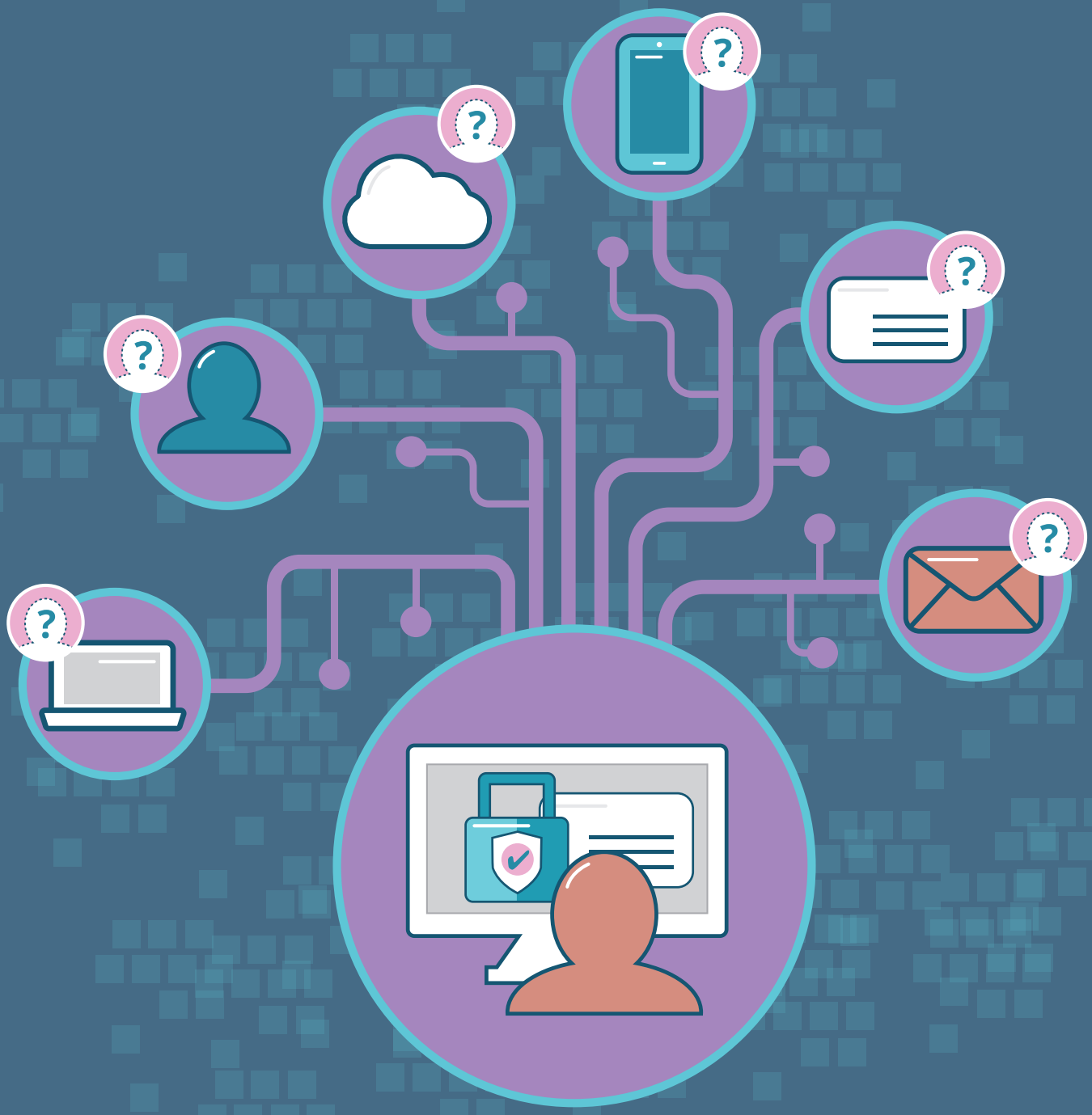
2nd Edition

The Anonymisation Decision-Making Framework: European Practitioners' Guide

Mark Elliot, Elaine Mackey
and Kieron O'Hara



UKAN Publications

The Anonymisation Decision-Making Framework 2nd Edition: European Practitioners' Guide

Mark Elliot, Elaine Mackey & Kieron O'Hara

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

Table of Contents

PREFACE

The need for well-thought-out anonymisation has never been more acute. The drive to share data has led to some ill-conceived, poorly-anonymised data publications including the Netflix (CNN Money 2010), AOL (Arrington 2006), and New York taxi (Trotter 2014) cases, underlining how important it is to carry out anonymisation properly and what can happen if you do not.

We published the original *Anonymisation Decision Making Framework* (ADF) in 2016 to address a need for a practical guide to anonymisation that gave more operational advice than the UK Information Commissioner's Office's (ICO) valuable *Anonymisation Code of Practice*. At the same time, we were concerned to be less technical and forbidding than the existing statistics and computer science literature (Elliot et al 2016a).

The framework was a long time in gestation. Its foundations were a twenty-year programme of research carried out at the University of Manchester, and the longstanding relationship between the University and the UK Office for National Statistics (ONS). More recently, the authors of this book have been partners in the UK Anonymisation Network (UKAN),[1] which drove the development of the framework and convinced us of the enormous demand for a guide in this space. One aim of UKAN, and indeed the ADF, was to integrate the many different viewpoints on the topic of anonymisation, and in particular to join up the legal and the technical perspectives. We would like to express gratitude for the contribution of the ICO, which provided the seed funding for the network from 2012-2014 and has been actively engaged with its development ever since.

In 2018, with funding from the Higher Education Funding Council for England, we began the project of updating the framework. There were two drivers for this: firstly the enactment of the EU's General Data Protection Regulation (GDPR) meant that we needed to review whether the legal elements of the ADF (which were based on the 1995 EU Data Protection Directive and the UK's Data Protection Act 1998) were still fit for purpose; and secondly, as we now had an extensive user community we wanted to hear their voices on how useful the ADF was and how it might be improved. So three workshops were held, first with legal experts, then with our user community, and finally with international technical experts. The input of these three communities was immensely valuable and constructive, resulting in an overhaul of the ADF'S ten-component structure, a rethink of some of the terminology and a decision to release the revised ADF in multiple formats (of which this Practitioners' Guide is the first).

Our view has always been that anonymisation is a heavily context-dependent process, and only by considering both the data and their environment as a total system (which we call the *data situation*) can one come to a well-informed decision about whether and what anonymisation is needed. Good technique is important but without a full understanding of the context, the application of complex

[1] UKAN provides services including training workshops and clinics for those who need to anonymise their data. These services can be accessed via the network website: www.ukanon.net.

disclosure control techniques or confidentiality models is a little like installing sophisticated flood defences: if you don't know whether you are based in the Mississippi flood plain or the Atacama Desert, you simply won't know how appropriate, valuable or effective your countermeasures will be. Given the importance of context, it is also important to understand that a fully formed anonymisation process includes consideration of the *ethics* of data sharing and the importance of *transparency* and *public engagement* (since these, as we will argue, make important contributions to the context of data sharing), and you will find as you work through the book that the framework incorporates these elements too.

This Guide is primarily intended for those who have data that they need to anonymise with confidence, typically in order to share it for some purpose in some form. Our aim is that you should finish the book with a practical understanding of anonymisation and an idea about how to utilise it to advance your business or organisational goals. To make this tractable we have focused on personal data and specifically on information presented in the form of a file or database[2] of individual level records (often referred to as microdata – a term that we will use here).[3] We should emphasise that this is not a recipe book with a set of canned instructions about how to produce a perfect anonymisation. Rather it is more like a guide to becoming a better cook. That said the Guide comes with some specific tools and templates to capture and evaluate your data situation and these we hope should make most data situations more tractable.

In this Guide we assume the regulatory context of current (2020) UK law;[4] other legal jurisdictions will impose different constraints on what you can and cannot do with data. Across jurisdictions there are differences in the interpretation of data protection legislation and in the meaning of key terms such as 'personal data'. A reader outside the UK (and European) context should interpret the legal matter in the Guide with this in mind. That said, the fundamental premise of anonymisation, that it is designed to control the risk of unintended re-identification and disclosure, will hold regardless of the legal context and therefore the *principles* that the framework provides should be universally applicable. In places, we have included pointers to other jurisdictions where we could do that without making the text cumbersome and future publications may attend to this issue in a more thorough manner.

On the ADF page of the UKAN website (www.ukanon.net), you will find a set of companion documents and files. These documents are referred to particularly in components 6 and 7 and provide more detailed treatments of more technical topics, a tool for carrying out a data situation audit and some synthetic data files on which to practice certain techniques.

[2] Thus, at this present time we do not consider unstructured data. However, the principles of the ADF do apply to this type of data and we envisage widening the scope of the ADF to incorporate an examination of it in future publications.
[3] We have also set aside the specialist topic of data about businesses, which of course may involve important issues of confidentiality, but need not be personal data. Such data has different technical properties and different legislation can apply to it.
[4] Which at time of writing aligns with EU law (specifically, GDPR) on application (under the UK Data Protection Act 2018).

We have decided to release this Guide as a freely available open access book rather than through a traditional publisher as we want to ensure that it is disseminated as widely as possible. We hope that you find it of value. We would welcome comments on the Guide at any time via our website www.ukanon.net. It is a living document and the ADF is in a state of continuous improvement; we will be updating it periodically. Although three authors have their name on the cover, we are grateful to those who have already contributed ideas, either through the workshops or commenting on drafts. Particular thanks to Claire Spencer for unfailing administrative support.

INTRODUCTION

In this chapter, we introduce the Anonymisation Decision-making Framework (ADF), explaining the thinking behind it and the principles on which it is founded. The ADF is made up of ten components, which we will describe in the next full section of this Guide; each component is a set of actions to achieve important goals. We outline how you might best use the ADF (given your skills and experience) in your anonymisation practice. But first, let us make explicit the three central terms featured in this book: *anonymisation*, *risk* and *sensitivity*.

Anonymisation, risk and sensitivity

A common error when thinking about *anonymisation* is to focus on a fixed end-state of the data. This is a problem because it leads to much muddled thinking about what it means to produce 'anonymised data'. Firstly, it focuses exclusively on the properties of the data whereas in reality the anonymity or otherwise of data is a function of both the data and their context. Secondly, it leads one into some odd discussions about the relationship between anonymisation and its companion concept *risk*, with some commentators erroneously (or over-optimistically) assuming that anonymity entails zero risk of an individual's being re-identified within a dataset.[5] Thirdly, viewing it as an end-state means that one might assume that one's work is done once the anonymisation process is complete and the end-state is produced, which in turn promotes a counterproductive mentality of 'release-and-forget'.

In many ways, it would be better to drop the adjectival form 'anonymised' altogether and perhaps talk instead of 'data that have been through an anonymisation process'. However, the constraints of the English language mean that this would sometimes lead to some quite tortuous sentences. So, in this book, we will use the term 'anonymised' but this should be understood in the spirit of the term 'reinforced' within 'reinforced concrete'. We do not expect reinforced concrete to be indestructible, but we do expect that a structure made out of the stuff will have a negligible risk of collapsing.

To that end, what we propose, and what is implemented by following the components of the ADF, is termed *functional anonymisation*.[6] Functional anonymisation does not assume that anonymisation can be zero-risk or irreversible; it is meant instead to bring anonymisation practice in line with the art of the possible, in particular by understanding that whether data are or are not anonymised is not a property of the data, but determined by the relationship between the data and the context(s) in which they are held. Given that, it is clear that risk cannot be totally eliminated, but rather we work to reduce the risk of re-

[5] This is the basis for example of Paul Ohm's much cited but ultimately ill-conceived attack on anonymisation (Ohm 2010).

