The Symbiosis between Data Protection and Open Data

A primer on how to reconcile EU data protection law with open data policies

data.europa.eu
The official portal for European data

This study has been prepared as part of data.europa.eu, an initiative of the European Commission. The Publications Office of the European Union is responsible for the management of data.europa.eu contracts.

For more information about this paper, please contact the following.

European Commission
Directorate-General for Communications Networks, Content and Technology
Unit G.1 Data Policy and Innovation
Email: CNECT-G1@ec.europa.eu

data.europa.eu
Email: info@data.europa.eu

Author:
Hans Graux

Last update: 31 July 2024
https://data.europa.eu/

DISCLAIMER
The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use that may be made of the information contained herein.

Luxembourg: Publications Office of the European Union, 2024

ISBN: 978-92-78-44365-8 doi: 10.2830/1874306 Catalogue number: OA-01-24-039-EN-Q

Contents

1. The symbiosis between data protection and open data

1.1. Introduction

Within the European Union (EU), the protection of personal data is recognised as a fundamental right under the Charter of Fundamental Rights of the European Union[1]. Broadly speaking, this implies that personal data – meaning any data that can be linked to a specific natural person, as will be more extensively discussed below – must be handled with appropriate care. As Article 8 of the charter notes, personal data 'must be processed fairly, for specified purposes, and on the basis of the consent of the person concerned or some other legitimate basis laid down by law'.

This high-level description in the charter of the right to the protection of personal data ('the right to data protection') has been elaborated on in a number of legal texts, most significantly the 2016 general data protection regulation (GDPR[2]), which succeeded the 1995 data protection directive[3]. The GDPR outlines the principal rules and requirements that apply to most personal data processing activities in the EU, including where personal data is collected, shared or reused. For instance, the GDPR contains obligations with respect to transparency (persons should be made aware of processing activities relating to their personal data), purpose limitation (processing activities must be limited to what was communicated to the affected persons) and data minimisation (personal data processing must be limited to what is strictly necessary to achieve the purposes communicated to those persons).

Within the EU, legislation on open data emerged more or less in parallel with the EU's data protection framework. The first public sector information directive (the PSI directive[4]) was adopted in 2003. It did not yet use the notion of 'open data' but it did provide a framework that encouraged (but did not yet require) public sector bodies to make their data available for reuse, for commercial or non-commercial purposes, under non-discriminatory and fair terms.

[1] European Parliament, Council of the European Union and European Commission, Charter of Fundamental Rights of the European Union, Publications Office of the European Union, Luxembourg, 2012 (OJ C 326, 26.10.2012, pp. 391–407, ELI: http://data.europa.eu/eli/treaty/char_2012/oj)

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (OJ L 119, 4.5.2016, pp. 1–88, ELI: http://data.europa.eu/eli/reg/2016/679/oj)

[3] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (OJ L 281, 23.11.1995, pp. 31–50, http://data.europa.eu/eli/dir/1995/46/oj)

[4] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information (OJ L 345, 31.12.2003, pp. 90–96, http://data.europa.eu/eli/dir/2003/98/oj)

The PSI directive was amended in 2013[5], in a revision that made reuse mandatory as a general rule, subject to a few exceptions. This amendment also provided the first reference to open data policies in its recitals, describing them as policies that 'encourage the wide availability and re-use of public sector information for private or commercial purposes, with minimal or no legal, technical or financial constraints, and which promote the circulation of information not only for economic operators but also for the public'.

The PSI directive was replaced in 2019 by the open data directive[6], which describes open data as 'data in an open format that can be freely used, re-used and shared by anyone for any purpose'. The open data directive further strengthened the obligation of public sector bodies to make their documents available as open data, wherever possible.

The legal frameworks relating to open data and data protection can apply cumulatively, namely when a document or dataset must be made available as open data, but simultaneously contains personal data. In those instances, it can be difficult for public sector bodies to find the right way to reconcile the two sets of obligations. Open data is fundamentally about making data freely available for reuse, whereas data protection focuses on implementing appropriate controls.

