" means, and we don't yet know enough details to define it precisely.

Before we go further, let's address a couple myths about anonymous data:

Myth 1: Data can be anonymized to the point where nobody will ever, not even with most powerful supercomputer, re-identify the personal data.
There is always a risk of data being re-identified, good anonymization can only make that risk smaller and hopefully infinitesimal.

Myth 2: Data can never truly be anonymous. No matter what you do it, there will be a decent chance an average hacker can re-identify it.
All the stories about data being easily re-identified are about data that was not well anonymized in the first place. Anonymization is more than just removing names, national ID numbers and other direct identifiers.

The world would be simpler if one of these extremes were true. Unfortunately we have to live in the gray area. There are many available methods to anonymize data but there is no way to 100% eliminate the risk of re-identification. That said, it is possible to make that risk extremely low.

Regulations like GDPR recognize this, which is why they use the phrase "reasonably likely". But until we know what will meet that standard, it's probably smarter to rely on consent for both anonymous and pseudonymous data. It'll be safer to take full advantage of the potential of each only later, once we have more information.

Personal data for internal (first-party) use

You'll need to gather consent for each specific data purpose. Then you'll need to tie that consent status to how you gather data. For example, if a user consents to on-site personalization for product or content recommendations, then you need your consent manager to talk to your personalization engine. You'll have these connections in place for all analytics and marketing uses. This will be painful to do manually, an automated approach will be more feasible.

Personal data for external (second- or third-party) use

You'll need to mention this external use in the consent. GDPR will also require a data processing agreement (DPA) with the third party. This is also a good idea for your organization. You'll want a plan for processing data subject requests with third parties. For example, if a user wants their data removed and you shared data with a third party, it's your responsibility to work with that third party to delete it. We'll discuss data subject requests more later but you should also check out our article and infographic.

Sensitive data

You'll need consent and an approved exception to use "special categories of personal data" that is: "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation".

In many contexts it is prohibited to collect such data. Exceptions include situations where an individual decides to make such data public. For other industries, like healthcare, sensitive data is allowed because it's unavoidable. Even if you can claim an exception, make sure you really need sensitive data and put extra work into properly processing it. Besides the legal implications, you can be sure that mishandling or leaks of sensitive data will cause a more severe reaction than for other kinds of data.

Do I need to worry about ad blockers?

According to the 2017 PageFair Adblock Report, ad blocking is growing 30% year on year. Ad block programs can block web analytics and other kinds of marketing efforts that aren't advertising.

What to do? Make sure you're using a reputable analytics platform that uses first-party cookies. Beyond that though, you'll just have to accept that you'll lose some data to aggressive ad blockers that block even analytics based on first-party cookies.

The good news is that this represents only 5-10% of traffic. In addition many hope that GDPR and ePrivacy will increase user confidence in how data is handled online leading to a decrease in the use of such ad blockers.

What about browsers on mobile and desktop, like Apple's Intelligent Tracking Prevention, incorporating ad blocking by default? For the time being, there is no need to worry about such software blocking the analytics methods we're talking about here.

