

Chap asks Facebook for data on his web activity, Facebook says no, now watchdog's on the case

Info collected on folk outside the social network 'not readily accessible'

By [Rebecca Hill](#) 114  [Reg comments](#) [SHARE](#) ▼

 Facebook-style crying emoji

Facebook's refusal to hand over the data it holds on users' web activity is to be probed by the Irish Data Protection Commissioner after a complaint from a UK-based academic.

Under the [General Data Protection Regulation](#), which came into force [on 25 May](#), people can demand that organisations hand over the data they hold on them.

Although a similar right existed in the UK before, crucially, it's now free to make these subject access requests (SAR) – and so many people decided to test the law.

Unsurprisingly, Facebook was a prime target, but its responses have failed to impress.

The crux of the issue is the data the firm slurps up via its [Facebook Pixel](#), the widely used tracking code on multiple websites and the subject of much debate during the heat of the Cambridge Analytica scandal.

Because, although the Zuckerberg offers people a way to access the data collected on the platform – for instance, ad preferences – these tools don't provide the information collected off it.

Michael Veale, who works at University College London, submitted a SAR to the social media giant on 25 May asking it to hand over the information it has collected on his browsing behaviour and activities off Facebook.

However, the firm declined to do so, effectively saying it was too difficult to locate the info within its humongous data warehouse.

Veale argued that this is unsatisfactory because – as it could be used to infer religion, medical history or sexuality – it is highly personal and sensitive data, and so made a formal complaint to the Irish Data Protection Commissioner (Facebook's European HQ is in Ireland).

In his complaint – shared with *The Register* – Veale said that he wanted to know whether Facebook has web history on him in medical domains and his sexuality.

"Both of these concerns have been triggered and exacerbated by the way in which the Facebook platform targets adverts in highly granular ways, and I wish to understand fair processing," he said.

Veale added that he had used the public tools Facebook offers, but that they had proved "insufficient".

The Irish DPC has now opened a statutory inquiry into the matter, telling Veale that it anticipated the case will be referred to the European Union's brain trust, the European Data Protection Board, as it involves cross-border processing.

"I hope to refute emerging arguments that the data processing operations of big platforms relating to tracking are too big or complex to regulate," Veale told *EI Reg*.

"By choosing to give user-friendly information (like ad interests) instead of the raw tracking data, it has the effect of disguising some of its creepiest practices. It's also hard to tell how well ad or tracker blockers work without this kind of data."

Getting into Facebook's Hive mind

Facebook slurps information about your device, the websites you visited, apps you used and ads you've seen via Facebook business tools and plug-ins, such as the Like button, on partner sites.

This is stored alongside an identifier for that person, whether you have an account or not, and whether you're logged in or not.

In a "Hard Questions" [blog post](#) in the aftermath of Mark Zuckerberg's [awkward testimony](#) in the US, Facebook said this information was used for safety and security, and to improve both its own and its partners' services.

But – as revealed earlier this year in an emailed response to activist Paul Olivier Dehaye [shared with with the House of Commons digital committee](#) – the firm said it can't share this with users.

The Social Network said the information was stored in a Hive data warehouse, which was "primarily for backup purposes and data analytics", noting that this kind of architecture was necessary due to the sheer volume of data created.

Data stored in Hive is kept separate from the relational databases that power the Facebook site, it said, and is primarily organised by hour, in log format.

However, Facebook said the information in Hive "is not readily accessible" as it isn't stored on a per user basis – rather it is log data stored in tables split into partitions.

Because it isn't indexed by user, in order to extract a user's data from Hive, each partition would need to be searched for all possible dates in order to find any entries relating to a particular user's ID.

"Facebook simply does not have the infrastructure capacity to store log data in Hive in a form that is indexed by user in the way that it can for production data used for the main Facebook site," Zuck's minions said.

'Staggeringly sensitive' info should be shared

Privacy campaigners have little time for this argument. As Veale noted in his complaint to the Irish DPC, this is "very clearly personal data".

Indeed, as anyone who has decided to clear their browsing history will know, a manual scan can elicit a fair amount of detail – so the application of machine learning over millions of users could be used to distinguish more nuanced patterns.

"Web browsing history is staggeringly sensitive," Veale said, pointing out it can be used to infer information on sexuality, purchasing habits, health information or political leanings.

He added that, even if it wasn't stored alongside a user ID, research has shown it is possible to re-identify web browsing histories to individual data subjects using only publicly available data.

"Any balancing test, such as legitimate interests, must recognise that this data is among the most intrusive data that can be collected on individuals in the 21st century," Veale said.

Moreover, Veale argued that this information – which will indicate which organisations hold data on them – forms a crucial piece of the jigsaw for people who want to understand who has access to their data and how it is used.

"This is a critical transparency tool to ensure legality of a complex data chain involving millions of organisations," Veale said, pointing out that Facebook has 2.2 million active installations of trackers.

'Don't blame the data subject for your data warehouse'

Veale also took issue with the claims made in Facebook's refusal – which also came a month after the deadline imposed on organisations under the GDPR.

For instance, it cited Article 12(5), which relates to requests that are "manifestly unfounded or excessive", in particular because of their repetitive nature.

But Veale has never made the request before and argued that the sensitivity of the data means it isn't manifestly unfounded.

Moreover, he pointed out that if the request is excessive, it is only because the amount of data collected and sent to Facebook is too large for one of the biggest companies in the world to retrieve.

"Which seems to be a breach of [GDPR's requirement for] data minimisation rather than my fault as a data subject requesting this data," he observed.

In response, the DPC said it had initiated a formal statutory inquiry about the complaint, which will examine whether Facebook has properly met its obligations and whether its response had contravened the GDPR.

The DPC confirmed to *The Register* that the inquiry had been initiated, but neither it nor the EDPB could comment further on an open inquiry.