Thus, if a dataset contains personal data – which can happen, as we will explain below, but is certainly not the general trend – a data provider may be reluctant to make it available as open data. After all, if there is indeed personal data in the dataset, the GDPR would usually apply, and both the data provider and the data reuser would need to take measures to ensure their compliance with the regulation. They would need to provide transparency on the use of the personal data, clearly define the purpose of the permitted use and ensure that the shared data is kept as minimal as possible.

In practice, this can be challenging. Transparency towards the persons that can be linked to the personal data (the 'data subjects' in the GDPR) on the reuse of their personal data may be difficult to ensure, since there is usually no contact data available that allows communication with them. Implementing the purpose limitation and data minimisation principles is equally complex for data providers, since the act of sharing personal data with a reuser also constitutes a form of personal data processing that must comply with the GDPR. As a result, the data provider and the data reuser would need to conclude agreements on what kind of use of the personal data is permitted and what personal data is strictly necessary to achieve that goal. This is not a particularly scalable approach and is somewhat at odds with open data policies.

As a result, the principal GDPR compliance strategy is to avoid the inclusion of personal data in datasets that are made available as open data. This can be done either by not making datasets containing personal data available at all, or by anonymising the personal data, so that the dataset no longer contains personal data and the GDPR no longer applies. Of course, either one of these strategies

[5] Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information (OJ L 175, 27.6.2013, p. 1–8, http://data.europa.eu/eli/dir/2013/37/oj)

[6] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (OJ L 172, 26/06/2019, p. 56–83, http://data.europa.eu/eli/dir/2019/1024/oj)

requires the possibility to determine what exactly qualifies as personal data, with a reasonable degree of accuracy – i.e. to determine whether a dataset is linkable to a natural person. Moreover, it must be possible to determine who is responsible when mistakes are made.

1.2. Formal problem statement and structure of this research report

This research report aims to examine how recent discussions with respect to EU data protection law can affect open data sharing practices, specifically in relation to two fundamental concepts: personal data and data controllership.

- **Personal data** is any information relating to an identified or identifiable natural person. In contrast, information is **anonymous** when it does not relate to an identified or identifiable natural person. Anonymous data includes data that never related to a natural person (i.e. data that was never personal data), along with data that was originally personal data, but which has been rendered anonymous[7] in such a manner that the natural person is no longer identifiable.

  The concepts of personal data and anonymous data are important for EU data protection law because they **determine whether the GDPR applies**: the processing of personal data (e.g. by collecting it, exchanging it, analysing it, enriching it, using it in an application or service, and so forth) must generally respect the requirements of the GDPR. The processing of anonymous data, on the other hand, falls outside the scope of the GDPR. For that reason, effective anonymisation (e.g. by aggregating datasets or by removing data from the dataset that allows identification) is a key strategy to convert a dataset with personal data into open data. Whether or not anonymisation is effective, however, is an assessment that can change over time, as will be explained further in this paper: new technological developments can, in exceptional circumstances, allow a dataset that was previously anonymous to become linkable to individuals.

- A **data controller** under the GDPR is the entity (e.g. a company or a public sector body) that decides on the purposes and means of data processing – i.e. what the personal data will be used for and what tools or methods will be used to process it. A public sector body is, for example, a data controller when making its personal data available to a reuser, or when anonymising the dataset; and a reuser is a data controller when using a dataset containing personal data for its own purposes.

  The concept of a data controller is important for EU data protection law, because the **data controller bears the bulk of the legal responsibilities** under the GDPR.

[7] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, 10 April 2014, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

On both of these topics – on how to draw the line between personal data and anonymous data, and on the notions of responsibility and controllership – new case-law has recently emerged at the EU level that affects how data providers should look at their data protection compliance risks.

There is little guidance on how this new case-law could affect open data providers. If there is a possibility that their non-personal open data requalifies as personal data, or that they might be held responsible as a data controller for future reuse of the data after making it available, this can act as a disincentive for making data available. This paper will explore whether this concern is realistic, and will identify any best practices for open data providers to mitigate the risk.

