The Anonymisation Decision-Making Framework 2nd Edition: Overview

Mark Elliot, Elaine Mackey & Kieron O'Hara

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

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 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*.

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.

The Data Situation Audit should determine the answers to three primary questions:

- **PQ1:** What in the data situation are you or your organisation responsible for (alone or jointly)?
- **PQ2:** Within that locus of responsibility is there a non-negligible disclosure risk that needs to be addressed?

- **PQ3:** How sensitive is your data situation?

If your data situation is simple, your answers to these questions may also be simple. However, if you identify complex stakeholder relationships in component 5 or more complex than expected data flows in component 2 then your anonymisation problem may inherit the complexity.

In principle, the outputs from components 1 and 2 will provide you with the basis for answering PQ1 (responsibility), the output from components 3 and 4 will provide you with the basis for answering PQ2 (risks to be addressed), and the method set out in component 6 answers PQ3 in terms of the likely cost/impact of a wrong decision at this point (sensitivity).

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.

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.

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).

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.

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.

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.

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.

Having selected your chosen methods, you should implement them. You may need to iterate between risk assessment and control to balance risk with data utility. Finally, having done this, repeat component 6. Loop through 6 and 7 until you have reached the point where the risk is negligible.

IMPACT MANAGEMENT

Much of what we have considered so far has framed risk management in terms of reducing the likelihood of an unintended disclosure happening, but it would be irresponsible not to prepare for the worst. 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.

Component 9: Plan What to Do if Things Go Wrong

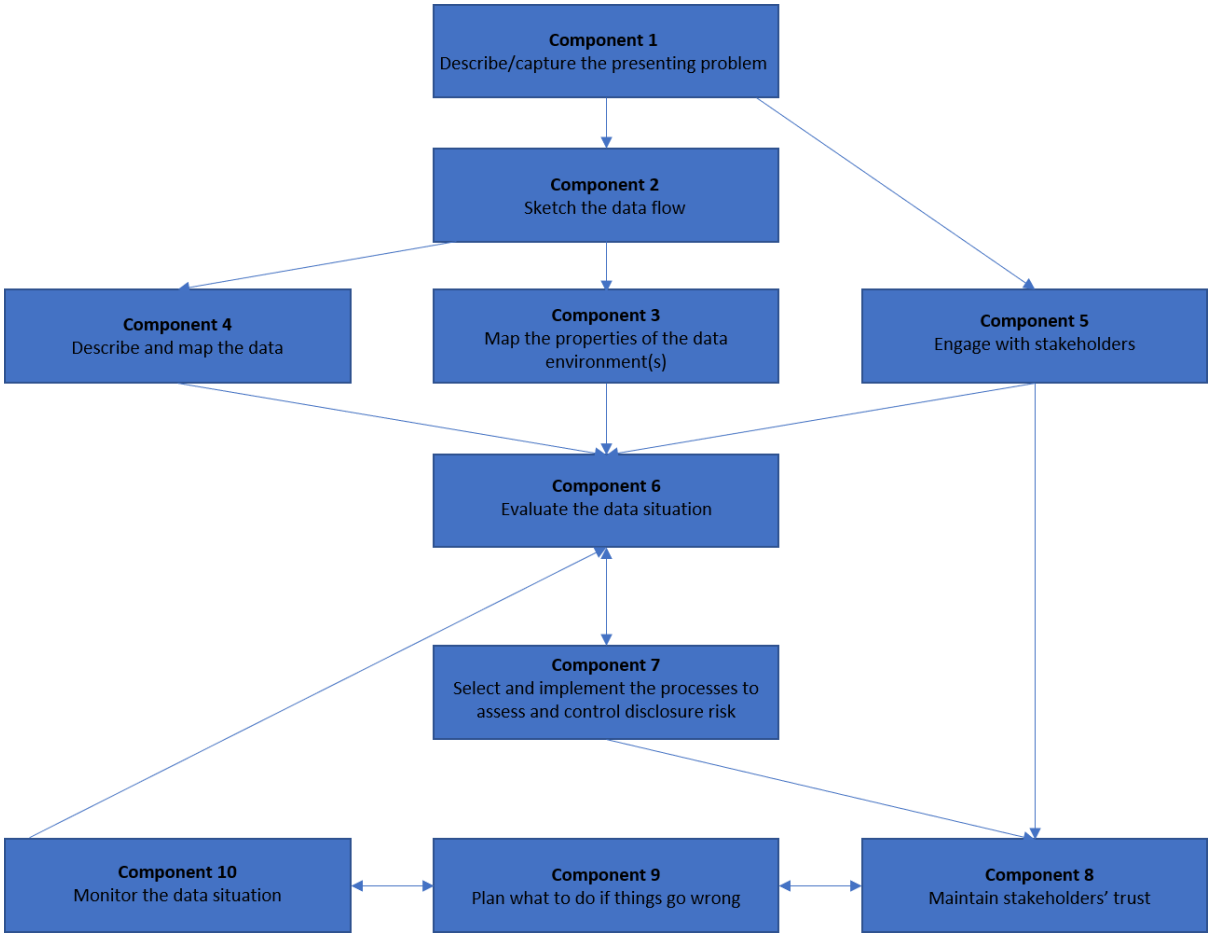
You have been diligent and followed all the steps above. However, you are managing risk, not eradicating it in its entirety. Residual risk means that an adverse event could happen. Put in place a crisis management policy covering four key areas: breach management, notification, review and communication. 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.

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. This would essentially mean revisiting component 6 and if necessary 8 and 9 as well.

WORKFLOW

One important point of practice is that although the components are presented as an enumerated list and indeed it does usually make sense to start at component 1 and work down the list, the framework is not rigidly sequential (and is certainly not a checklist). This affects some components more than others. For example, stakeholder engagement (component 5) might happen at any point of the process and evaluating your data situation (component 6) may be repeated several times as you propose and revise changes to your data situation. A typical workflow is shown below.


Component 1
Describe/capture the presenting problem
Component 2
Sketch the data flow
Component 4
Describe and map the data
Component 3
Map the properties of the data environment(s)
Component 5
Engage with stakeholders
Component 6
Evaluate the data situation
Component 7
Select and implement the processes to assess and control disclosure risk
Component 10
Monitor the data situation
Component 9
Plan what to do if things go wrong
Component 8
Maintain stakeholders' trust