European Law Blog

Search queries and anonymisation: How to read Article 6(11) of the DMA and the GDPR together?

Sophie Stalla-Bourdillon Bárbara da Rosa Lazarotto

European Law Blog

Published on: Apr 03, 2024

URL: https://europeanlawblog.pubpub.org/pub/2uxr4anu

The Digital Markets Act (DMA) is a regulation enacted by the European Union as part of the European Strategy for Data. Its final text was published on 12 October 2022, and it officially entered into force on 1 November 2022. The main objective of the DMA is to regulate the digital market by imposing a series of by-design obligations (see Recital 65) on large digital platforms, designated as "gatekeepers". Under to the DMA, the European Commission is responsible for designating the companies that are considered to be gatekeepers (e.g., Alphabet, Amazon, Apple, ByteDance, Meta, Microsoft). After the Commission's designation on 6 September 2023, as per DMA Article 3, a six-month period of compliance followed and ended on 6 March 2024. At the time of writing, gatekeepers are thus expected to have made the necessary adjustments to comply with the DMA.

Gatekeepers' obligations are set forth in Articles 5, 6, and 7 of the DMA, stemming from a variety of data-sharing and data portability duties. The DMA is just one pillar of the European Strategy for Data, and as such shall complement the General Data Protection Regulation (see Article 8(1) DMA), although it is not necessarily clear, at least at first glance, how the DMA and the GDPR can be combined together. This is why the main objective of this blog post is to analyse Article 6 DMA, exploring its effects and thereby its interplay with the GDPR. Article 6 DMA is particularly interesting when exploring the interplay between the DMA and the GDPR, as it forces gatekeepers to bring the covered personal data outside the domain of the GDPR through anonymisation to enable its sharing with competitors. Yet, the EU standard for legal anonymisation is still hotly debated, as illustrated by the recent case of SRB v EDPS now under appeal before the Court of Justice.

This blog is structured as follows: First, we present Article 6(11) and its underlying rationale. Second, we raise a set of questions related to how Article 6(11) should be interpreted in the light of the GDPR.

Article 6(11) DMA provides that:

> *"The gatekeeper shall provide to any third-party undertaking providing online search engines, at its request, with access on fair, reasonable and non-discriminatory terms to ranking, query, click and view data in relation to free and paid search generated by end users on its online search engines. Any such query, click and view data that constitutes personal data shall be anonymised."*

It thus includes two obligations: an obligation to share data with third parties and an obligation to anonymise covered data, i.e. *"ranking, query, click and view data,"* for the purpose of sharing.

The rationale for such a provision is given in Recital 61: to make sure that third-party undertakings providing online search engines *"can optimise their services and contest the relevant core platform services."* Recital 61 indeed observes that *"Access by gatekeepers to such ranking, query, click and view data constitutes an important barrier to entry and expansion, which undermines the contestability of online search engines."*

Article 6(11) obligations thus aim to address the asymmetry of information that exist between search engines acting as gatekeepers and other search engines, with the intention to feed fairer competition. The intimate relationship between Article 6(11) and competition law concerns is also visible in the requirement that gatekeepers must give other search engines access to covered data *"on fair, reasonable and non-discriminatory terms."*

Article 6(11) should be read together with Article 2 DMA, which includes a few definitions.

1. Ranking: *"the relevance given to search results by online search engines, as presented, organised or communicated by the (…) online search engines, irrespective of the technological means used for such presentation, organisation or communication and irrespective of whether only one result is presented or communicated;"*
2. Search results: *"any information in any format, including textual, graphic, vocal or other outputs, returned in response to, and related to, a search query, irrespective of whether the information returned is a paid or an unpaid result, a direct answer or any product, service or information offered in connection with the organic results, or displayed along with or partly or entirely embedded in them;"*

There is no definition of search queries, although they are usually understood as being strings of characters (usually key words or even full sentences) entered by search-engine users to obtain relevant information, i.e., search results.