[6] (Elliot et al 2018) for a detailed exposition of the concept of functional anonymisation and a rebuttal of Ohm's argument.

identification of individuals from functionally anonymised data to a negligible level.[7]

This means that in turn we have to consider the notion of *risk*. Since Amos Tversky and Daniel Kahneman's seminal work in the 1970s, it has been clear that humans are quite poor at making judgements about risk and are subject to numerous biases when making decisions in the face of uncertainty (e.g. Tversky & Kahneman 1974). One aspect of this is the tendency to confuse the likelihood of an event with its negative impact (or disutility). To complicate matters further, where risks are dependent on human action, these biases themselves factor into the risk profile. So, for example, if we can convince a data intruder[8] that the likelihood of a re-identification attempt succeeding is negligible, then they are less likely to put the necessary effort in to attempt it (in the same way as putting motion lights on your house, or even keeping a loud dog, can cause a burglar to pass by down the street) and thus we have controlled the risk beyond what we can measure objectively.

Thinking about the impact side of risk brings us to the third key concept, *sensitivity*, which tends to be understood as roughly proportional to the potential harm of any confidentiality breach.[9] However, as we will see, sensitivity is a larger concept than this and encompasses other issues such as how the data were collected, and what reasonable expectations a data subject might hold about what will happen to data about them. So a key point in the anonymisation decision-making process is evaluating how sensitive the overall data situation is (we address this in component 6 of the ADF).

Throughout this Guide we deal with the complexity of the confidentiality risk by developing a bivariate system. The term *risk* will simply refer to our assessment of the likelihood of an adverse event (most often the re-identification of a data unit) given all of the information currently at our disposal. The second concept, *data situation sensitivity*, will be used to describe those elements which affect the impact of an adverse event.

Anonymisation, then, is a process of risk management but it is also a dynamic decision-making process answering the practical question of whether we should release these data or not, and if so in what form? Considering all the elements

[7] We use the term 'negligible' – others, including the ICO, use the term 'remote'. By 'negligible' we mean *"the risk that a reasonable person would ignore".* As I go about my daily life, I ignore the risk of being hit by a meteor even though it is possible that that could happen. I do not ignore the risk of being hit by a car. The former is negligible, while the latter is not.

[8] A relatively minor point of terminology: different disciplines and sectors working in this area have different terms for the 'bad guys'. You will see 'adversary', 'intruder', 'attacker', 'snooper', and other similar terms used. These are essentially synonymous. Our preference is for 'intruder'.

[9] Throughout this Guide we will be using the term 'confidentiality breach' to refer primarily to re-identification or a similar disclosure resulting from a failure to anonymise successfully. We will not be referring to other data protection issues referred to in GDPR and elsewhere such as *unauthorised processing*, *unlawful processing*, *accidental loss*, *destruction* or *damage*. Nor will we be referring to matters such as the duty of confidence which fall under the heading of confidentiality.

involved, that decision can appear complex and rife with uncertainties. It does require thinking about a wide range of heterogeneous issues from ethical and legal obligations to technical data questions. Bringing these disparate elements under a single comprehensible framework is what the ADF is all about.

Disclosure control

Let us introduce another term at this point: *disclosure control.* This term is sometimes incorrectly used interchangeably with anonymisation. However, it denotes a more specific practice than anonymisation. Disclosure control is simply the manipulation of data or their environment to reduce the risk of disclosure of personal information from those data.

One form of disclosure control is *statistical disclosure control* (SDC) – a suite of sophisticated tools and techniques for assessing and controlling disclosure risk through data manipulation based on statistical models. This is sometimes confusingly just referred to as 'disclosure control', but SDC is not the only data-focused disclosure control technique. There are also formal models – the most well-known being the oddly-named *differential privacy*, which is a mathematical technique based on information theory.

The key point is that when we refer in the Guide to disclosure-controlled *data*, we simply mean data to which such a manipulation has been applied (whether it has been successful in its aim or not). This is particularly relevant when we are describing the importance of data flows in component 2.

A distinction that has become more prominent in the SDC literature is between *input* and *output* disclosure control. Inputs are data moved into an analytical environment and outputs are the products of the data analysis. From the ADF point of view this is a distinction without a difference. The output environment (usually publication which is often open or quasi-open) is just another environment and the analytical products are themselves data. It is true that the details of the risk assessment processes will be quite different (with output SDC this is often called 'Output Checking') but essentially it is still the movement of data from one environment to another. In component 7 when we discuss these more technical processes we will be focused more on inputs than outputs because this is the far more difficult problem to solve.[10]

The legal context for this Guide – GDPR and the UK's Data Protection Act

The General Data Protection Regulation (2016/679) frames the legal context for the protection and movement of data across and outside of Europe (for EU and UK citizens' data) and for the British reader, the UK's Data Protection Act 2018 (DPA) which incorporated the main text of the GDPR when it came into law. The DPA provides useful interpretation to GDPR's provisions and should be read alongside GDPR. EU countries will each have their own national Data Protection

[10] For detailed discussion about the types of output disclosure control, (Ritchie & Elliot 2015, Griffiths et al 2018).

Act which address a number of derogations[11] from GDPR, concerning topics such as national security, age of consent for children, supervisory authorities, sanctions, etc.

GDPR is underpinned by a bipartite model; data is either personal data or anonymous information. It has little to say about anonymous information other than: (i) it is information that does not relate (or no longer relates) to an identified or identifiable natural person, and (ii) it is out of scope of the legislation. In other words:

> 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. (Recital 26, GDPR 2016/679)[12]

To learn more about the notion of anonymous information under GDPR, we need to understand what is meant by the terms 'identified' and 'identifiable' which underpin the concept of personal data. The Regulation defines personal data as "any information relating to an identified or identifiable natural person" (Article 4(1)):

1. an identified natural person is one who can be identified *directly* from the data;
2. an identifiable natural person is one who can be identified *indirectly* from the data.

Directly identifying data is a relatively straightforward notion, associated with the presence of formal identifiers such as name, address and unique common reference numbers. The notion of identifiable data and its application to real world data situations is however more complex. To unpack the complexities of applying the term 'identifiability' in practice we need to consider the definition of 'pseudonymisation' as described by GDPR, approaches to thinking about anonymisation and GDPR's *means reasonably likely to be used* test.

GDPR introduces to data protection legislation the term 'pseudonymisation', defined as:

> … the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. (Article 4(5))

Given this description we might suppose that pseudonymisation is a process for moving data from a state of being identified to a state of identifiability. It is a

[11] GDPR contains a number of *derogations* that enable Member States to apply a degree of discretion on applying certain provisions.
[12] All future references to the text of GDPR will just give the article or recital number.

process involving protecting the direct identities attached to the data, i.e. removing, masking or replacing direct identifiers, and employing technical and organisational measures to ensure that whatever you have done to that data cannot be easily undone. Under GDPR, personal data that have undergone a process of pseudonymisation are still considered personal data, since it only requires a look-up table or details of the encryption algorithm to break the protection.

Thus far we have referred to the law from a data-centric perspective: the description given of pseudonymisation focuses on the data, employing technical and organisational measures to prevent the resulting identifiable data from being linked back to the original identified data, so the resulting (identifiable) data should still be considered personal data. However, how we describe the *resulting data* once formal identifiers are removed is not predetermined by the pseudonymisation process. From a functional anonymisation approach one has to take account not only that of the data themselves (and what processes they have undergone, including pseudonymisation) but also of the data's context. To do this, we need to apply the *means reasonably likely to be used* test (MRL) described in Recital 26.

> 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 …, 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. (Recital 26)

We set out in components 6 and 7 how you may go about determining the MRL. A key point to reiterate is that the risk of re-identification of useful data is never zero – this is one of the principles underpinning the ADF, and held by many in the academic community and across National Statistical Institutes (Rubinstein 2013). It was also the approach taken by the ICO in its 2012 Code of Practice on Anonymisation.

> The DPA [Data Protection Act 1998] does not require anonymisation to be completely risk free – you must be able to mitigate the risk of identification until it is remote. If the risk of identification is reasonably likely the information should be regarded as personal data – these tests have been confirmed in binding case law from the High Court.[13]

You may wonder why, if anonymous information is out of scope of GDPR, data protection legislation is an important consideration to the topic of anonymisation. This is because:

1. Determining whether negligible risk has been achieved is a major challenge.

[13] ICO Anonymisation Code of Practice: 6, https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf.

2. Anonymisation involves, by definition, the processing of personal data.

Or, more clearly, if you get anonymisation slightly wrong, what remains is personal data.

The principles behind the ADF

The ADF incorporates two frames of action: one technical, the other contextual. The technical element of the framework will enable you to think about both the quantification of disclosure risk and how to manage it. The contextual element will enable you to think about and address the factors that affect that risk. These include the particulars of your data situation such as the data flow, legal and ethical responsibilities, governance practices, your responsibilities after share or release, and your plans for the rare event where things go wrong.

The framework is underpinned by a relatively new way of thinking about the re-identification problem that posits that you must look at both the data and their context to ascertain realistic measures of risk. This is called the *data situation approach*. The approach will be explained in more detail later, but the basic intuition is that all data are held in some sort of context, which we call the *data environment*. When we consider all the environments in which some data are held together, we will call this aggregated context (which, for reasons which will become clear, is relevant to determining whether the risk of re-identification is negligible) the *data situation*. Formally, a data situation is *the aggregate set of relationships between some data and the set of their environments.*

Perhaps it seems obvious that the environment(s) in which data are to be shared and released is(are) important, but for many years the data confidentiality field focused almost exclusively on the data themselves. Consequently, risk was seen as originating from, and largely contained within, the data, and is why 'release-and-forget' was seen as acceptable. As the examples given in the Preface of AOL, Netflix and New York taxis show, release-and-forget is far from acceptable. With a few notable exceptions (e.g. Duncan & Lambert 1989, Elliot & Dale 1999, Reiter 2005) researchers in this area have not concerned themselves with issues such as how or why a re-identification might happen, or what skills, knowledge, or other data a person would require to ensure the attempt was a success. Consequently, the models researchers built to assess re-identification risk, whilst statistically sophisticated, have at best been based on implicit assumptions about the data context and at worst completely detached from any real-world considerations.

To address these failings, there have been attempts to describe and theorise about context beyond the data. This has usually taken the form of intruder scenario analysis that we will consider in more detail in component 7 of the ADF. Scenario analysis began the process of shifting attention away from the traditional question 'how risky are the data?' towards the more critical question 'how might a disclosure occur?' (and therefore 'how might we reduce its risk?'). The data situation approach builds on this and broadens our understanding to include the actions of other key agents, other data within the environment and previously neglected considerations such as information governance processes. The basic premise is that you cannot guard against the threat of re-identification unless you

have a clear idea of what it is you are guarding against, and this requires you to consider the whole data situation, not just the data.

What this means for you is that your assessment and management of disclosure risk should include reference to all the components of the ADF, including your data, other external data sources, legitimate data use and potential misuse, governance practices, and your legal, ethical and other ongoing responsibilities. The ADF is a total system approach, based on four principles of functional anonymisation: *comprehensiveness*, *utility*, *realistic risk* and *proportionality*.

Comprehensiveness Principle: ***You cannot decide whether or not data are safe to share/release by looking at the data alone, but you still need to look at the data.*** This principle encapsulates the data situation approach outlined above, where risk is seen as arising from the interaction between data, people and (the soft and hard) structures that shape that interaction (such as national policies on data sharing and access, the legal framework, IT systems, governance practices, cultural attitudes to data sharing and privacy, etc.). You do also need to know your data – which means being able to identify the critical properties of your data and to assess how they might affect risk. This will feed into decisions about how much data to share or release, with whom and how.

Utility Principle: ***Anonymisation is a process to produce safe data but it only makes sense if what you are producing is safe <u>useful</u> data.*** You may wonder why we talk about the need to balance data utility with data safety in the anonymisation process. It is easy after all to think about anonymisation solely in terms of producing safe data, but if you do that you could paradoxically end up taking a risk for no actual benefit (or worse).

