The Anonymisation Decision-Making Framework 2nd Edition: European Legal Context (GDPR)

Mark Elliot, Elaine Mackey & Kieron O'Hara

Published in the UK in October 2020 by UKAN, University of Manchester, Oxford Road, Manchester, M13 9PL

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The Anonymisation Decision-Making Framework (ADF) provides a way of thinking about anonymisation and the reuse of personal data that breaks out of the constraints of overly technical or overly legal framings of the problem. The EU's General Data Protection Regulation (GDPR) was always intended to facilitate proper and appropriate data sharing and reuse as well as protecting data subjects, and effectively anonymising your data whilst still remaining compliant with GDPR is possible, given a suitable framework and set of tools. The ADF provides a mechanism for realising both of these ambitions.

This short document summarises the legal context as given by GDPR of the ten components of the ADF, separated into the three activities of the *data situation audit* (components 1-6), *disclosure risk assessment/control* (component 7), and *impact management* (components 8-10). For further details, see our *Practitioners' Guide*.

Disclaimer

The information provided in this document does not, and is not intended to, constitute legal advice. All information and content in this document are for general informational purposes only. Information in this document may not constitute the most up-to-date legal or other information. Please consult a lawyer or information management professional for advice concerning any particular matter.

THE DATA SITUATION AUDIT

The data situation audit is essentially a framing tool for understanding your data situation (the data and their environment), and therefore to help scope the anonymisation process appropriately for you to share your data safely. It will help you to clarify the goals of the anonymisation process and will enable the more technical aspects of it (component 7) to be planned and conducted more rigorously. Even if you determine that for your data situation, it is not possible to reach a standard of functional anonymisation, for example because of the presence of additional relevant data, a data situation audit will help you effectively assess and manage risk.

Component 1: Describe/Capture the Presenting Problem

Your first task in understanding the data situation is to capture the presenting problem, which is a top-level description of what you are trying to achieve or do. The presenting problem might be specific (e.g. you wish to share an extract from a specific database with another organisation) or it might be more general (e.g. you have an obligation to publish some data as part of a commitment to transparency). An important element of this is to establish what the (intended) use of the data is. Clarity about the use will make specifying the data much easier and feed into considerations in all of the later components.

Legal context: Anonymisation involves by definition the processing of personal data and capturing the presenting problem will in all likelihood require you to think about:

- the lawfulness and fairness of your processing; and
- what your responsibilities for the planned processing are.

To ensure the lawfulness of your processing you need a valid legal basis. For general processing, you will need one of the legal bases set out in Article 6, most of which require that processing is necessary for a specific purpose. If you are processing special category data (see component 4) as part of your general processing you will also need an Article 9 legal basis.

Also of importance to anonymisation is the principle of purpose limitation (Article 5(1)(b)). This requires the purpose for which you process personal data to be compatible with the purpose for which they were originally collected. This, in turn, leads us to the issue of fairness, a well-established principle of data protection law. Fairness when applied in the context of anonymisation requires that this type of processing and the end use of the anonymised data is in keeping with data subject's reasonable expectations. When considering data subjects' reasonable expectations about the reuse of data it is helpful to think about the following (captured in component 6 using the Expectations sensitivity template):

- The data (to be anonymised)
- The context in which you collected the data
- The relationship you as the collector of the data had with the data subjects
- Whether consent for reuse was obtained
- What you told data subjects about how you would process their data

On the issue of responsibilities for processing, Article 5(2) introduces a responsibility for demonstrating compliance. What this means in practice is that data controllers and processors must now also be able to demonstrate compliance (Articles 24 and 28). To do this you essentially need to provide evidence that you have appropriately assessed and managed data protection risk (including risk of loss or destruction of data as well as re-identification). One way of demonstrating compliance is to undertake a Data Protection Impact Assessment (DPIA) and the ADF's Data Situation Audit can be an important part of that.

Component 2: Sketch the Data Flow and Determine Your Responsibilities

Most data situations are dynamic, that is they involve a set of processes by which data are moved from one data environment to another. These environments may be within a single organisation or across different organisations or perhaps an environment is global (data publication). Sketching a data flow from its origin will allow you to visualise the outline of your data situation. The next step is then to refine your focus and this will critically relate to what you are responsible for. In components 3 and 4, you will build up this outline adding key information about your data situation including features of the data environment and the data.

Legal context: Determining your responsibilities across a data flow is critical to ensuring your processing is compliant with GDPR. However, determining who is responsible for what is not always straightforward. The movement of data across multiple environments can complicate the question of your responsibilities in respect of those data, and whether they are strategic, operational or both. The key to resolving this is to examine the flow of data and consider the following questions:

- Roles: Are you acting under your own (organisation's) direction or under the direction of another organisation?
- Data provenance: Where have the data come from, and where are they going?
- Data classification and perspective: What is the status of the data (personal data or anonymous information) for all stakeholders along the data flow?