As mentioned above, Article 6 (11) imposes upon gatekeepers an obligation to anonymise covered data for the purposes of sharing it with third parties. A (non-binding) definition of anonymisation can be found in Recital 61: *"The relevant data is anonymised if personal data is irreversibly altered in such a way that information does not relate to an identified or identifiable natural person or where personal data is rendered anonymous in such a manner that the data subject is not or is no longer identifiable."* This definition echoes Recital 26 of the GDPR, although it innovates by introducing the concept of irreversibility. This introduction is not surprising as the concept of (ir)reversibility appeared in old and recent guidance on anonymisation (see e.g., Article 29 Working Party Opinion on Anonymisation Technique 2014, the EDPS and AEPD guidance on anonymisation). It may be problematic, however, as it seems to suggest that it is possible to achieve absolute irreversibility; in other words, that it is possible to guarantee an impossibility to link the information back to the individual. Unfortunately, irreversibility is *always conditional* upon a set of assumptions, which vary depending on the data environment: in other words, it is always relative. A better formulation of the anonymisation test can be found in section 23 of the Quebec Act respecting the protection of personal information in the private sector: the test for anonymisation is met when it is *"at all times,* ***reasonably foreseeable in the circumstances*** *that* [information concerning a natural person] *irreversibly no longer allows the person to be identified directly or indirectly."* [emphasis added].

Recital 61 of the DMA is also concerned about the utility third-party search engines would be able to derive from the shared data and therefore adds that gatekeepers *"should ensure the protection of personal data of end users, including against possible re-identification risks, by appropriate means, such as anonymisation of such personal data,* ***without substantially degrading the quality or usefulness of the data****"*. [emphasis added]. It is however challenging to reconcile a restrictive approach to anonymisation with the need to preserve utility for the data recipients.

One way to make sense of Recital 61 is to suggest that its drafters may have equated aggregated data with non-personal data (defined as *"data other than personal data"*). Recital 61 states that *"Undertakings providing online search engines collect and store aggregated datasets containing information about what users searched for, and how they interacted with, the results with which they were provided."* Bias in favour of aggregates is indeed persistent in the law and policymaker community, as illustrated by the formulation used in the adequacy decision for the EU-US Data Privacy Framework, in which the European Commission writes that *"[s]tatistical reporting relying on aggregate employment data and containing no personal data or the use of anonymized data does not raise privacy concerns"*. Yet, such a position makes it difficult to derive a coherent anonymisation standard.

Generating a means or a count does not necessarily imply that data subjects are no longer identifiable. Aggregation is not a synonym for anonymisation, which explains why differentially-private methods have been developed. This brings us back to when AOL released 20 million web queries from 650,000 AOL users, relying on basic masking techniques applied to individual-level data to reduce re-identification risks. Aggregation alone will not be able to solve the AOL (or Netflix) challenge.

When read in the light of the GDPR and its interpretative guidance, Article 6(11) DMA raises several questions. We unpack a few sets of questions that relate to anonymisation and briefly mention others.

The first set of questions relates to the anonymisation techniques gatekeepers could implement to comply with Article 6(11). At least three anonymisation techniques are potentially in scope for complying with Article 6(11):

- global differential privacy (GDP): "GDP is a technique employing randomisation in the computation of aggregate statistics. GDP offers a mathematical guarantee against identity, attribute, participation, and relational inferences and is achieved for any desired 'privacy loss'." (See here)
- local differential privacy (LDS): "LDP is a data randomisation method that randomises sensitive values [within individual records]. LDP offers a mathematical guarantee against attribute inference and is achieved for any desired 'privacy loss'." (see here)
- k-anonymisation: is a generalisation technique, which organises individuals records into groups so that records within the same cohort made of k records share the same quasi-identifiers (see here).

These techniques perform differently depending upon the re-identification risk at stake. For a comparison of these techniques see here. Note that synthetic data, which is often included within the list of privacy-enhancing technologies (PETs), is simply the product of a model that is trained to reproduce the characteristics and structure of the original data with no guarantee that the generative model cannot memorise the training data. Synthetisation could be combined with differentially-private methods however.