- There is little point in sharing or releasing data that do not represent whatever they are meant to represent sufficiently well. There are two possible outcomes that arise from low utility and neither are happy ones: either the data are of little or no use to their potential users and you will have wasted your time and resources anonymising them, or the data could lead to misleading conclusions which might have significant consequences if the data are used for decision-making or problem-solving.
- Low-utility data may still retain some re-identification risk but in the absence of demonstrable utility you will lack any justification for taking that risk.

So, remember, anonymisation is a process that cannot be understood independently of the intended use(s) of the data.

Realistic Risk Principle: ***Zero risk is not a realistic possibility if you are to produce useful data.*** This is fundamental. Functional anonymisation is about risk management, nothing more and nothing less; accepting that there is a residual risk in all useful data inevitably puts you in the realms of balancing risk and utility. But the trade-off of individual and societal benefits against individual and societal risks is the stuff of modern life. This also brings into focus the issue of stakeholder engagement. There is no agreement on how to have a conversation with data subjects and the wider public about this issue and there are concerns (not

unfounded) about causing unnecessary worry by drawing attention to confidentiality risks. At the same time, it is worth recognising that people are capable of balancing risk and utility in much of their daily lives whenever they cross a road, drive a car etc. We consider these issues in more detail in components 5 and 8 of the ADF.

Proportionality Principle: ***The measures you put in place to manage risk should be proportional to that risk and its likely impact.*** Following the realistic risk principle, the existence of risk is not necessarily a reason for withholding access to data. However, a mature understanding of that risk will enable you to make proportionate decisions about the data, who should have access and under what conditions. However, disseminating as anonymous data which remains personal data is a breach of the Data Protection Act.

A final point about the ADF is that it is applicable beyond being a process for the negligible risk standard. You may accept that given your use case and the environment that the data in question are and can only be personal. However, the ADF can still help you in your obligations to data minimisation. The approach that the ADF embodies constitutes good practice in data management regardless of how you classify the data. The structures and data flows that you model are reusable, because these data environments will not go away. So your compliance with regulation, good ethical practice and your relationships with your stakeholders can all be facilitated going forward by using the ADF to audit your data.

Structure of this Guide

In this Introduction, we have introduced some of the core concepts relevant to our approach to confidentiality and functional anonymisation. We have also provided a high-level overview of the Anonymisation Decision-Making Framework, explaining the thinking behind it and the principles upon which it is founded. The ADF is a general approach to the process of anonymisation which will help you to identify and address the key factors relevant to your particular data situation. In the rest of this Guide, we work through each of the ten components of the ADF in detail. The approach taken is practical, with worked examples and advice on how to operationalise each component. As we have prioritised accessibility of the narrative over precision and completeness, some of the more technical aspects of disclosure risk assessment and control (for example synthetic data generation) are necessarily passed over but in many cases highly technical processes are unnecessary, and when they do prove useful it is generally better to work with an expert on their application.

THE FRAMEWORK

The ADF has ten components, which are grouped together into three activities. These are:

- ***A data situation audit*** (components 1-6). This activity will help you to identify and frame those issues relevant to your data situation. You will encapsulate and systematically describe the data, what you are trying to do with them and the issues thereby raised. A well-conducted data situation audit is the basis for the next core activity.

    1. Describe/capture the presenting problem
    2. Sketch the data flow
    3. Map the properties of the data environment(s)
    4. Describe and map the data
    5. Engage with stakeholders
    6. Evaluate the data situation

- ***Risk analysis and control*** (component 7). Here you assemble the processes that you will need to employ in order to both assess and manage the disclosure risk associated with your data situation.

    7. Select and implement the processes you will use to assess and control disclosure risk

- ***Impact management*** (components 8-10). Here you consider the measures that should be in place before you share or release data to help you to communicate with key stakeholders, ensure that the risk associated with your data remains negligible going forward, and work out what you should do in the event of an unintended disclosure or security breach.

    8. Maintain stakeholders' trust
    9. Plan what to do if things go wrong
    10. Monitor the data situation

How you use the framework is likely to depend on your level of knowledge and skills as well as the role you play in your organisation. Some might use it for knowledge development purposes, to understand how a confidentiality breach might occur and its possible consequences, or to develop a sound understanding of the important issues in the anonymisation process. Others might use it directly to support management of the risk of a confidentiality breach, reducing risk to a negligible level.

Anonymisation is not an exact science, and even using the ADF at this level you will not be able to avoid the need for complex judgement calls about when risk is sufficiently low given your data situation. The ADF will help you in making sound decisions based on best practice, but it is not an algorithm; it is an approach whose value depends on the extent of the knowledge and skills you bring to it. You may still need expert advice on some parts of the anonymisation process, particularly with the more technical risk analysis and control activity. However, even in such a position the ADF can still be very useful; you and your expert will have more

fruitful discussions, make quicker progress and will be more likely to produce a solution that works for you if you properly understand your data situation. Consider the ADF as a member of your team; it will not solve all your problems, but will provide graded support appropriate to your own level of expertise.

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 in Figure 1.

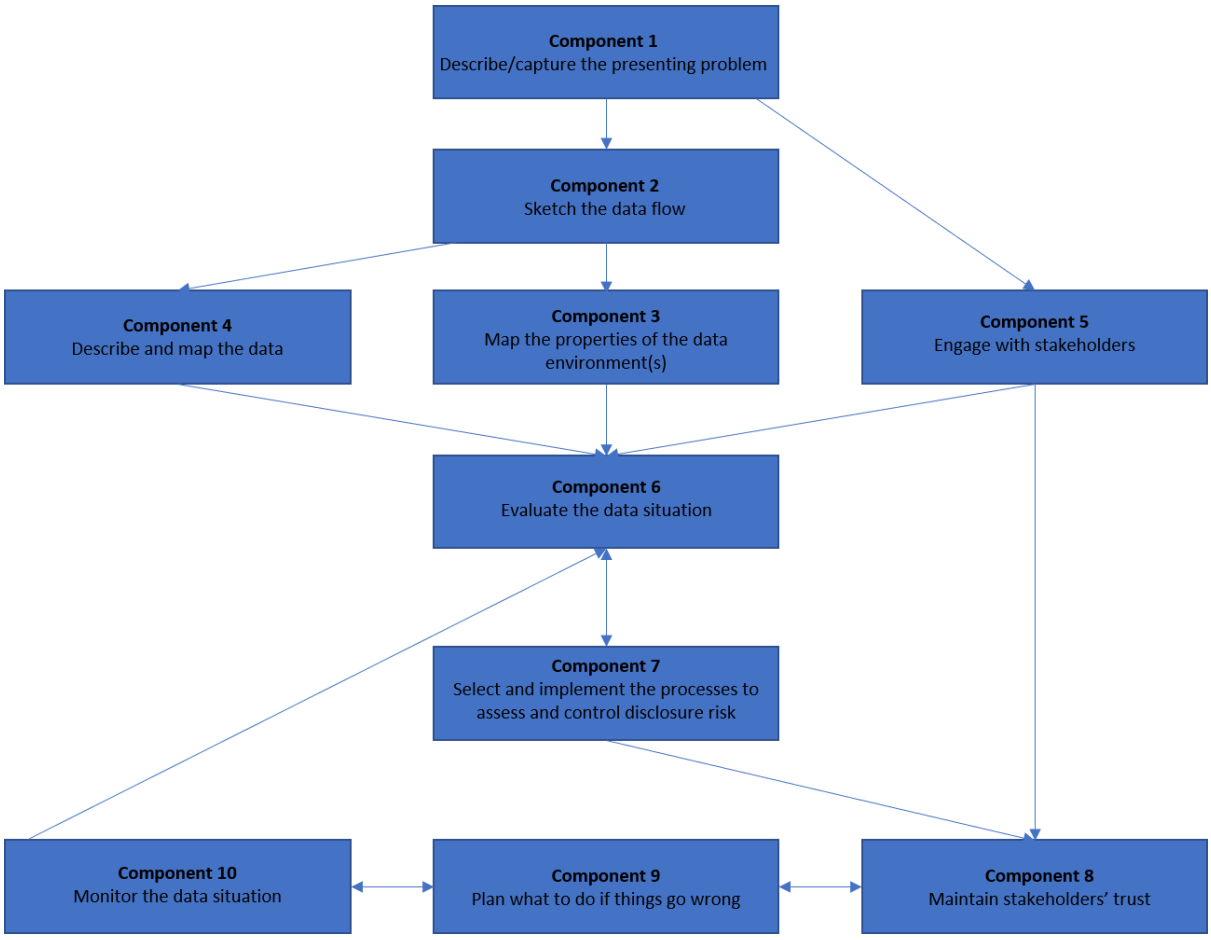


Figure 1: ADF workflow

Each component will be described in this Guide, beginning with an overview of its purpose, and then relevant legal notes based on the UK/European legislation in GDPR. Then the various decision-making steps relevant to most uses of the component to deal with the types of data we focus on in this Guide will be set out.

A final point before we launch into the framework in detail: in all likelihood, you will need to adapt the ADF to suit your own needs. Some aspects may be more important than others for your particular data situation. All the important considerations are there but you will need to think how they relate to and affect

each other. Most importantly, in applying the framework you should keep clear in your mind that the objective is to disseminate *safe useful data*. In your data situation, *you* will be the best judge of what they might be.

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).

The data situation audit consists of ADF components 1-6.

1. *Describe/capture the presenting problem*
2. *Sketch the data flow and determine your responsibilities*
3. *Map the properties of the environment(s)*
4. *Describe and map the data*
5. *Engage with stakeholders*
6. *Evaluate the data situation*

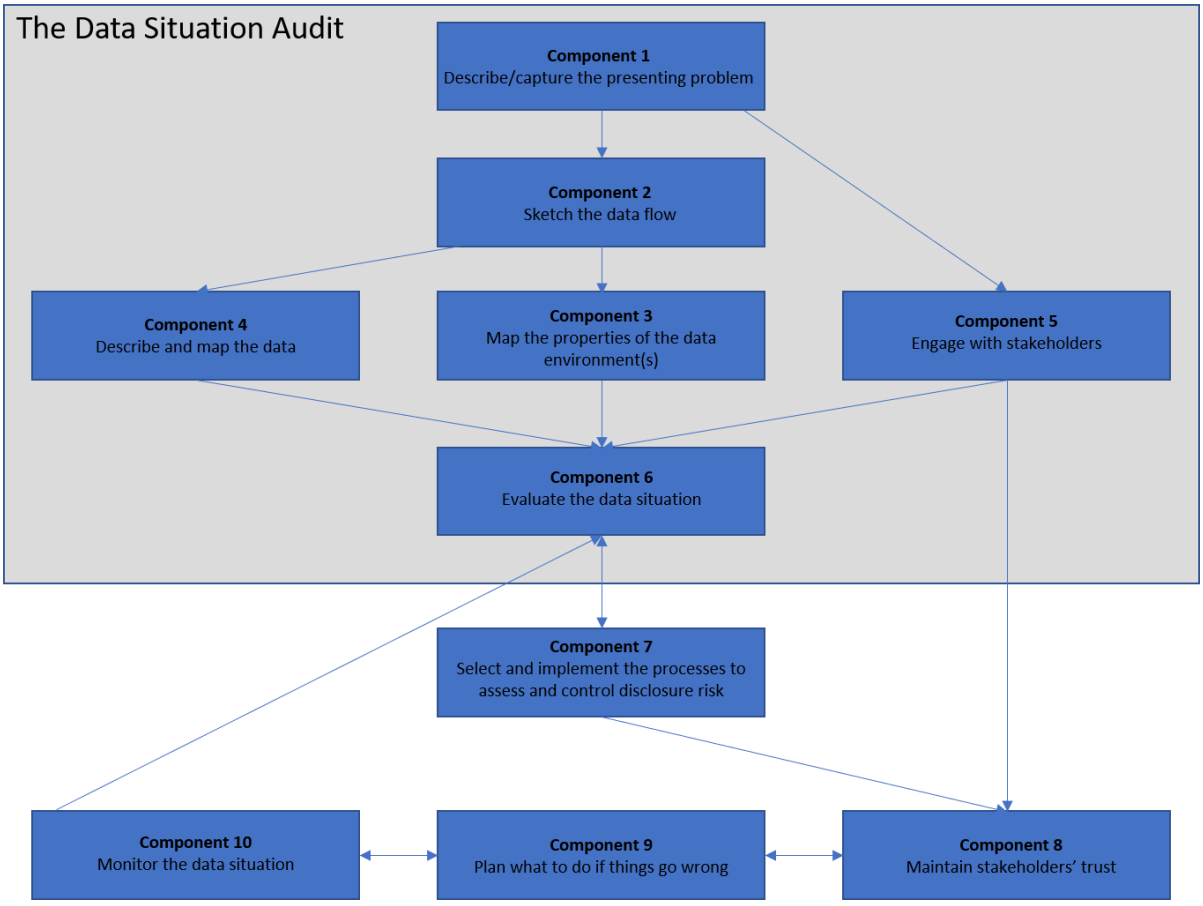


Figure 2: The data situation audit

Component 1: Describe/Capture the Presenting Problem

Overview: *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.[14] *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*

[14] Article 5(1)(b) states "further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes."