The legal research paper thus aims to examine how open data providers can consider data protection compliance with respect to their datasets. Can they be sure that a dataset is free from personal data, and what are their responsibilities? We will explore what factors affect the answer to these questions, based on existing guidance from data protection authorities and on the case-law of EU courts.

2. Personal data and anonymisation

2.1. Introduction of the central concepts

The central building block of EU data protection law is the notion of personal data, defined in the GDPR as 'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'.

In many cases, the concept is fairly intuitive: a person's name, physical address, phone number and email address all clearly 'relate' to an identifiable natural person, and thus constitute personal data. However, there are many situations where there is a margin for discussion. Does, for example, a customer identification number that is used only by a specific store qualify as personal data? What about the licence plate number of a vehicle? These questions require an interpretation of the law: what does it mean for data to 'relate' to a person, when is that person 'identifiable' and does it matter who is capable of identifying the person?

The GDPR itself provides some guidance on the topic. Recital 26 indicates that 'to determine whether a natural person is identifiable, **account should be taken of all the means reasonably likely to be used**, such as singling out, **either by the controller or by another person** to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, **account should be taken of all objective factors**, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. **The principles of data protection should therefore not apply to anonymous information**, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes' (emphasis added).

The general test for determining whether information is identifiable is thus whether there are means 'reasonably likely to be used', by any person. Applying this rule to the example of vehicle licence plate numbers, the outcome would be that these constitute personal data (assuming that some of the vehicles can be linked to natural persons and not exclusively to companies), since there is an entity (namely the administration issuing the numbers) that can fairly easily look up who the affected natural persons are. The fact that these persons might not be driving the vehicle is irrelevant: the GDPR requires that the information relates to an identified or identifiable natural person; not that the same natural person is involved in a specific activity.

A qualification as personal data does not imply that the dataset cannot be lawfully shared or reused. Health data, for instance, is a category of personal data whenever a patient is identifiable. Moreover, it is a category for which specific safeguards apply under the GDPR, due to its inherent sensitivity. Nonetheless, the processing of health data is permitted under the conditions set out in the GDPR, which even contains specific exemptions in the context of scientific research. In this manner, the GDPR tries to strike a sound balance between protecting the privacy of patients and ensuring that healthcare research can build on the available data.

There is fairly detailed guidance[8] available on the notion of personal data, issued by the European data protection authorities, on how exactly to apply the test of identifiability, including a number of examples. The guidance stresses that a mere hypothetical possibility to single out the individual is not enough to consider the person as 'identifiable'. The criterion of 'all the means likely reasonably to be used' must be realistic and consider factors such as the means available to link data together, but also the way the processing is structured and the advantage expected from identifying the persons involved.

In the United Kingdom (where data protection law is based on and equivalent[9] to the GDPR), the test of the 'motivated intruder' is used for this purpose. A motivated intruder is described as 'a person who starts without any prior knowledge but wishes to identify an individual from whose personal data the anonymous information is derived. The test assesses whether the motivated intruder is likely to be successful. It assumes that a motivated intruder is someone that:

- is reasonably competent;
- has access to appropriate resources (e.g. the internet, libraries, public documents); and
- uses investigative techniques (e.g. making enquiries of people who may have additional knowledge about an individual, or advertising for anyone with that knowledge to come forward).

The intruder is therefore someone who has the:

- motives to attempt identification;
- means to succeed; and

[8] Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data, 20 June 2007, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[9] Commission Implementing Decision (EU) 2021/1772 of 28 June 2021 pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the adequate protection of personal data by the United Kingdom (C/2021/4800, OJ L 360, 11.10.2021, p. 1–68, http://data.europa.eu/eli/dec_impl/2021/1772/oj)

- intent to use the data in ways that may pose risks to your organisation and the rights and freedoms of individuals whose data you process'.

If such a motivated intruder would be capable of linking the data to a natural person, that data will be considered personal data and EU data protection law would need to be complied with.