- Could it be that *only* global differential privacy meets Article 6(11)'s test as it offers, at least in theory, a formal guarantee that aggregates are safe? But what would such a solution imply in terms of utility?
- Or could gatekeepers meet Article 6 (11)'s test by applying *both* local differential privacy and k-anonymisation techniques to protect sensitive attributes and make sure individuals are not singled out? But again, what would such a solution mean in terms of utility?
- Or could it be that k-anonymisation following the redaction of manifestly identifying data will be enough to meet Article 6(11)'s test? What *does it really mean* to apply k-anonymisation on ranking, query, click and view data? Should we draw a distinction between queries made by signed-in users and queries made by incognito users?

Interestingly, the 2014 WP29 opinion makes it clear that k-anonymisation *is not able to mitigate* on its own the three re-identification risks listed as relevant in the opinion, i.e., singling out, linkability and inference: k-anonymisation is not able to address inference and (not fully) linkability risks. Assuming k-anonymisation is endorsed by the EU regulator, could it be the confirmation that a risk-based approach to anonymisation could ignore inference and linkability risks? As a side note, the UK Information Commissioner's Office (ICO) in 2012 was of the opinion that pseudonymisation could lead to anonymisation, which implied that mitigating for singling out was not conceived as a necessary condition for anonymisation. The more recent guidance, however, doesn't directly address this point.

The second set of questions Article 6(11) poses is related to the overall legal anonymisation standard. To effectively reduce re-identification risks to an acceptable level, all anonymisation techniques need to be coupled with context controls, which usually take the form of security techniques such as access control and/or organisational and legal measures, such as data sharing agreements.

- What types of context controls should gatekeepers put in place? Could they set eligibility conditions and require that third-party search engines evidence trustworthiness or commit to complying with certain data protection-related requirements?
- Wouldn't this strengthen the gatekeeper's status though?

It is important to emphasise in this regard that although legal anonymisation might be deemed to be achieved at some point in time in the hands of third-party search engines, the anonymisation process remains governed by data protection law. Moreover, anonymisation is only a data handling process: it is not a purpose, and it is not a legal basis, therefore purpose limitation and lawfulness should be achieved independently. What is more, it

should be clear that even if Article 6(11) covered data can be considered legally anonymised in the hands of third-party search engines once controls have been placed on the data and its environment, these entities should be subject to an obligation not to undermine the anonymisation process.

Going further, the 2014 WP29 opinion states that "*it is critical to understand that when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data.*" This sentence, however, now seems outdated. While in 2014 Article 29 Working Party was of the view that the input data had to be destroyed to claim legal anonymisation of the output data, Article 6(11) nor Recital 61 suggest that the gatekeepers would need to delete the input search queries to be able to share the output queries with third parties.

The third set of questions Article 6(11) poses relates to the modalities of the access: What does Article 6(11) imply when it comes to access to data, should it be granted in real-time or after the facts, at regular intervals?

The fourth set of questions Article 6(11) poses relates to pricing. What do fair, reasonable and non-discriminatory terms mean in practice? What is gatekeepers' leeway?

To conclude, the DMA could signal a shift in the EU approach to anonymisation or maybe just help pierce the veil that was covering anonymisation practices. The DMA is actually not the only piece of legislation that refers to anonymisation as a data-sharing safeguard. The Data Act and other EU proposals in the legislative pipeline seem to suggest that legal anonymisation can be achieved, even when the data at stake is potentially very sensitive, such as health data. A better approach would have been to start by developing a consistent approach to anonymisation relying by default upon both data and context controls and by making it clear that, as anonymisation is always a trade-off that inevitably prioritises utility over confidentiality; therefore, the legitimacy of the processing purpose that will be pursued once the data is anonymised should always be a necessary condition to an anonymisation claim. Interestingly, the Act respecting the protection of personal information in the private sector mentioned above makes purpose legitimacy a condition for anonymisation (see section 23 mentioned above). In addition, the level of data subject intervenability preserved by the anonymisation process should also be taken into account when assessing the anonymisation process, as suggested here. What is more, explicit justifications for prioritising certain re-identification risks (e.g., singling out) over others (e.g., inference, linkability) and assumptions related to relevant threat models should be made explicit to facilitate oversight, as suggested here as well.

To end this post, as anonymisation remains a process governed by data protection law, data subjects should be properly informed and, at least, be able to object. Yet, by multiplying legal obligations to share and anonymise, the right to object is likely to be undermined without the introduction of special requirements to this effect.