- *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.*[15]

* * *

Capturing the presenting problem can sometimes be simple but it often requires some thought. The goal is to come up with as clear as possible a statement of what you want to do with the data in question.

1.1 Perspective

The very first question we must ask is "Who are you?" This is not a cue for existential searching but rather to understand your roles and responsibilities in the data situation. You could be an individual within an organisation who is responsible for managing data flows *within* that organisation; we describe this as an *internal data situation*. Alternatively, you could be an individual who represents the whole organisation in terms of data flows *between* your organisation and another organisation (sharing) or the outside world in general (publication or dissemination); these are *external data situations*. Whether your data situation is internal or external will inform the granularity of data flow that you capture in component 2.

External data situations tend to be represented by coarsely grained data flows, with individual environments that may represent whole organisations. Internal data situations tend to be represented by finely grained data flows with environments that may be teams or even individual servers/computers. This does not mean that you ignore details if you are looking at an external data flow but you will tend to focus on specific relevant details. For example if you are sharing data with a large company you may not be concerned with every detail of their

[15] ICO provide a DPIA template: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/. Data situation audits can feed in at the organisational level or at the level of a data situation or a combination of both.

data processing operation but you may drill down on the specific elements that are relevant to the share.

1.2 The core presenting elements

To refine the presenting problem, you need to clarify three interrelated elements:

- **Why:** the reason that you wish to share or release the data (*the rationale*).
- **Who:** the individuals, groups and/or organisations who will use the data (*the users*).
- **What:** what the users of the data (might) want to use it for (*the use case*), and what uses the environment for the data affords (contrasting a safe setting from uncontrolled re-dissemination).

Working through these three questions will help you with decisions about both what data you can safely share or open and what are the most appropriate means by which to do this.

In most cases considering these questions will resolve the presenting problem into one of four types, depending on how reuse is framed.

a. Sharing of data with another party.
b. Dissemination of data (publication or release).
c. Continued use of data after some deadline for destruction.
d. Re-use by the same organisation of data for a different purpose other than that for which they were collected.

There are numerous rationales for sharing or releasing the data. Perhaps it provides useful information for stakeholders or about your organisation, offers new insights/perspectives on a topic, offers a benefit to particular groups, supports the more effective/efficient use of a service, or maybe you have received an FOI (freedom of information) request and you need to be able to anonymise the data in order to be able meet that request. Thinking through why you are disseminating your data automatically brings in the other two questions, the 'who' and the 'what' of access.

Your potential users may be a single organisation, a defined group or several different user groups. You may decide to provide different data products via different dissemination routes.[16]

Direct consultation with your potential data users (as part of component 5) is one method for understanding the use case and can take many forms. Methods available include interviews, focus groups, web surveys or a call for written feedback, the last of which you could administer directly through a website or via a third party. The exact nature of the type of activity you might carry out will depend on the number and type of users and the drivers of the programme to share/release. Are they internal or external to your organisation? Are you responding to a contractual or statutory obligation, or are you trying to increase

[16] If you make available different data products via different dissemination routes you will need to take account of the risk of disclosure for each in combination with the others. See component 7 for further discussion

the utility of your data? Is it a drive for transparency and goodwill, or do you hope to provide an income stream?

However you decide to engage with your users, it is helpful from the outset to identify who they are and how they will use your data, although this is not always possible as the detailed use case may emerge over time. Certainly, data released for one reason and for a particular user group may in future be used serendipitously for purposes not initially envisaged, and by new groups of users. Whilst you may not be able to determine all possible uses for your data immediately, you should try to keep abreast of how they are being used. That there will be some benefit to the reuse of data is axiomatic in today's 'big data' climate. The demand for data seems insatiable. Clarifying the questions to be answered by your data, or what needs it is hoped they will meet, is a good place to start when thinking about exactly what data to release and how they should be specified.

1.3 The variable presenting elements

It could be that data or the data environment are specified in the presenting problem. For example, specific data may have been requested or the receiving environment will definitely have to be an independently managed secure processing facility. Where the data are specified, we can then focus on the environment: 'what environmental controls would make such a share sufficiently safe?' In the latter case where the environment is specified, the presenting problem becomes data focused: 'given the receiving environment is it possible to specify a dataset which meets the use case and for which the risk of the share is sufficiently low?' Or it could be that both the data and environment are loosely conceived for the presenting problem and you need to work out, through a process of propose and revise, iteration and reiteration, what data and environment features will allow you to safely and responsibly meet the use case.

1.4 Risk management or full anonymisation?

The final element often specified in the presenting problem is the level of risk you are working to. The key question is whether you are endeavouring to functionally anonymise the data – so that within the receiving environment the users are not processing personal data – or you accept that downstream the data will be personal data but are trying (in line with the principle of data minimisation) to minimise the risk of re-identification without necessarily reaching the standard of negligibility.

Component 2: Sketch the Data Flow and Determine Your Responsibilities

Overview: *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.[17] *The law provides a description of two types of processing role:* data controller *and* data processor*. Data controllers*[18] *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.*[19] *In contrast, a data processor*[20] *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.

Table 1 sets out how processing roles and responsibilities, data provenance and data classification interrelate.

---

[17] For ICO guidance on roles and responsibilities, https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf. The European Data Protection Supervisor published guidelines in November 2019 on determining which role you occupy, https://edps.europa.eu/sites/edp/files/publication/19-11-07_edps_guidelines_on_controller_processor_and_jc_reg_2018_1725_en.pdf. See also DLA PIPER blog, https://blogs.dlapiper.com/privacymatters/eu-new-guidelines-on-the-concepts-of-controller-processor-and-joint-controllership/.

[18] In GDPR a data controller is "the natural or legal person, public authority, agency or other body, which alone or jointly with others, determines the purpose and means of the processing of personal data" (Article 4(7)).

[19] For ICO guidance on roles and responsibilities, see https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/controllers-and-processors-1-0.pdf.

[20] In GDPR a data processor is "the natural or legal person, public authority, agency or other body, which acts on behalf of the controller" (Article 4(8)).

	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*	*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.*

	Data controller (DC)	***Data processor (DP)***	***User (persons who are not a DC or DP)***
	are no longer identifiable.		

Table 1: What to consider when determining your responsibilities

* * *

In the Introduction, we defined the term *data situation* as *the aggregate set of relationships between some data and the set of their environments.* For example, your organisation itself will constitute an environment, whilst any proposed share or dissemination would constitute another environment. Each environment will have a different configuration of the same core features: people, other data, infrastructure and governance processes.

Data situations can be *static* or *dynamic*. A static data situation is where there is no movement of data between environments; a dynamic data situation is where there is such movement. By definition all data shares or dissemination processes take place within dynamic data situations in which data are intentionally moved from one environment to another. A dynamic data situation might be relatively straightforward involving the movement of data from just one environment to another environment. Often though, it is more complex, involving multiple environments.[21]

For example, in Figure 3, we see data flowing through a number of different environments. Most of the environments are within Organisation A, but in data environment n+1, the data are transferred to Organisation B, perhaps as a result of a data share, or because Organisation B is processing data on behalf of A. The data situation as a whole, then, consists of all the environments and their data flows aggregated, whether in Organisation A or B. The data controller is whoever determines the purpose and essential means of processing, and may be one or more people situated in Organisation A, or Organisation B, or both. The data situation is not restricted to the set of responsibilities of a particular organisation, or a particular occupant of a role.

[21] Actually, there are many more variations in data situations than this distinction suggests. We do not here consider issues arising from multi-party computation for example. However, for the purposes of exposition we will restrict ourselves to the relatively simple case of a unidirectional sharing/dissemination process.

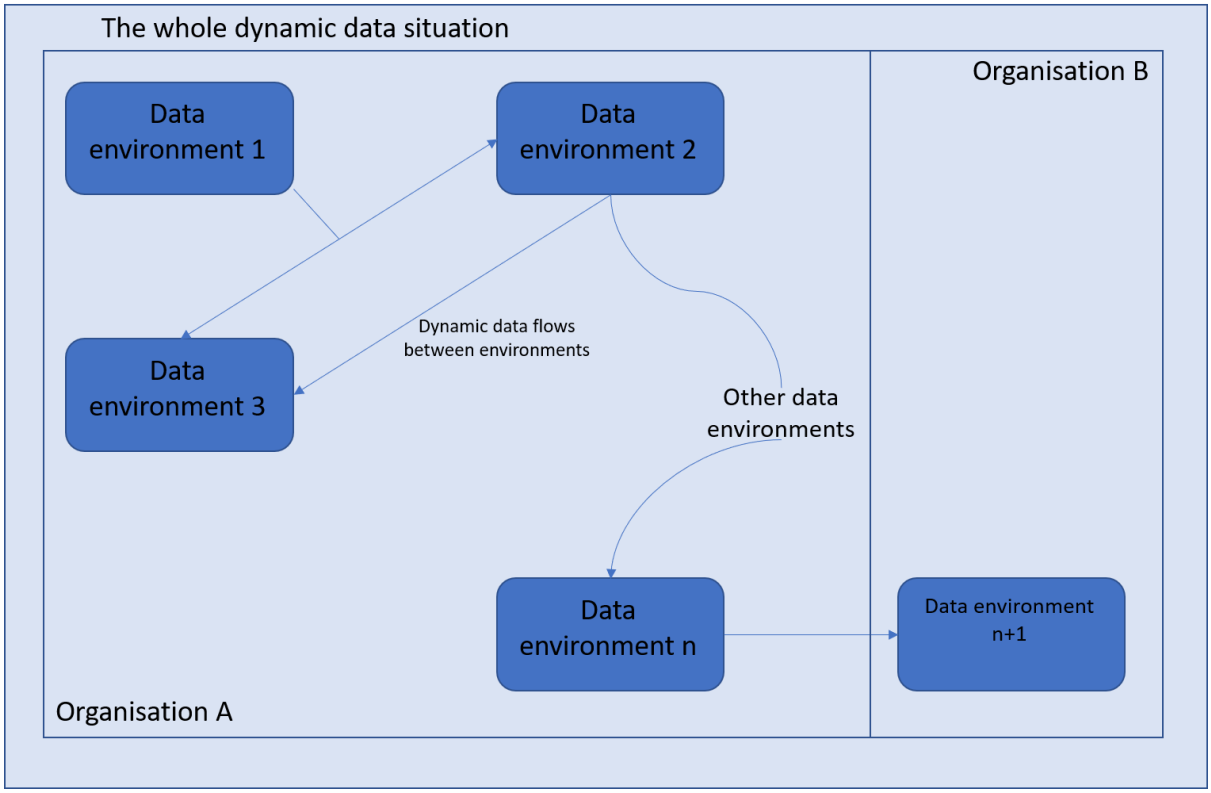


Figure 3: Data environments in a data situation

At this stage we want to familiarise you with the idea of data moving between environments. Whilst data environments can be thought of as distinct contexts for data they are interconnected by the movement of data (and people) between them. As we have said previously, by mapping the data flow from the point at which data are collected to the point after which they are shared or released, you will be able to define the parameters of your data situation. Before we illustrate this idea further in the examples given below, it is worth reiterating that determining responsibilities even in a simple data flow can be complex and should be considered on a case by case basis. The examples below serve merely to illustrate how various factors will come into play when working out who is responsible for what along a data flow. Those factors include, in addition to who is determining the essential means and purpose of processing:

- The specification of the data to be shared
- The specification of the receiving environment, i.e. who will have access to the shared data and how and what other information is in the receiving environment. How to specify an environment is addressed in component 3.