2.2. Recent case-law on personal data and anonymous data

Despite the available guidance, there are still frequently cases presented to courts at the EU level where the notion of personal data must be interpreted. Hereunder, we will examine a particularly relevant example, Case C-319/22 of 9 November 2023[10], commonly known as the **Scania case**. Under EU law, vehicle manufacturer Scania was required to make certain information on their vehicles available to repairers and to independent operators, to allow maintenance. They did so via a website that allowed individual vehicles to be looked up based on their vehicle identification number (VIN). The VIN of a vehicle and its technical characteristics could not be linked by Scania to a natural person as Scania had no records of who owns or uses these vehicles. Thus, for Scania this dataset did not contain personal data.

In its ruling, the Court of Justice noted that, while the VIN and vehicle data held by Scania were indeed not inherently personal data, they become personal data if the person who holds a piece of the data can reasonably associate it with a specific natural person. This is the case when that person also has access to the registration certificate of a vehicle, which contains the VIN along with the name and address of the holder of that certificate (who may be a natural person). That was indeed the case for the independent operators, who would thus have the means to link a VIN to an identified or identifiable natural person.

The case is noteworthy because it reinforces that the assessment of what constitutes personal data must be done in the context of the processing activities and the means reasonably available to the entities that are processing the data. The dataset held by Scania contained purely vehicle information and did not constitute personal data as long as it was held only by Scania. However, once shared with independent operators who have additional information at their disposal, that dataset becomes personal data subject to the GDPR.

The case is thus an example of the dynamic nature of the assessment: whether a dataset constitutes personal data cannot always be determined merely by examining the data as such. It requires a consideration of the broader processing context, including additional resources that the data might be reasonably combined with in the course of its foreseeable use.

[10] Judgment of the Court of Justice of 9 November 2023, Gesamtverband Autoteile-Handel eV v Scania CV AB (Scania), C-319/22, EU:C:2023:837

2.3. Relevance and impact for open data providers

2.3.1. Bad news? Not necessarily

The Scania case is relevant for open data providers too, because it relates to the dynamic nature of the assessment of personal data: whether a specific dataset contains personal data or not, to some extent depends on the processing context. One of the key elements in the case was the question of whether the recipient of a dataset would be able to use additional data sources to turn non-personal data into personal data. Scania's dataset didn't inherently contain personal data, but given the ability of reusers to combine it with other data sources, it did need to be qualified as personal data when shared.

At first sight, this dynamic nature of the personal data assessment might appear to be bad news for providers of open data. After all, given the unbounded nature of open data, which is by definition shared with as few constraints as possible, how could a data provider possibly keep track of what additional data sources a recipient might have at their disposal, and whether they might be usable to turn an uncontroversial open dataset that contains no personal data into a GDPR compliance risk?

However, reviewing these decisions of the EU courts carefully, the outcome does not seem so negative. The courts consistently apply the criterion of whether there are means 'reasonably likely to be used' to identify natural persons on the basis of the data. The test emphatically is not whether it is conceptually, hypothetically or theoretically possible for a dataset to be linked to natural persons, nor does the law require a data provider to be all-knowing of the potential activities that a recipient might undertake with respect to the shared data.

2.3.2. Dealing with the Scania decision in practice – assessment and anonymisation

A clear takeaway from the Scania case is that EU data protection law requires serious reflection from a provider of non-personal data on the likely and foreseeable use of the data by a recipient, and on the possibility that they might have means 'reasonably likely to be used' to identify natural persons. This was the main lesson of the Scania case: in that situation, non-personal vehicle data was indeed likely (and indeed inevitable) to become personal data upon disclosure to independent operators. It was foreseeable, and in fact known to Scania as the data provider, that those independent operators had direct means available to link the formerly non-personal data to individuals – making it personal data in the context of data sharing.

Thus, the lesson that emerges is not that data providers must consider every possible use of their datasets; nor is the lesson that no dataset is safe from GDPR compliance concerns. Instead, **a diligent assessment is required** on the **nature and reasonably foreseeable use of the dataset**, before making it available as open data.