Step 3: Engage with visitors about data collection

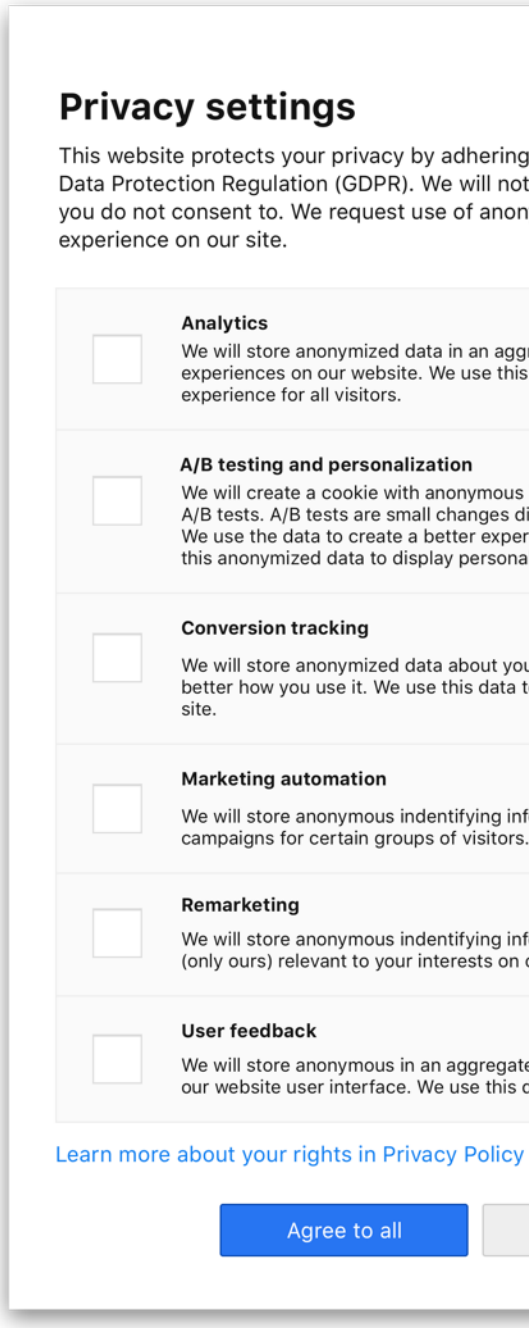
Link consent status to data collection

If your goal requires anything beyond anonymous data, you'll need user consent. If you don't know exactly how to ask for consent and what that means, check out our article and infographic on the topic.

Simply put, you'll have to mention each kind of data you'll be collecting and why (specific data purpose). It's up to you how you ask, as long as you explain everything in clear, everyday language. No confusing legalese is allowed.

Once you have the consent set up, you still need a few more things to complete the process:

- You need to store the consent record in a database somewhere and tie it to a user identifier (first-party cookie, device or advertiser ID, fingerprint).
- You need to tie that consent status to how the user is treated on your site. Collecting data, sending it to third parties and all associated tags, etc. need to agree with the specific purposes the user consented to.
- You need to have a system in place to resolve and pass on data subject requests (if the request involves communicating with a third party.



If you have analytics now, this means inserting a new gatekeeper in the middle of all those complex data connections. This will be easier for organizations just now building online analytics capability, but it's still no small task. Luckily many software vendors are starting to respond to the need with specialized software to perform these functions automatically. For example, Piwik PRO's Consent Manager interacts with all other parts of our platform, automatically ensuring consistent treatment of consents within the platform.

Will "Do Not Track" affect my analytics?

"Do Not Track" is an option available in browsers that can send an automatic message to websites: "I don't want to be tracked in any way, no first or third party cookies, fingerprinting, etc." It's essentially an automatic "no" to all consents. It's been estimated that 10-20% of visitors have Do Not Track enabled. What does this mean for you after GDPR?

This depends on how you interpret the Do Not Track browser setting. However, until we know more about GDPR enforcement it's safest to treat it as an **objection to data processing**. So you won't be able to collect personal data on anyone with it enabled. You also won't be able to ask for consent to override this setting. You can't deny them access based on this feature, all you can do is inform them of what affects it will have on how they use your site.

The good news is this is likely to get looser with time. Even the relatively rigid Electronic Frontier Foundation Do Not Track Policy does not recommend interpreting the setting so strictly - they allow for consent to override Do Not Track, for example. It's hard to imagine GDPR and ePrivacy being interpreted in a more limiting manner, but we don't have enough information right now.

Step 4: Maintain the trust of your users

Respond to changes and data subject requests

Consent changes

You already have current consents stored, so for this stage you need a place where visitors can see their consent status and change it. It can be a similar dialog to the one where they first gave consent.

Most importantly, it needs to be **easy to find and to understand**. If a user feels like you're hiding it, you'll lose some of that trust you gained by politely asking for consent.

Data subject requests

Allowing users to exercise their data subject rights will be a different story. **It won't be completely automated even in the best case scenario**.

You'll need a place for users to exercise those rights, explaining to them what each one means. This could be the same page as for managing consent or a separate dedicated page. A good place for this would be on the privacy policy page.

As soon as you receive any request you'll need a system for tracking request status. You'll have 30 days to fulfill requests so you'll need to stay on top of them. This part will look something like a task manager with status updates, reminders and assignment to a responsible person.