2.1 Example data situation 1: simple share

In this first example, we look at the data flow between environments of data that have been subject to disclosure control.

Imagine that a franchised public transport provider PubT collects personal data from its customers relating to public transport usage. Call this environment 1. As part of its processing activities, PubT plans to share a disclosure-controlled version of the data with the Local Authority of Barsetshire, which wants to use it to support

(better) provision of public transport. PubT agrees to share an extract of the data – or perhaps is obliged to as part of its service agreement. To do this:

i. It restricts the data by removing direct identifiers, e.g. the customers' names and addresses, and by reducing the detail on several key variables. However, it leaves some other key variables – which are of particular interest to Barsetshire – unchanged. We imagine that the detail left in the data extract (to be shared with Barsetshire) is such that within the PubT environment it is still personal data – because PubT have the means to re-identify the data.

ii. It controls the share environment using a Data Sharing Agreement,[22] which: (a) specifies how Barsetshire can hold the data, analyse it and who can access it; (b) proscribes Barsetshire from sharing or releasing any part of the data without the prior agreement of PubT; (c) requires Barsetshire to keep the data securely and evidence its destruction at an agreed time and (d) allows PubT to audit Barsetshire in respect to its processing of the data.

After this contract is signed, the dataset is passed to Barsetshire. We call Barsetshire's arrangements environment 2.

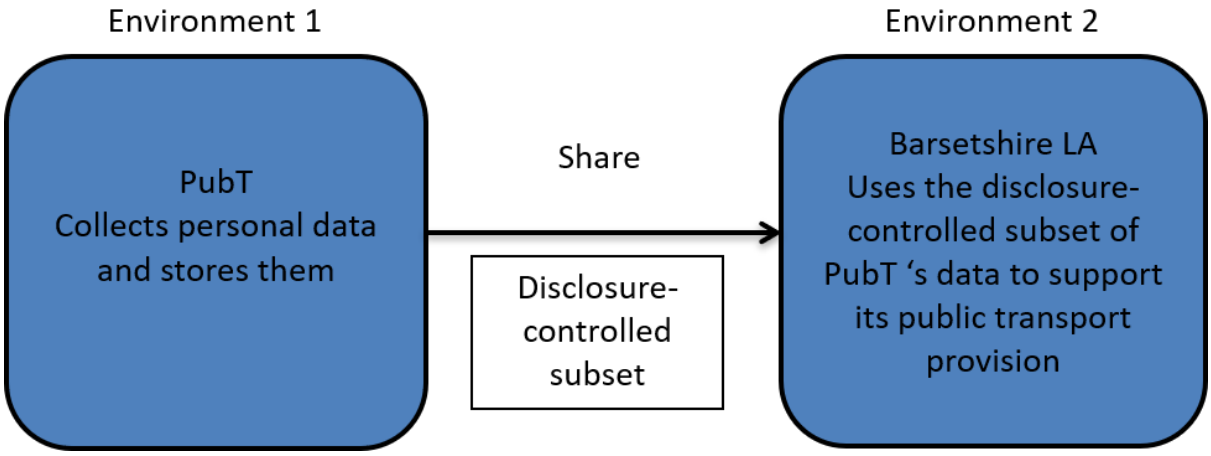


Figure 4: Data flow between two environments in a simple share

Figure 4 illustrates the intentional movement of data from environment 1 to environment 2. The data flow between PubT and Barsetshire defines the parameters of the data situation. By using a Data Sharing Agreement to place governance and infrastructure controls on environment 2, PubT manages some of the disclosure risk associated with the data situation. The (disclosure-controlled) data within Barsetshire's environment might be considered low (and possibly even negligible) risk in that environment even though they contain some detailed key variables. This is because by placing controls on the environment they are

[22] A Data Sharing Agreement should set out a common set of rules to be adopted by the organisations involved in the share. It should cover such issues as: (i) the purpose(s) of the share; (ii) the recipients of the share and the circumstances in which they will have access; (iii) the data to be shared; (iv) data security; (v) retention of shared data; and (vi) sanctions for failure to comply with the agreement. See the ICO's Data Sharing Code of Practice, https://ico.org.uk/media/for-organisations/documents/1068/data_sharing_code_of_practice.pdf.

effectively managing the 'who' and 'how' of access. These factors can reduce the risk considerably, and in these circumstances the data now held by Barsetshire may be functionally anonymised, and not personal data. Whether this has been achieved will also depend on the details of the data and of other data that Barsetshire holds. At this point, it is likely that a risk analysis is necessary – and indeed this is precisely what the data situation audit leads you towards: are there any feasible scenarios (i.e. "means reasonably likely to be used" in GDPR terms) by which data might be re-identified.

So we will assume for now that the data held by Barsetshire are functionally anonymised and therefore not personal. The key question becomes: is Barsetshire simply a user in this data situation? Well, yes and no! The data that they themselves are holding are not personal and therefore – by definition – for those data they are a user. However, Barsetshire has been involved in determining the purposes for which the data that PubT hold are processed. Although there are many possible complex nuances in how Data Sharing Agreements are written, the default assumption in this arrangement should be that Barsetshire becomes a joint data controller for PubTs data – *in respect of the processing required for this particular use*.

Once you have digested that, then one question that arises is why bother to anonymise at all; why not just share the personal data? There are several reasons why not. Firstly, sharing personal data is a different form of processing to sharing data in a manner that means that it is not personal. If the data are anonymised then the use case is almost certainly statistical/research and therefore will not be deemed incompatible with the original purpose of collecting the data (Article 89(1)) and therefore will be lawful (assuming the fairness criteria are also met). Secondly, because the data are functionally anonymised, PubT can share the data with confidence and will be meeting their own obligations in law. Thirdly, the fact that Barsetshire has data controller status means it is dividing the risk of the share proportionately, so that PubT will be more willing to share. Fourthly, Barsetshire will not take on all of the responsibilities of data controller for PubTs data – for example it will not have to meet data subject access requests (because the data to which it has access are functionally anonymised and therefore by definition it will not be able to identify individual data subjects). Barsetshire's data controller responsibilities will be specified in the Data Sharing Agreement (which will therefore be a controller-to-controller agreement). In essence though, they can summarised *as treat the data with same care as you would the original personal data*. Finally, functional anonymisation protects *all* parties in this arrangement: PubT, Barsetshire and the data subjects, whilst allowing useful data to be shared for public benefit.

2.2 Example data situation 2: data release as open data

Consider now a second situation, which captures the notion of open data which is core to the transparency driver that underpins many policy-driven data releases.

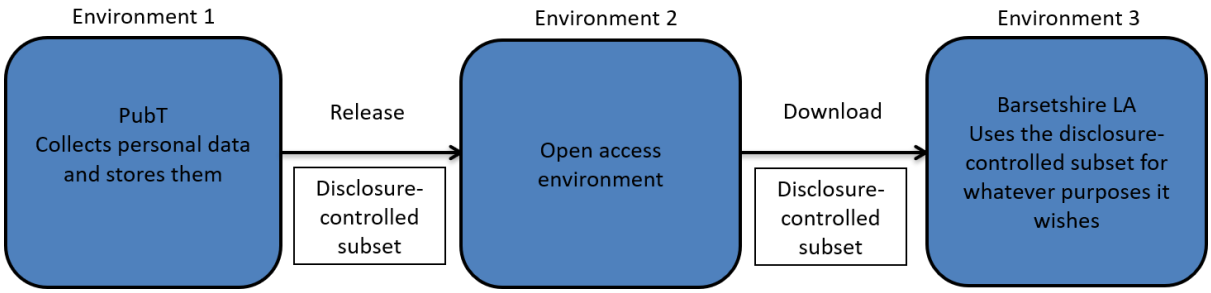


Figure 5: Data flow for open release environments in a simple share

In this less restricted environment the risk associated with these data may well not be considered low. PubT are no longer controlling the who and how of data access. Where access is truly open, anyone who has access to the Internet can download the data and do whatever they want to do with them. Barsetshire in this arrangement are simply one of the users of the data, downloading it from the open access environment. They have no responsibilities, and are not even under an obligation to tell PubT about their use of the data. The risk of this release falls completely onto PubT (and the data subjects). Finally, release into the open data environment means that in principle the data could be linked to other data held anywhere in the world, which is only going to increase the number of feasible scenarios (means reasonably likely to be used) for re-identification.

What should be clear from this is that, in order to functionally anonymise the data in data situation 2, it is necessary for PubT to ensure that the data qualify as anonymous information in and of themselves, which implies that they will be considerably more restricted than in data situation 1. This may seem obvious, but failure to understand the basic point that data releases need to be appropriate to their release environment is the primary cause of the well-publicised examples of poorly anonymised datasets such as the AOL, Netflix and the New York taxi driver open datasets mentioned in the Preface. Put simply, the same data that are functionally anonymised in the controlled environment of Barsetshire, will be personal data in an open release.

2.3 Example data situation 3: simple share with secondary open release

Let's develop the concepts outlined in the first two data situations further and imagine that Barsetshire would like to release some analytical output from the data openly. For example, it might want to publish aggregate cross-tabulations of public transport use by key demographics as part of a transparency initiative. Aggregate outputs are still data and so such a release extends and indeed complicates the data situation. The third environment in the chain is the open environment. The new picture of the data flow is shown in Figure 6. This is an example of what we call a *two-step data situation*. The first step is completely described by data situation 1 above, but the second step is effectively a new data situation and needs separate analysis.

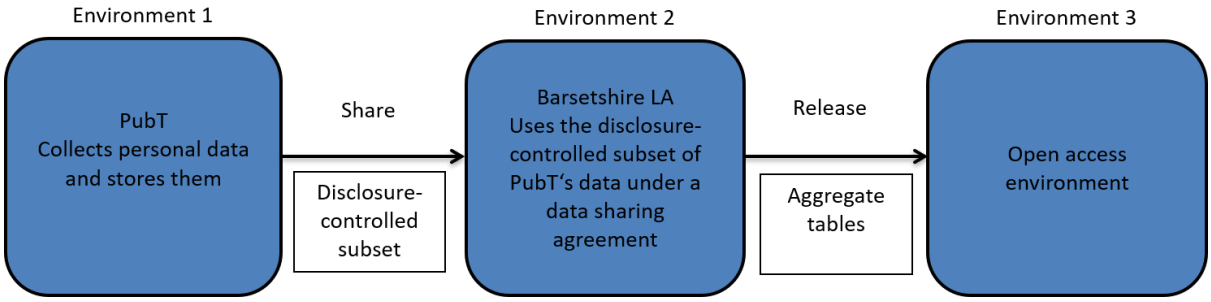


Figure 6: Data flow between multiple environments

The data flow between PubT, Barsetshire and the open access environment defines the parameters of Barsetshire's data situation for the anonymised public transport data.

Barsetshire, as stipulated in its Data Sharing Agreement with PubT, cannot release the anonymised data given to it in its original form without permission from PubT. As joint data controllers, Barsetshire and/or PubT should: (i) carry out a disclosure risk analysis of the intended release of output into the open access environment and (ii) if necessary employ further disclosure control to the data in order to reduce the risk of data to negligibility. The exact process for this will be specified in the Data Sharing Agreement but it would be common practice for Barsetshire to carry out risk assessments of its own before proposing specific outputs to PubT.