Although data provenance and the class of data (i.e. personal data or anonymous information) do not of themselves determine one's responsibilities they are closely tied to them, although not straightforwardly.

For example, it is commonly but mistakenly believed that if one's involvement in the data flow is downstream from data collection, the origin of the data has little to do with you. Similarly, another false assumption is that if you don't have access to some personal data, then you cannot have the position or responsibilities of the data controller. Neither statement is correct.

To explain why, we need to consider how responsibilities are specified in data protection law. The law provides a description of two types of processing role: data controller and data processor. Data controllers are those that determine the essential means and purpose of the processing, e.g. deciding what data to collect, from whom and what are the legal grounds for doing so. Two organisations can act together as joint data controllers – this arrangement should transparently set out what the agreed roles and responsibilities are for complying with GDPR for each organisation. In contrast, a data processor acts on behalf of the controller; this arrangement may give the data processor a degree of autonomy in respect to the non-essential means of processing. Anyone one else processing data who does not fit into one of these two roles is commonly classed as a 'data user'. Following this, let us now qualify the two belief statements above.

1. If your involvement in a data flow is downstream from data collection, the origin of the data is important. Understanding the origins of data can help you understand what your role is in processing it and the role of others as you map out the data flow. As a general rule, unless you have data controller responsibility for the data in question you will need instructions from those with controller responsibility to allow you to process it, including anonymising the data (which is a type of data processing).
2. If you do not have access to the personal data you may still have (data controller) responsibility. You may have controller responsibilities for personal data you do not have access to if you have determined the purpose and/or the essential means of the processing.

The table below sets out how processing roles and responsibilities, data provenance and data classification interrelate.

	Data controller (DC)	**Data processor (DP)**	**User (Persons who are not a DC or DP).**
Role	Determines the purpose and essential means of processing. The DC may have responsibility alone or jointly with others.	Acts under the direction of DC. May have autonomy over non-essential means of processing.	Has no role in determining the purpose or means of the processing of personal data
Provenance	A DC might collect data, direct the collection of data, or compel under a statutory requirement the sharing of data. This means the origin of the data may be the controller's own Data Environment (DE) or DEs upstream from them.	A DP might collect data on behalf of a DC or have data shared with it under the direction of the DC. This means the origin of the data may be the processor's own DE or (many) DEs upstream.	Provided with access to functionally anonymised data by DC or DP (on instruction of DC).
Data type	You can still be a DC even if you do not have access to personal data. Commonly, though, the DC holds the means for identifying data subjects. Implementing appropriate technical and organisational measures can involve (amongst other things) keeping directly identifying information (often referred to as keys) and the attribute data separately. If the DC destroys the keys for a dataset, the question of whether the data are personal still or anonymous information would need to be properly assessed through a data situation audit. It should not be assumed if keys are destroyed that the data are no longer identifiable.	The data for the DP may be identifiable (either directly or indirectly) and therefore classed as personal data. For data held that is indirectly identifying the risk of identification might have been mitigated such that the risk of re-identification is considered very low but above negligible meaning it is still classed as personal data.	For the receiver of the data to be considered a user the data must be functionally anonymised. This may be achieved through either restrictions on the data or restrictions on a combination of the data and environment.

Component 3: Map the Properties of the Environment(s)

Once you have sketched out the data flows, you can map the properties of each environment in terms of the four data environment elements (agents, other data, governance and infrastructure).

Legal context: GDPR does not explicitly address the issue of environments or data context; it does specify that appropriate technical and organisational measures are required to ensure appropriate security (Article 5(1)(f)). In addition, as outlined in the Introduction of this Guide, the Regulation states that to determine identifiability, account should be taken of all objective factors to assess the Means Reasonably Likely, such as cost, time and available technologies (Recital 26). The MRL, we would suggest, can best be assessed using the organising concept of the data environment, i.e. by considering human action, the availability of other data and the presence or absence of governance processes and infrastructure.

Component 4: Describe and Map the Data

The next layer of information to add is the data themselves. You will describe the data within each environment across a range of parameters: data structure, data type, variable type, population, dataset properties, variable and topic sensitivity. These parameters relate to risk in terms of either the likelihood or the impact of a breach.

Legal context: GDPR introduces new types of data as (potentially) personal data including location data, online identifiers and genetic data. It also lists special category data which need greater protection because they pose particular risk to the rights and freedoms of data subjects. These are:

- racial or ethnic origin
- political opinions
- religious or philosophical beliefs
- trade union membership
- genetic data
- biometric data (where used for identification purposes)
- health
- sex life
- sexual orientation.