- With respect to the **nature of the dataset**, a data provider should determine whether it objectively contains directly or indirectly identifiable data – i.e. data that could be easily linked to a natural person. The datasets in question might contain names, contact data, private residence addresses, unique identification numbers, personal property information, and so forth. **Such datasets should not be made available as open data, unless they are effectively anonymised first**.

  Anonymisation entails that data is generalised, modified or removed, to the point that there are no reasonably foreseeable techniques or technologies that can undo the anonymisation and link the data to specific natural persons. Anonymisation is challenging, but there are a multitude of techniques that are recommended[11] by European data protection authorities, ranging from relatively simple options such as the randomisation of data (e.g. by replacing some of the original data with semi-randomly generated variations), to much more advanced mathematical models. Once effectively anonymised, the dataset can be published as open data.

- With respect to the **reasonably foreseeable use**, the central question for a provider is whether there are means reasonably likely to be used by reusers of the dataset to convert non-personal data into personal data, by linking it to datasets that they might have. This requires reflection on the content of the dataset, and on what the plausible use of the dataset would be. As noted in the Scania decision, the GDPR does not require clairvoyance or perfect insight into hypothetical use, but it encourages consideration of the most obvious reuse cases and whether they involve data becoming linkable to natural persons.

  This will most often be the case when publishing information about property: for example, a database specifying the registered sales value of houses at a specific address provides only data on property, not people. However, the obvious use of such a dataset is to combine it with residence information, thus resulting in an assessment of the value of the property of specific households, which would constitute personal data. In such cases, making the dataset available as open data without prior anonymisation may be inadvisable.

3. Data protection responsibility and controllership

3.1. Introduction of the central concepts

As with any legal framework, it is important to understand who will bear the responsibility for complying with data protection law. In the EU, much of this question is linked to the notion of a data controller. A data controller is defined in the GDPR as the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of processing of personal data. A data controller (or simply 'controller') determines what personal data will be collected, what will be done with it and how it will be processed.

Identifying the data controller for any personal data processing activity is important, since the data controller is most commonly designated as the entity that has to comply with the GDPR, and that has to be able to prove its compliance. By way of examples, the data controller must be able to demonstrate on which legal basis (such as consent or a legal obligation) it processes personal data, it is responsible for informing data subjects on how and why their data will be processed, it must respond

[11] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, 10 April 2014, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

to data subject rights requests (such as requests to access or delete data) and will be the first target of enforcement by data protection authorities and courts.

As EU guidance[12] on the notion of data controllership stresses, controllers must be identified for a specific processing operation or for a set of operations. The assessment is thus not inherently linked to a dataset. It is perfectly possible for a single dataset to have multiple independent data controllers with fully distinct responsibilities: one entity might be the data controller for collecting and maintaining the data, whereas others are independent data controllers when accessing and using the data for their own purposes.

This is a crucial point for the open data context. A public sector body will generally be the data controller for creating and maintaining a dataset containing personal data, and for using it for its own purposes. If it chooses to make that open data available to a reuser, then that reuser will use the dataset for their own purposes, entirely separate from the purposes of the data provider. In that case, the reuser is an **independent data controller**, and the personal data is transferred between the public sector body and the reuser in a controller-to-controller model. **This is the standard scenario in open data use cases**.

It can also happen, in rare cases, that the public sector body holding the data works with one or more organisations to **jointly determine the purposes and means of a certain processing activity, making them joint controllers**. This can occur, for example, when a data provider wants to work with a third party (such as a university) to develop a new service using the personal data, or when organisations (such as two separate government departments) want to combine their personal data for a common goal (e.g. by creating an enriched database that both of them can use in the future).

The qualification as independent controllers or joint controllers matters significantly from a data protection compliance perspective, because controllers (including joint controllers) bear most of the responsibilities and liabilities under EU data protection law. For open data providers in particular, it is usually preferable not to be qualified as a joint controller with a reuser of a dataset containing personal data, since this could make them jointly liable for the processing activities of that reuser.