2.4 Example data situation 4: simple share with secondary controlled release

In this example, we consider the data flow across environments involving personal and de-identified[23] data. Imagine that Barsetshire collects public health data for its area. It has powers under legislation to share some of that data with the Department of Social Affairs (DoSA) to support its work on health promotion. The data share is formalised under a Data Sharing Agreement which stipulates that Barsetshire and DoSA are joint data controllers for those data. This does not necessarily mean that the division of responsibilities will be equal between the two organisations – the Data Sharing Agreement should stipulate which organisation is responsible for what. We will call Barsetshire council environment A.

DoSA as part of its remit for health promotion (and in accordance with its agreement with Barsetshire) creates a de-identified subset of the data and makes it available within a secure setting for reuse by approved and accredited researchers. DoSA is environment B. The secure setting is designed in such a way as to ensure that the data are functionally anonymised. It places restrictions on who can access the data, how they can be accessed, and what auxiliary information can be brought in and out of the secure lab environment. The secure lab is environment C.

[23] The term de-identified as employed here refers to data that have had all direct identifiers removed or masked. This is not a synonym for 'anonymised' and de-identified data may still be identifiable. Nor is it the same thing as 'pseudonymised' which has a broader technical meaning within GDPR.

An approved and accredited researcher carries out her data analysis in the secure lab producing statistical output, such as regression models, that she will need as she writes up her research. These outputs are first checked by secure lab staff to ensure that they are not disclosive, in which case they are passed as 'safe'. The researcher duly writes up and openly publishes her research, which contains some of the analytical output. The publication platform of the research is a fourth environment, which we call environment D.

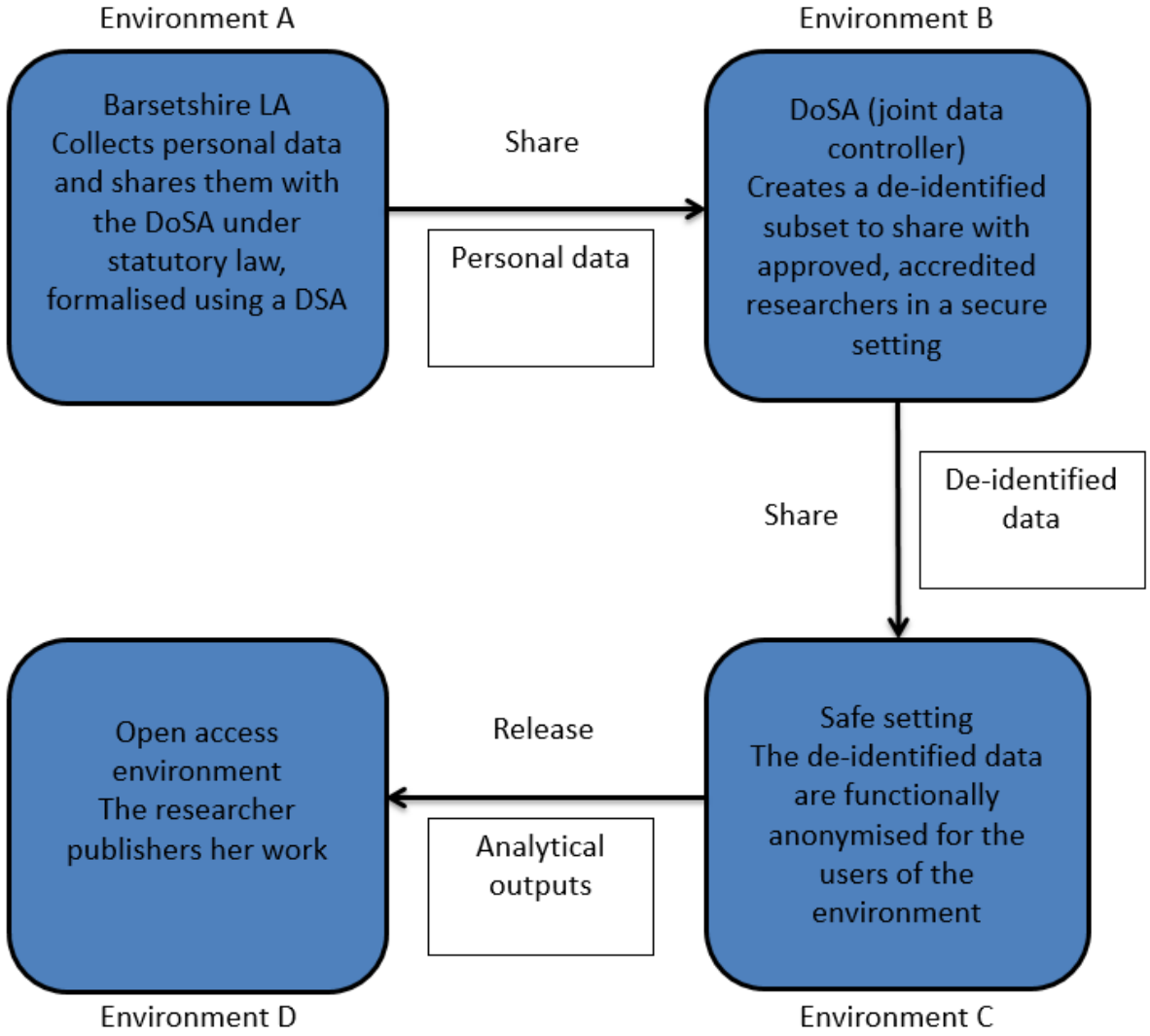


Figure 7: Movement of data across multiple environments

As in the first two examples, one of the key issues which we particularly want to highlight is that data in one environment may be considered functionally anonymised, but in a different environment (such as the researcher's publication) this may no longer be the case. Hence, in this example the researcher's analytical outputs have to be checked and verified as 'safe' by the secure lab before she can take the data away with her.[24]

[24] The set-up we are describing here falls within the core purpose of the *Five Safes* model (Ritchie 2017, Arbuckle & Ritchie 2019).

It is worth stressing that whilst a data situation might be complex, it should not be considered a problem so intractable that you feel it safer not to *even consider* sharing or releasing your data. Of course, that might be the conclusion that you come to after you have worked through the ADF but it should not be the starting position. You should not lose sight of the enormous range of benefits that can and do come from sharing and opening up data.

2.5 Refining your focus

Often the data flows that you map will turn out to be a lot more complicated than you first envisaged and this can be daunting. Sometimes the mapping process will highlight what you already knew: that the situation that you are in is complicated. One way of seeing through the complexity is to refine the focus of your analysis.

Let us define your *focal data situation* as the data and environment that is the object of concern for your analysis. For example, if you are considering sharing the data with another organisation, that organisation will be the *focal data environment*, and the movement of the data you are considering sharing into that environment will be the focal data situation. Hence, your goal here is to refine your focus appropriately, and to think seriously about the focal data situation. In order to do this, you need to consider three further constructs.

- **Locus of Control:** Over which elements of the data flow do you have decisional control and over which do you have operational control? Decisional control indicates the ability to set policies, make rules and regulations for processing. Operational control means the capacity to make day-to-day decisions about processing.
- **Locus of Responsibility:** What elements of the data flow are your direct responsibility and over which do you have indirect responsibility? Responsibilities may come through legal requirements, ethical considerations, and/or agreements. Direct responsibilities are ones where only your own actions are in scope. Indirect responsibility is where the actions of others are in scope. Identifying these responsibilities will enable you to clarify which of these applies.
- **Relevance:** What elements of the data flow affect the focal data situation?

Note that the notions of responsibility and control are related to the GDPR concepts of 'data controller' and 'data processor' but they are not definitive of them. Often a processor will have operational control and direct responsibility but not decisional control or indirect responsibility. However, situations do arise where processors delegate or subcontract processing tasks to sub-processors, in which case they will also incur some indirect responsibilities. A controller has, by definition, decisional control and may have operational control and direct and/or indirect responsibility. A second point is that the notions of control and responsibility apply to data whether it is personal or not whereas the controller/processor distinction is explicitly tied to personal data.

If you consider data situation 1, PubT has decisional and operational control and direct responsibility over its own data, and Barsetshire has some decisional control and indirect responsibility. Over the functionally anonymised version of the data

Barsetshire has operational and some decisional control and direct responsibility and PubT has decisional control and indirect responsibility. This sort of mirror image is one pattern that will be quite common in controller-to-controller arrangements.

A key point at this stage is to ensure that your locus of control and your locus of responsibility align. If they do not then you may have a problem for which anonymisation is not the solution. Consider the following example: The Government Office for National Data (GOND) passes data to the National Research Data Service (NRDS), a unit at the University of Barsetshire, which, amongst other functions, provides access to those data to researchers under controlled conditions. In this situation, GOND has indirect responsibility and strategic control over the release of data for publication (i.e. into the open environment). NRDS on the other hand has operational control and direct responsibility for such releases. So here, we have alignment of responsibility and control and we can therefore define the anonymisation questions for both entities for the data situation regarding the data release.

GOND's anonymisation question concerns the policy for output disclosure control for data release: should it be the same for all data or should it contain procedures for determining policy on a case-by-case basis? NRDS's anonymisation question is how it should operate the GOND policy (whether general or case-by-case). How much resource does it need to do that? What training does it need for staff who will be involved in checking outputs? It may be that GOND specifies some of those elements in the policy, for example that the training of output checkers should be of a particular form. They might do this because they perceive that it is necessary to meet their legal responsibilities or perhaps because stakeholder engagement (component 5) has told them that practice harmonisation across a range of services is desirable. This does not create an intrinsic problem, it is simply a small redefinition of the anonymisation problem to be solved.

The final element that needs to be considered in determining focus is *relevance*. A data situation may be relevant if it affects the risk of the current focal data environment but falls outside the locus of control. GOND may also produce statistical output from the data in question which it releases into the public domain; this is relevant because any output produced by a researcher using the NRDS might overlap with this output, which raises the possibility of both differencing and reconstruction attacks on the output. Note that in considering this possibility, GOND must consider both its own actual data releases and the potential data releases by researchers using the NRDS, as both lie within its loci of control and responsibility. In principle, it could decide to cease or restrict its own future data releases. However, for the current anonymisation question, it should take into account its own data releases when setting the policies for the operation of the NRDS. For NRDS, GOND's data releases are relevant to its decision making, but it has no control over or responsibility for them. The extent to which it has to take them into account may be explicitly represented in the policy that GOND has set for the output disclosure, or it may be that doing so is simple good practice. It may decide to carry out intermittent checks on the output that is being produced from its lab against the possibility of reconstruction attacks.

In one sense, GOND data releases are themselves part of the global data environment. However, they are a very particular piece of that environment because they are drawn from the same data source as the outputs from the data service and so present a very particular threat. In short, NRDS will mark GOND's data as especially *relevant* as they represent a very particular test of its operational practice.

So by the end of this component, using the concepts of responsibility, control and relevance you will have refined your focal data situation and by doing that clarified the specific anonymisation question that you want to address.

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

Overview: *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.*

* * *

The environments in which data are stored and otherwise processed have a large – if difficult to quantify – impact on risk. In this component, we specify each environment in the data flow according to the four elements. Depending on its complexity, this might not be possible at the same level of detail all along the data flow. This will relate to your focus. If an environment is outside of your locus of control then you simply may not have access to all of the information. However, it is vital that you should be able to specify in full detail all of the environments within your focus and any elements that you have determined in component 2 to be relevant (i.e. having a direct impact on risk or sensitivity). The exception to this is where the data environment is not yet precisely defined for a proposed share or release, i.e. those situations where the determination and construction of the properties of the destination environment is part of the presenting problem.

