Online identifiers in security logging & monitoring

# Online identifiers in security logging & monitoring

It can sometimes be counter-intuitive to think of IP addresses, cookies, and log data as personal data. However there are good reasons why it is important that we do so when we design, implement, and operate our online services. Put simply, it is easiest for us to assume that any information we capture and process from our public-facing services might contain personal information and protect this information accordingly.

## What are online identifiers?

Online identifiers are anything that could be used to track someone as they interact with our online services. This can include their IP address(es), any cookies that we (or 3rd parties we use) set on their devices, information placed into local storage on their device, username(s) associated with our services, and things like third-party authentication tokens etc. It could also include metadata captured about a device interacting with our services if this information is sufficiently different to allow devices to be reliably identified.

## Why are online identifiers treated as personal data?

If there is any way to tie an online identifier to an individual, then that identifier needs to be treated as though it is personal data. The way this mapping might be achieved is unimportant - it could be because the user later provides personal data to us as part of using a service (thus providing a link between all of the activities that their IP or session cookie has done with their identity), or if there is a legal route available to us to unmask the identity behind an identifier - such as by a lawful request to an ISP to uncover the person associated with a dynamic IP address at a particular time. For more information on this see the ICO [key definitions](https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/) and '[Recital 30](https://www.privacy-regulation.eu/en/recital-30-GDPR.htm)' from the [Article 29 Working Group](https://en.wikipedia.org/wiki/Article_29_Data_Protection_Working_Party). There is also an informative article [here](https://www.fieldfisher.com/en/services/privacy-security-and-information/privacy-security-and-information-law-blog/can-a-dynamic-ip-address-constitute-personal-data)).

## What does this mean for our services?

We need to think carefully about what metadata about a user's interaction with our service we are capturing, how long we are retaining that information, and who will have access to it. We need to be clear in our privacy notices on our services about the information we capture as part of a user's interaction with them - including 'anonymous' interactions, such as just browsing information about the services. Metadata like this must be included in the scope of privacy impact assessments for our services.

This only applies to our externally-facing services; it does not apply to our internal services, although it is undoubtedly good practice to apply the same approach.

## What does this mean for security logging and monitoring?

Under the updated data protection legislation we are still able to log and monitor the use of our services to help defend them against cyber security attacks, and misuse (such as fraud).

[Recital 49](https://www.privacy-regulation.eu/en/recital-49-GDPR.htm) notes that the processing of personal data (to the extent that is strictly necessary and proportionate) to ensure the security of a system which forms the underlying lawful basis for why the Ministry of Justice (MoJ) processes this type of data for this purpose. Thus we are still able to log and monitor external interactions with our services to look for evidence of cyber security attacks, and to enable us to act to protect those services - such as by blocking an IP address associated with known malware, or which is trying to perform a denial of service against us.

However we must be careful that we do not over-retain such log information, or share it with those who do not need to see it, without lawful justification. We must also ensure we act in a proportionate way with this data.

The MoJ CISO is ultimately responsible for all logging and monitoring systems which have been implemented for cyber security purposes, and as such is the Information Asset Owner for all logging and monitoring data.

By default we will retain raw logs in direct relation to security logging and monitoring purposes for at least 90 days and a maximum of 2 years. The variation in between is as defined and required by legislation, regulation (such as the Law Enforcement Directive) or certification compliance (such as [PCI-DSS](https://en.wikipedia.org/wiki/Payment_Card_Industry_Data_Security_Standard)). Retention for periods longer than 2 years requires MoJ CISO approval.

Aggregate data from logging systems (such as number of particular types of events, total numbers of visits to sites, etc) can be retained indefinitely, so long as care has been taken to remove potentially unique or identifying information from the retained information set.

## Protecting log files and log data

Default permissions must be set on logging and monitoring systems such that only ops staff for that service and the MoJ's security operations team have access to the data in them. All access to the raw logging and monitoring data must also be logged.

Bulk exporting from such logging systems is prohibited by default as sensitive logs should be analysed programmatically in-situ. Bulk exporting should be prevented by default technical/access controls where possible. If a bulk extract from a logging system is required (for example, into a more complex analytical system or in a wider migration) then this requires the approval of the MoJ CISO.