Finally you need to carry out the request. For example, if a user requests the deletion of all data you have on them, you need to do so whether that data is only in-house or was also sent to a third-party partner. You'll need to contact **all third parties** that had a hand in processing the data and make sure they follow through on the request.

This last part will be hard to automate. You'll need to think about how much manpower it will take to handle these requests. How many users will you likely be dealing with? What's a reasonable estimate for the amount of requests you'll get? At the beginning these will be rough estimates, to be replaced later when we have experience with GDPR.

What are data subject rights?

Data subject rights are a new set of individual rights that GDPR grants all residents in the European Union. They are intended to improve individual control over personal data and internet privacy. Worried? Don't be. They will require extra work from data controllers and processors but they are reasonable requests. On a positive note, they may increase consumer trust in online analytics and advertising, making for a more healthy relationship between internet users and advertisers.

We have a great article and infographic if you need a refresher.

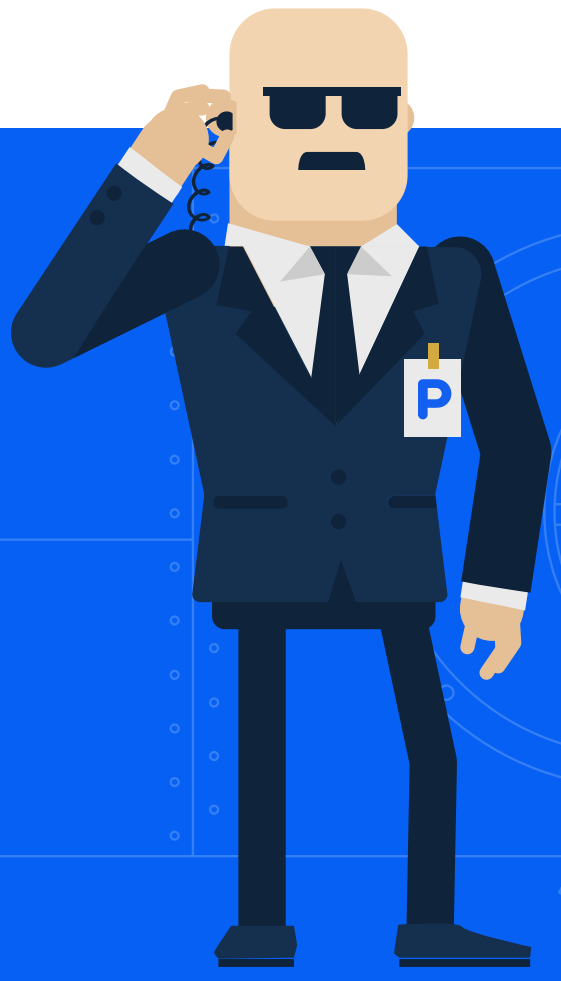
Conclusion: A virtuous cycle

Now that you've seen the full analysis cycle, you may want to think about that original analytics goal again. Ask yourself a couple simple questions:

- Can I get a similar result with less data?
- What features do I need from my analytics and marketing platforms to get all this done?

After going through this process for all your analytics goals, you'll have most of what you need for a Data Processing Impact Assessment (DPIA). You'll also know any holes that need to be filled in your toolset.

Piwik PRO has been developing our GDPR-compliant analytics and marketing suite for years. We've developed a Consent Manager to help deal with many of the issues mentioned above. Don't hesitate to get in touch. We'd be happy to share our expertise and show you what our products offer organizations of all kinds.



See how our integrated platform helps our clients stay on the safe side of web and mobile analytics.

Schedule a demo

About Piwik PRO

AdTech and MarTech experts founded Piwik PRO in 2013 due to the lack of an analytics stack that was both high performance and privacy-friendly. Our suite of products marries privacy by design, flexible hosting and full data ownership with enterprise-level features and support.

The Piwik PRO team consists of seasoned analytics experts and engineers who have advised on and delivered a wide range of successful implementations. Acting as your technology partner, we share our expertise, tailor our products and services to match your particular goals, and support you from start to finish.

Contact

NORTH AMERICA	DACH	BENELUX	EMEA
+1 (888) 444 0049	+49 2203 989 620	+31 858 881 458	+48 71 716 69 50

Web:
https://piwik.pro

Email:
sales@piwik.pro