Processing special category data is prohibited unless one of the ten exceptions detailed in Article 9 applies. Five of the exceptions only apply if your processing has an authorisation in member state law. In the UK this authorisation is set out in the DPA.

Criminal offence (personal) data is not classed as special category data although there are additional safeguards around processing it. That is, for processing criminal offence data you require a lawful basis under Article 6 and either legal or official authority under Article 10. The DPA sets out the specific conditions providing lawful authority.

Component 5: Engage With Stakeholders

Your use case and the data situation will include interactions with various stakeholders, including data subjects, data providers, data consumers, customers, clients, and, for public service providers, citizens. Your ability to attain and preserve a good reputation for trustworthy data stewardship depends on retaining the trust of these stakeholders. Engage with stakeholders to understand their expectations in their dealings with you. What do they trust you to do? Are there groups who are likely to resist the use of data proposed in the use case? Can you engage with them to address their concerns, manage expectations, or even to change their minds? Or, if you decide to go ahead in the face of resistance, is that sustainable in terms of either political capital or your business case? While you may understand that these questions are important you may question their relevance to anonymisation. The primary relevance is a consequence of the realistic risk principle of functional anonymisation (that zero risk is not a realistic possibility if you are to produce useful data). As risk is greater than zero an adverse outcome is possible. Engagement with stakeholders will mitigate the damage of such an outcome, and therefore lowers the risk it poses.

Legal context: While engaging with stakeholders is indeed an ethical issue that goes beyond a legal requirement, data protection legislation does make provision for ensuring that processing for one particular stakeholder, the data subject, is fair and transparent as well as lawful. So in addition to a legal basis for processing (as described under component 1) processing should: (i) be used compatibly with the purpose of collection, in a way which is consistent with data subjects' reasonable expectations; (ii) be fair; and (iii) not be hidden or presented in a misleading way. The idea that data subjects should know about your processing activities and any related risks, and therefore requires from you some level of engagement, is touched on in various places in GDPR, including Article 6, Recital 50 with respect to the processing of data collected for a different purpose, Articles 12, 13 and 14 with respect to the provision of transparent information for the exercise of data subject rights, and Articles 34 and 36 with respect to the communication of a breach and risk. Article 35(9) requires that data subjects or their representatives be consulted as part of a Data Protection Impact Assessment, "where appropriate"; the anonymisation use case should make it clear whether it is appropriate for you.

Component 6: Evaluate the Data Situation

At this point in the ADF process, you should have a diagram of your data situation detailing the data flow which identifies for each focal data environment roles and responsibilities and the properties of both the data and environment(s). Now you carry out an evaluation of all of the elements. Can you proceed to share/release the data or do you need to assess the risk in more detail and/or put in place further controls on that risk? The ADF provides tools and techniques for aiding this decision.

Legal context: Articles in Chapter 4 of GDPR address the issue of risk and mitigation. The notion of processing being necessary and proportionate is a key concept. In particular:

- Article 25 requires that at the time of planning and at processing, both the principles of data protection and appropriate technical and organisational measures are implemented to ensure that only personal data necessary for a specific purpose are processed.
- Article 32 requires data controllers and processors to take account of risk and implement a level of security appropriate to that risk.
- Article 35 introduces as a new obligation the requirement to conduct a Data Protection Impact assessment in cases where the processing of personal data is envisaged to likely result in a high risk to the rights and freedoms of natural persons. What constitutes a case of high risk processing is categorised very broadly in GDPR. These categories are expanded, and greater detail provided, in Guidance from the Article 29 Working Party on Data Protection and the ICO. Article 35(7) directs that the assessment should contain at least:

    i. A description of the proposed processing activities and purpose of processing;
    ii. An assessment of the necessity and proportionality of the processing;
    iii. Assessment of the risk to the rights and freedom of data subjects; and
    iv. The measures proposed to address the risks and also demonstrate compliance with the Regulation

Essentially underpinning GDPR – and its interpretation by most regulators – is the general notion of due diligence. By capturing all of the elements of your data situation in components 1-5 and then evaluating it, you will strengthen your due diligence case.

DISCLOSURE RISK ASSESSMENT AND CONTROL

Risk assessment and control should usually be an iterative, not linear, process. There is rarely a single possible solution; the risk analysis might suggest changes to the data specification which, once experimentally applied to the data, require a fresh risk analysis. Furthermore, there are several types of risk assessment, and you should be strategic in how you apply them. Some are quite resource-intensive and therefore should only be applied to near-final versions of the data if they are needed at all (assuming your budget is limited).

This process will be constrained by the use case and the resources available. As ever, our goal is to produce data that meet the requirements of the use case. The use of resources to address potential risks should be proportionate to the likelihood and impact of a breach.