Essentially, when sharing open data that contains personal data with third party recipients, it is important for data providers to remain within the scenario of independent data controllership, rather than joint controllership.

It is on this latter topic in particular – the distinction between independent and joint controllers – that relevant new case-law has recently emerged.

[12] European Data Protection Board, 'Guidelines 07/2020 on the concepts of controller and processor in the GDPR', 7 July 2021, https://www.edpb.europa.eu/system/files/2023-10/EDPB_guidelines_202007_controllerprocessor_final_en.pdf

3.2. Recent case-law on data controllership

There have already been multiple decisions from the Court of Justice on the notions of controllership and joint controllership that we won't explore further in this report. However, a recent ruling does merit specific attention, notably Case C 604/22 of 7 March 2024[13], commonly known as the **IAB Europe (IAB) case**.

Briefly summarised, IAB, a digital marketing and advertising association, developed a set of guidelines to facilitate the lawful processing of personal data for advertising purposes. After an investigation, the Belgian data protection authorities imposed corrective orders and sanctions against IAB as a data controller. IAB disputed this qualification, arguing that it merely provided guidelines, rather than deciding on purposes and means.

In its decision on this case, the Court recalled that a **person who exerts influence over the processing of personal data, for his or her own purposes, and who participates thereby in the determination of the purposes and means of that processing, may be regarded as a controller**. Under that criterion, the Court ruled that IAB should be regarded as a joint controller with its members.

The case is particularly noteworthy because the Court applied the qualification of joint controllership in a situation where IAB did not engage in any personal data processing itself, and mainly defined the rules under which the processing of personal data should occur, thus 'exerting an influence' on the data processing activity, which warranted a qualification as a joint data controller. The fact that there was no equal responsibility over the data processing between the sector body and the individual members was not seen as decisive for the qualification.

3.3. Relevance and impact for open data providers

3.3.1. Do open data providers 'exert influence' on reusers?

For open data providers, a common concern when making personal data available for reuse to a third party is the risk of being held responsible or liable under EU data protection law for the processing activities of the reuser. Cases such as the IAB ruling are relevant because they derive two elements from the criteria defined in the GDPR to determine controllership, including joint controllership.

- Firstly, the element of '**exerting an influence**' on the data processing activities, which is not explicitly stated in the formal definition of controllers in the GDPR. From an open data perspective, this is an unpredictable element: if an open data provider publishes a dataset that contains personal data, would this be sufficient to argue that it 'exerts an influence' by doing so, because providing input data inherently affects the means that are used to process data? This question is not addressed by the case-law.

[13] Judgment of the Court of Justice of 7 March 2005, IAB Europe v Gegevensbeschermingsautoriteit, C-604/22, EU:C:2024:214

However, based on the considerations of the Court in the IAB decision, **it does not seem plausible in most open data scenarios to rule that a data provider 'exerts an influence' on the processing activities of the reuser, in the sense of the GDPR**. An open data provider generally limits itself to making data available. This is significantly different from the IAB context, where IAB actively coordinated the rules under which personal data could be processed.

- Moreover, the second element that the Court's jurisprudence derives from the GDPR is the element of personal interest: the influence must be exerted '**for its own purposes**'. In an open data context, the data provider has no 'personal interest' in the reuse of its data. Since the data provider has no common interest with the reuser, then no joint controllership can reasonably occur. The data provider makes data available in the public interest (since open data policies are a matter of public policy), but this interest is not personal, nor is it shared with the interests of a reuser.

3.3.2. No joint controllership by default for open data providers

Based on this analysis, the IAB ruling actually acts as a shield against excessive liabilities for open data providers, by protecting them against the risk of a qualification of joint controllership with a data reuser on two fronts. Firstly, by implementing an open data policy, it emphatically does not exert an influence on the data reuser. Rather the opposite: since open data policies aim to minimise reuse constraints, open data presents a strong argument that there is no influence on the reuse.