3.1 Features of the environment

The environment features you need to specify are the following:

- **Other data** are any information that could be linked to the data in question, thereby enabling re-identification. There are four key types of other data: personal knowledge, publicly available sources, restricted access data sources, and other similar data releases. A vital question is whether those other data are themselves identified or identifiable.
- **Agents** are those people and entities capable of acting on the data and interacting with it along any point in a data flow.
- **Governance processes** constrain how agents' relationships with the data are managed. This includes formal governance such as data access controls, licensing arrangements and policies which prescribe and proscribe agents'

interactions and informal behaviour through norms and practices, for example risk aversion, a culture of prioritising data privacy, etc.

- **Infrastructure** denotes the structures and facilities that allow the data to flow and shape the data environment, including security infrastructure and wider social and economic structures. At its narrowest level, infrastructure is best thought of as the set of interconnecting structures (physical, technical) and processes (organisational, managerial) that frame the data environment. At its broadest level, infrastructure can include intangible structures, such as political, economic and social structures, that influence the evolution of technologies for data exploitation, as well as data access, sharing and protection practices.

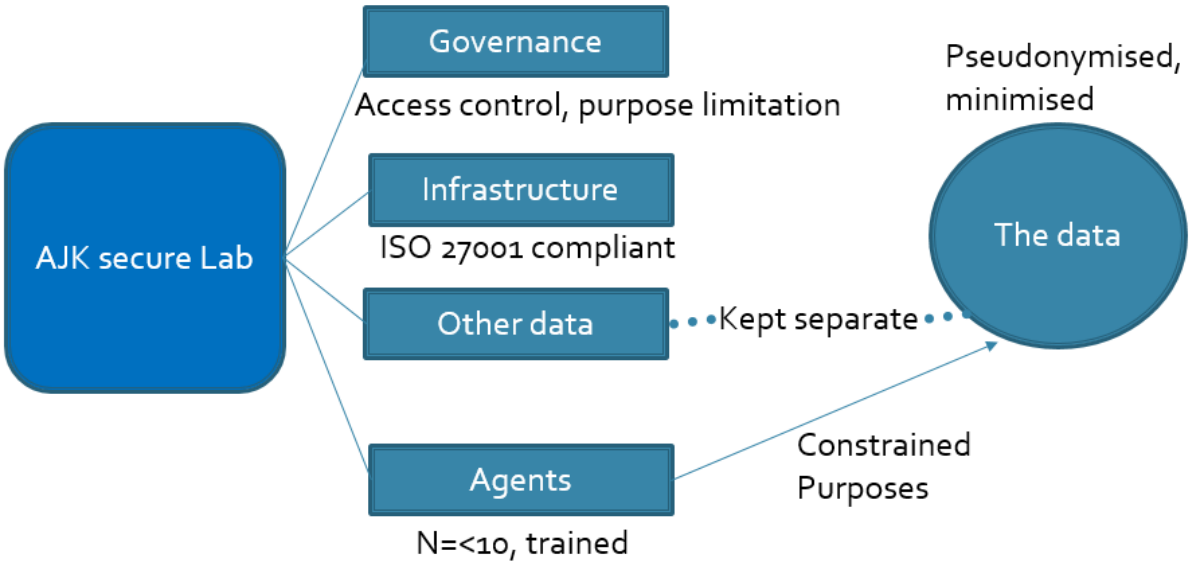


Figure 8: An example specification of a data environment

Figure 8 shows an example of what a first pass at an environment specification for a secure data lab might look like. As a prescription of a prescribed share, it would feed into the evaluation to be done in component 6 and/or the risk analysis that might be done in component 7 in one of two ways. Either (i) it could be used to specify/constrain plausible intrusion scenarios, or (ii) it could be used for the purposes of carrying out comparative data situation analyses.

3.2 Capturing the environment features

Table 2 shows a completed example of the template for capturing environment features which you can find on the resources page of the UKAN web site. This will help you capture the properties of each of the environments along your data flow.

Data environment properties	**Examples**	**AJK secure Lab**
Other data	* Publicly accessible sources e.g. registers, media * Restricted access sources	The data is kept completely separate from all other data on a standalone server.

Data environment properties	Examples	AJK secure Lab
	* Other similar data releases	
Agents	* Specified group * Small group contained or dispersed across teams or organisation * Large group contained or dispersed across teams or organisation * No specified group, i.e. potentially anyone in the world, open data	No more than ten staff will access the data whilst it is in the lab. They are all trained in data protection and disclosure risk for outputs and have received specific training regarding lab expectations.
Governance	Policies and procedures: access controls, purpose limitation, breaches, output checking, authentication, data destruction	A full set of policies and procedures are in place covering these topics.
Infrastructure	* Hardware * Software * Physical security	The infrastructure has been assessed as being ISO27001 compliant, and certification has been applied for.

Table 2: An example template for capturing data environment properties

Component 4: Describe and Map the Data

Overview: *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.[25]

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

[25] For a fuller discussion on processing special category data, see the ICO's *Lawful Basis for Processing Special Category Data*, https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/special-category-data-1-0.pdf.

and either legal or official authority under Article 10. The DPA sets out the specific conditions providing lawful authority.[26]

* * *

When thinking about whether, and how, to share or release your data safely, a key consideration will obviously be the data themselves. In this component, we set out a top-level examination of data focusing on the data's type and properties. Identifying these features will be relevant for both components 6 and 7 of the ADF later. In addition, it is possible at this stage to make some straightforward decisions that will simplify the more detailed processes you will go through later. However, the main purpose of this component is to get a picture of the data in much the same way that a data analyst might explore a dataset before building a more complex statistical model.

As we will illustrate, the data features *status*, *type* and *properties* are central to the issue of disclosure control, whilst other points are useful indicators as to the *summary likelihood;* a term we will define in more detail in component 6.

1. **Data subjects:** Who are the data about and what is their relationship with the data?
2. **Data type:** What form are your data in, e.g. statistics or text? What level of information about the population do the data provide, e.g. are they microdata or aggregated?
3. **Variable/information types:** What is the nature of each variable within the data? For instance, are they direct identifiers, indirect identifiers or targets? This will be a more complex matter if the data are unstructured. In that case you may want to grade variables as definitely present, probably present and possibly present.
4. **Dataset properties:** What are the top-level properties of the dataset, e.g. its age, quality, etc.?

We consider now each of these four features in turn and in doing so highlight their relevance to the question of anonymisation.

4.1 Data subjects

In most cases, who the data are about is a straightforward question. However, data can be indirectly informative about people when they are directly about something else. Similarly, you should also be mindful that data which are directly about one group of data subjects may also be indirectly informative about another group; for example patient record data for a particular GP practice could be informative about the practice's GPs (e.g. their prescribing practices). Similarly, data about workplaces might be informative about individuals (and *vice versa*). Another complex area is when something somebody has written or produced is personal data about the writer/producer. The recent case involving a student's

[26] For further information on this see ICO's *Guide to the General Data Protection Regulation*, https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf.

exam paper suggests that at least sometimes that such information might constitute personal data, in that case about the student,[27] although the ICO's guidance suggests that it might depend on the details of the processing.[28] A precautionary approach would be to treat these borderline cases as personal unless your jurisdictional regulator indicates otherwise.

You should also consider whether the data subjects in the collective represent a vulnerable group and the extent to which they have given consent to any of the data processing involved in your data situation and/or the extent to which they are aware of it.

4.2 Data type: what type of data do I have and what type of data should I share/release?

If you have collected your own data about people then it is likely to be in the form of individual unit records, or *microdata*. Such data are commonly stored in digital databases as single records of information where the rows represent a single population unit (person, household, etc.) and the columns represent the information (variables) you have collected about them. For the purpose of sharing or releasing data you may decide not to make available an anonymised version of the microdata, but instead to aggregate your data and make it available as an anonymised table, graph or map. To assist you in making decisions about what type of data to share or release let us consider the particular disclosure risks associated with each.

For aggregate data, particularly for small geographical areas such as census output areas or postcode sectors, attribution disclosure[29] and disclosure by differencing[30] are considered to be particularly problematic (Smith & Elliot 2008, Duncan et al 2011).

For microdata, re-identification disclosure is a particularly challenging problem because of the difficulty of determining which variables or combinations of variables might make an individual unique in a dataset and therefore stand out as vulnerable to re-identification (we consider this further in the next subsection).

4.3 Variable type

Variables can be identifiers (direct or indirect), or targets. Most datasets will have a mix of all three types. At this stage you are not attempting to form fully specified scenarios but simply explore the data, sorting variables into appropriate types. Some variables are considered to be standard indirect identifiers – for example sex and age are routinely included in most key variable sets, while others you may

[27] https://www.irishtimes.com/news/ireland/irish-news/irish-student-wins-support-in-legal-battle-over-exam-script-1.3161854.
[28] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-the-meaning-of-relates-to/#pd5.
[29] This is where there are zeroes present or where they are inferable from some combination of the variables in the aggregated data.
[30] This is where different codings for the same variable are overlain to produce intersecting categories with small numbers.

not yet be sure about. The purpose of this is to understand the scope of the anonymisation problem.

A *target variable* is usually one that a reasonable person would consider to be sensitive and not widely available (and therefore might well be the target for an intruder). In general, special category data is often considered sensitive and therefore a target. Identifying the target variables will inform you about likely harm that will arise from a disclosure and may also inform the construction of disclosure scenarios (see component 7). Examples include a personal salary, or sensitive medical data such as HIV status.

The overarching point is that if you are dealing with sensitive data then the risk is higher both in terms of the likelihood of a deliberate attempt to access the data and the impact of an attempt if successful. As well as impacts on the data subjects, unintended disclosure of sensitive data is also likely to be more damaging to the data controller than disclosure of non-sensitive data, because the impact on public trust and reputation is likely to be greater. If you do not have data subjects' consent to release or share a disclosure-controlled version of data containing sensitive data then the risk to your organisation's reputation, in the event of a subsequent breach, is amplified further. Put another way, if you do not have consent then your overall data situation is more sensitive.

Direct identifiers are any attributes or combination of attributes that are structurally unique for all persons in your data, such as unique reference numbers like NHS numbers or US social security numbers. In most dynamic data situations you will suppress (or possibly remove or replace) the direct identifiers as a first step. Therefore, being clear about which variables are direct identifiers is important.[31]

	Name	Address	Hospital ID	Gender	DOB	Date of admission	Medical test 1	Medical test 2
Direct?	Yes	Yes	Yes					
Indirect?				Yes	Yes	Possibly	Possibly	Possibly
Target?						Possibly	Possibly	Possibly

Table 3: Some example variables from medical data

Indirect identifiers in contrast can be any attributes or, more likely, combination of attributes that, whilst not structurally unique, are likely to be unique for at least some individuals in your dataset *and* in the population. An example of indirect identifiers might be the combination of age, marital status and location variables. Whilst these are not immediately obvious identifiers individually or even in combination (a single 43-year-old in London, say), a single 43-year old at a particular postcode could easily be unique. A sixteen year old widower living in rural Scotland will almost certainly be unique relative to this rare combination of attributes and thus at far greater risk of re-identification. The important point is that rare combinations can crop up and create a risk of someone spontaneously re-identifying individuals. Some variables might be classified as indirect variables

[31] For a more detailed discussion about what might constitute a direct identifier see the 1st edition ADF book pages 17-18.

or not depending on the scenario (see component 7 for a discussion of scenario construction).

4.4 Dataset properties

The properties of a dataset can *potentially* increase or decrease the risk of disclosure. We say 'potentially' because, as acknowledged, at this stage we are doing no more than getting to know the data without reference to the data environment.

The use of a general risk indicator, such as the one described below, merely acts as a guide to help you think about which data properties require particular attention when you are doing further analysis. It is a precursor to, not a substitute for, the requirement for a careful analysis of your data.