Component 7: Select and Implement the Processes You Will Use to Assess and Control Disclosure Risk

If your assessment in component 6 is that the risk may be unacceptable (non-negligible) then you should employ disclosure risk assessment and control methods. This component is about selecting those methods. The choice of methods should be proportionate to the risk. Options for assessment include penetration (or intruder) testing, data analytical risk assessment and comparative data situation analysis. Options for control include controls on the data (suppression, noise addition etc.) and controls on the environment (access and licensing). Alternatively, in choosing the controls you apply to the data, you might select ones that satisfy a confidentiality model definition such as differential privacy or k-anonymity In that case data analytical risk assessment may not be needed, if the level of risk assured by the model is sufficiently low.

Legal context: As discussed in previous components there is a requirement on data controllers and processors to ensure that technical and organisation measures are implemented appropriate to the processing being carried out. For assessing re-identification risk there are three important connecting concepts underpinning GDPR notion of identifiability.

i. The motivated intruder – characterised by the ICO as someone "who wishes to identify the individual from whose personal data the anonymised data has been derived."
ii. The idea that a claim of re-identification should be more than a 'lucky guess' and should carry with it a reasonable degree of confidence. This is quite a tricky concept to work with but the essential idea here has three parts:
    a. The claim of re-identification is correct.
    b. The claimant is confident that it is correct.
    c. That confidence is well grounded in empirical and/or statistical evidence.
iii. A description of a level of risk of identification. As noted in the Introduction of the Guide and underpinning the realistic risk principle, risk may be remote, but never zero.

We apply the three concepts (motivated intruder, degree of confidence and level of identification risk) to the means reasonably likely to be used test to determine identifiability.

IMPACT MANAGEMENT

Much of what we have considered so far has framed risk management in terms of reducing the likelihood of an unintended disclosure, but impact management requires a plan for reducing the impact of such an event should it happen.

Component 8: Maintain Stakeholders' Trust

Maintaining the trust of your stakeholders implies behaving in a trustworthy way but also involves engaging with them to make your trustworthiness evident. Communication need not be frequent, but you should be transparent, and provide updates if the situation changes. Provide stakeholders with a responsible point of contact, so that they can communicate any concerns directly. The importance of this should not be underestimated, because your freedom of action to deal with a problem will be conditioned by the extent of stakeholders' trust in you.

Legal context: See component 5.

Component 9: Plan What to Do if Things Go Wrong

You have been diligent and followed all the steps above. However, you are managing risk, not eradicating it in its entirety. Residual risk means that an adverse event could happen. Put in place a crisis management policy covering four key areas: breach management, notification, review and communication. Consider carefully a range of likely breach scenarios, who the agents will be, their goals, means and whether these goals exacerbate or ameliorate the impacts of the breach. Make clear what your own goals are and how these will interact with those of the other agents. Then consider the set of possible actions that you could take for each permutation.

Legal context: If things go wrong and there is a confidentiality breach, you may be required to report it to the supervisory authority in your country (in the UK this is the ICO). Article 33(1) stipulates when and how a personal data breach should be reported and under what conditions – i.e. notification is required "unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons". GDPR addresses a wider range of breach types than is dealt with in this Guide. In additions to re-identification and disclosure it considers a breach to include unauthorised data access, security incidents, and loss, damage or destruction of data.

If you need to notify the supervisory body, the notification should include a description of the nature of the breach, categories of data, number of people involved, assessment of the potential impact, initial steps taken to mitigate that impact and details of the designated point of contact. Communication with the data subjects affected by a breach is a requirement under Article 34 when said breach is likely to result in a high risk to the rights and freedoms of those data subject. There are three exemptions to this: (i) the controller has applied technical measures such as encryption that would render the breached data unintelligible; (ii) the controller has taken subsequent measures to ensure high risk is an unlikely outcome; or (iii) notification is considered to involve disproportionate effort (in which case public communication is required).

Component 10: Monitor the Evolving Data Situation

Risk is neither exactly calculated nor constant. You should produce and implement a policy for monitoring the risk and consider adjusting the data situation if it changes significantly. Review the data situation periodically and assess whether any of the elements have changed. Keep a log of all such changes and assess whether the net effect of all changes requires a full case review.

Legal context: Monitoring the data situation comes down to the key interrelated issues of perspective and responsibility, data and environment, so while data may be considered of very low risk to the processor or even functionally anonymised for the end user they remain personal data and the responsibility of the data controller(s). This is important because it means that even when data are considered anonymised for particular agents they continue to be someone's (i.e. the data controller's) responsibility and are captured under the framework of GDPR.

As developed in component 6, a DPIA (of which a Data Situation Audit could form a key part) as a living document could provide a useful mechanism for monitoring your evolving data situation.