Secondly, the data provider's public policy purpose in making the data available for reuse is fully distinct from the purposes of a reuser of the data. For that reason, cases like the IAB ruling matter to open data providers, since they illustrate the importance of distinguishing between the interests and objectives of the data provider and those of the data reuser.

4. Conclusions for open data providers

4.1. Summary of the main findings and the implications

As the summaries of case-law above show, the applicability of EU data protection law is not always easy to determine. In most instances, the assessment is not that complex: a dataset either clearly contains personal data (such as the names of natural persons, their addresses, and so forth) or it clearly does not. Either way, it will be fairly obvious whether EU data protection law applies. When a dataset contains personal data, it is advisable to anonymise it prior to making it available as open data, since free reuse of a dataset containing personal data can be difficult to reconcile with EU data protection law.

In some cases, however, as the case-law shows, whether a shared dataset contains personal data can be a contextual question to which the answer may evolve over time. This can put data providers in a difficult position, since there is always a certain degree of risk to be managed. If seemingly anonymous data becomes personal data over time, for example because a recipient of the data is able to identify natural persons by linking the shared data with external data sources, this creates a GDPR compliance challenge.

In the same way, data providers will usually want to avoid being qualified as joint data controllers with the recipients of their personal data. The case-law notes that this requires that they do not exert any influence over the data processing activities of the data recipients, and that they may not have a joint purpose in common with them that would justify a qualification of joint controllership. As the analysis above shows, this usually does not cause problems in open data scenarios, since the open data provider generally does not concern itself with the ambitions of the reuser.

Moreover, by implementing a few recommendations, open data providers can significantly reduce any compliance concerns.

4.2. Recommendations to facilitate compliance

The case-law shows that the risk of data qualifying as personal data, or of a data provider qualifying as a joint data controller, is contextual. This makes it difficult to define conclusive risk management strategies. Nonetheless, a few elements can clearly be derived that can serve as good practices for facilitating compliance.

- **Know your data.** While EU case-law by its very nature focuses on complex situations, most datasets can be assessed more simply. Prior to making a dataset available as open data, a data provider should examine the contents, and determine whether it contains information that directly enables the identification of a natural person (and thus contains personal data), or whether it contains information that could be fairly easily linked to a natural person by a third party with access to additional information (which makes a qualification as personal data more likely in the future). Examples of the latter include information linked to a location, a file or a physical object, which is sufficiently granular or detailed to allow a third party to link it to a natural person, even if the original information lacked that link. In instances where a qualification as personal data is likely, anonymisation of the dataset should be strongly considered before sharing it as open data.

- **Clearly separate the interests of data providers from the interests of reusers of datasets containing personal data.** In general, supporting reusers of open datasets is beneficial and advisable. However, there are some safeguards to be considered. As the discussion of the IAB ruling showed, if a dataset contains personal data, then it is important for a data provider not to exert any influence over the reuse, and to ensure that it has no personal interest in the reuse, since these are elements that can trigger a qualification as a joint data controller. Thus, a data provider can support a reuser, but when doing so, it should take care to maintain a healthy separation between its own public policy objectives and those of the reuser. The exception is, of course, when both sides feel that they are engaged in a joint initiative for which a qualification as a joint controller is acceptable or even desirable.
- **Periodically re-evaluate the data and the risks.** Assessments of EU data protection law are dynamic, and a qualification of personal data can evolve over time. The contents of datasets might change, for example upon being expanded with new data that a third party can link to natural persons – making it personal data. Moreover, changes in technology (e.g. the increasing power and flexibility of artificial intelligence tools) can increase the likelihood that previously anonymous data can suddenly be linked to natural persons, again resulting in the unexpected application of EU data protection law.

The case-law and the recommendations above show that compliance with data protection rules can sometimes be challenging. However, it is also important to recognise that neither EU data protection law nor the existing case-law requires a data provider to predict perfectly what a reuser might do, or to accept responsibility and liability for the actions of that reuser. Through the implementation of the recommendations above, however, compliance risks can remain quite manageable in practice.

Publications Office
of the European Union