- **Data quality:** It is generally accepted that all data will contain some level of error. Error can originate from the data subject, data collector and/or the data collection process (Bateson 1984). In general, as a data holder, your aim is to ensure that the error level in your data is small; after all there is little point in sharing or releasing data that do not represent whatever they are supposed to represent well because they are distorted by their low quality. However, a small level of error, inherent in all data, has some advantages as it offers a degree of natural data protection (Duncan et al 2011) and indeed some disclosure control measures introduce error for precisely this reason.
- **Age of data:** The older the data, the harder it is to identify people correctly from them.[32] This is because people's circumstances change over time as, for example, they move location, change jobs or get married. Thus, older data may also acquire a more basic level of data protection.
- **Hierarchical data**: This type of data contains information for members of a group who are linked with one another and is a common source of disclosure risk in business or household data. The data are considered riskier because they provide (more) information that might make a data subject unique in a dataset and as such potentially identifiable. For example, the combination of age and sex of all members of a household will be unique for most households above a relatively modest size (Duncan et al 2011).
- **Longitudinal data**: This is data about a defined population that have been collected over time and then linked. These are considered riskier because of the potential to capture possibly unique changes in information over time such as seemingly unremarkable changes to a data subject's marital status, economic position, employment, health or location that may stand out against the longitudinal patterns of other data subjects. Again, this may

[32] The inverse problem with older data relates to an increased risk of associating incorrect information with people identified within a dataset because the information is out of date.

increase the likelihood of a data subject being unique in a dataset and as such potentially identifiable.[33]

- **Population or sample data**: Population data includes census data and data for all people in a particular administrative group such as benefit claimants or hospital patients. It is considered more risky because there will be little uncertainty as to who is represented in the dataset.

4.5 Capturing the data features

In the ADF companion document *Data Features Template*,[34] you will find a tool for capturing the above features and perhaps recording any top level actions that you might make. For example, you could decide that you will release a flat rather than hierarchical file or that date of birth will be recoded to single year of age. These decisions simplify the technical work required (in component 7).

[33] Analytically, longitudinal data could be treated as longwave dynamic data (i.e. the slowest form of data that updates over time). In practice, however, they are analysed as static datasets and therefore in most data situations the longitudinal element is treated as another property of the dataset. It does however lead to some different intrusion scenarios as it is necessary to consider the likelihood of an intruder also having access to longitudinal data.

[34] All of the ADF companion documents can be found on the ADF page of the UKAN website (www.ukanon.net).

Component 5: Engage With Stakeholders

Overview: *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.*

* * *

Effective communication can help build trust and credibility, both of which are critical to difficult situations where you need to be heard, understood and believed. You will be better placed to manage the impact of a disclosure if you and your

stakeholders have developed a good working relationship, and you understand their expectations of you. This is a key point worth emphasising: *effective engagement with stakeholders enables you to constrain the impact of a disclosure*. If we think of risk as a function of the likelihood of an event happening and the negative impact if it did, engagement reduces the risk by reducing the second input to the risk function. Indeed, the comments of stakeholders may also help with the first input too, if it brings you more information (for example, about the data about them in other datasets, especially public ones). Engagement is a necessary part of a functional anonymisation process. At the extreme, it may be that once you have received and come to understand stakeholders' objections to the data share, you decide to scale back or even cancel your plans.

For the purposes of the ADF, your stakeholders are the actors who might influence your data situation, with some ability to alter the data flow within the various environments. They therefore have some capacity to make your job easier or harder. Their interaction may be direct or indirect. For instance, a data subject can affect the flow of data very directly, by refusing or withdrawing consent for some processing to go ahead. On the other hand, an investor might be influential over a longer timescale; if media stories of poor data management cause her to lose confidence, then the organisation's resources may be reduced. Your clients and partners may help create new data environments within your data situation, or, if they have concerns about your data management, may close down or restrict your freedom within existing environments.

The process of developing your data situation audit will help identify who your stakeholders are. Your clients and customers are not the only groups with an interest in your business or activity. Others who may be affected include data subjects within your datasets, those sharing or wishing to share your data, the general public, partner organisations, investors, people in other units within your organisation, the media, funders, regulators and special interest groups (e.g. privacy advocates, open data advocates, consumer representatives, etc.). Any of these stakeholders may affect your data situation in positive and negative ways, at different scales and over different timescales.

The sector you are in will also bring its own set of stakeholders with it. So, for example, stakeholders in the health sector in the UK will include groups such as the Department of Health and Social Care, local authorities, hospital trusts, patients and patients' groups, service users, suppliers, funders, commissioning groups, quality assessors, special interest groups, community groups, the wider public health workforce, pharmaceutical companies and the media. These stakeholders may have conflicting interests in the data that might be released, which will clearly impact on your data situation.

5.1 The importance of developing trust with stakeholders

Earning and maintaining the trust of your stakeholders involves two related aspects, which sometimes are difficult to balance. The first is being trustworthy, and true to your word. The second is persuading others that you are trustworthy; their trust is basically their belief that you are trustworthy. The first involves your decisions around data management, including risk management and anonymisation decision-making, and so can be delivered technically, through

conscientious practice. You control your trustworthiness. The second, which is the focus of this component, requires you to craft an accurate narrative which expresses and communicates your trustworthiness to your stakeholders, so that they understand that your actions are trustworthy. You cannot control someone's trust in you, although you can influence it (for good or ill). You have to earn it. Indeed, the crafting of the narrative may be an iterative process performed in collaboration with your stakeholders, as you find out what they consider trustworthy and adapt your aspirations accordingly.

It is quite possible to be trustworthy while your stakeholders do not believe you to be so, and you have to bear the cost of this failure to communicate. Or, nearly as bad, if your stakeholders trust you for the wrong reasons (they have misunderstood your intentions for the data, through some kind of miscommunication), then their trust will not be robust, and could be shattered in a moment once the reality of what you are doing emerges. Hence your narrative needs to be full and accurate, and you need to listen to your stakeholders. Don't say just what they want to hear, but listen to their concerns and address them.

Being trustworthy and earning trust involve working to deliver your commitments to others. Naturally, you are committed to obeying the law, and legal compliance, particularly with GDPR, is an important part of the narrative of your trustworthiness, but it cannot be the whole story. Your stakeholders will also expect that you will take their interests into account to a proportionate extent when you make your decisions about processing the data. It is therefore important that you understand their interests and how they expect you to behave.

Trustworthiness is a virtue; it is a vital part of ethical data processing, about which much has been written recently, particularly in the context of ethical AI or ML (Jobin et al 2019). However, the ADF is not so much about ethical data processing, as ensuring that your data situation is manageable and tractable. It so happens that if, in addition to observing the law, you behave ethically beyond the requirements of the law and show respect for your stakeholders, then your life as a data manager will be easier and more rewarding, so behaving ethically is a win-win.

5.2 About your stakeholders

Depending on the circumstances and the public interest in your data, many if not all of your stakeholders are likely to have an interest in your data, their use and reuse, whether confidentiality is a high priority in your organisation, and whether assurances of confidentiality are well-founded. However, what they would like to hear about these topics may differ. For example, data subjects and the wider public are likely to want to know the *what* of your processing activities, such as what data, in which environment(s). In contrast, specialist interest groups and the media may also want to know about the *how* of your processing activities, such as how the data are manipulated or how you determine an environment to be safe. The key is to engage with your stakeholders to determine what they would like to know about your processing activities, most obviously so that you can put your point of view across (and perhaps adjust your practices in response to reasonable criticism), but also so that you understand their information needs

immediately when you find you have to pick up the phone. You can do this in much the same way that you engage with your user groups by, for example:

- **A web or mail survey:** You could develop a short survey to be delivered via your website or through a mail-out. Bear in mind you may need to tailor the survey to different stakeholder groups.
- **Going out and talking:** You may want to tailor the mode of discussion for particular target stakeholders, e.g. holding face-to-face meetings with funders, holding focus groups with representatives of civil society or the general public, etc. Networking with your peers in data management will usually reveal what issues are concerning them; if enough of your peers have the same problem, then it may become your problem soon enough. They may also be able to share solutions.
- **A little research:** One way to identify concerns is to look at the type of FOI requests you and similar organisations in your sector receive. Identify common themes and whether particular stakeholder groups are associated with particular themes.

You should try to understand your stakeholders' points of view – they are not the enemy, although they are not always your friend! They have their own agendas, which they will rightly prioritise and pursue. You need answers to important questions such as:

- Are their motivations financial, business-related, political, emotional, or something else? You need to communicate to them relevantly (e.g. telling a privacy group that your activities will increase your profits is unlikely to be persuasive to it, whereas an investor might quite like that argument).
- Are they (generally) supportive of or negative towards your organisation? Messages to someone who is generally negative will typically need to be handled with care, as you cannot expect goodwill or cooperation; messages might be passed on with a damaging spin. Messages to a supportive stakeholder can take more common ground for granted. Some words can trigger negative reactions in unsympathetic interlocutors: 'risk', 'threat', 'exploitation'. And always think what the message would look like quoted out of context.
- What information do they (already) have, and what might they know that you do not know? This will affect the detail of your messaging, and the amount of explanation it has to include.
- How do they receive and consume information? No point tweeting to an individual who is not on Twitter. If you want feedback or a conversation, it's a bad idea to write a letter.
- Are they an organisation, a particular individual, a group of concerned individuals, or a diverse cross-section of a heterogeneous population? An organisation will have specified contact points; an individual will have preferences that you can find out. Less formal groups may need a means of collective communication, such as via a website, social media or an

advertisement. You need to determine whether it is possible to get any feedback from a less formal group at all.

- If they are representatives of a wider population, whom do they represent? How legitimate are they? Your messaging needs to reflect the interests and concerns of the population, but also take into account the particular attributes of the representatives. If they are self-appointed and not legitimate, do you need to engage with them at all?
- What ability do they have to affect your data situation? Directly, or indirectly? Substantially or marginally? In the short term or over longer timescales? These questions affect the importance of engaging with them sooner rather than later, and the effort you put into engagement. If their effect is indirect, then the conversation you have with them might be more protracted, and less focused. If their effect can be significant and short-term, then you may need to have phone numbers or emails addresses to hand for an emergency.
- How interested are they in you? Do your actions crucially affect their workflow? Or will you have to work hard to get their attention? If you want their feedback, you need to put effort and resource into getting it. If they mean more to you than you mean to them, then getting their attention and persuading them to engage is a priority.
- Are they familiar with your sector, your language, your jargon and your assumptions? Do they share them? If they do, you can use your common language. If they do not, you will have to use plain language and spell things out.
- Are you able to converse in two-way interaction, or will you have to broadcast to a large group? How easily and quickly can you assess the response to your messaging? An intimate conversation allows nuance. Immediate feedback means you can correct misunderstandings quickly. A broadcast needs to be correct first time, because the message can't be withdrawn, and clarifications may not be picked up by the same audience. A broadcast message also needs to be simple and clear.
- What risks are there in engaging with them? In not engaging? The greater the risk of engagement the more carefully it needs to be done, or maybe just don't bother.

Determining the next step once you know what your various stakeholders want to hear from you may or may not be straightforward. As we have already said, being open and transparent is always preferable but you may not be able to meet all your stakeholders' requests for information, either because they affect your disclosure control or because they create their own confidentiality issues. On the other hand, you will have had to draw up a Data Protection Impact Assessment for the processing, and many corporate data processors are likely to find themselves under pressure to release their DPIAs in future if data privacy remains a critical issue or vulnerability. You certainly should not assume that a DPIA will necessarily remain confidential.

5.3 Planning communication and engagement

Plan how you will talk to and engage with your various stakeholders. Below is a list of pointers that you may wish to capture in your engagement plan (it is not an exhaustive list).

5.3.1 Identify your key stakeholders

This is an obvious point but you need to make sure you capture all those likely to have a stake in your data processing activities; this might be a wide range of groups. We have listed common types of stakeholder above, although the final list will be dependent on your activities, your organisation, the sector you belong to etc.

You are the judge of who the important stakeholders are, and how central they are to your operations. Your need for trust is going to be roughly proportionate to the sensitivity of your processing and the vulnerability of that processing to the withdrawal of stakeholder support. And not all stakeholders are equal – adjust your efforts accordingly, and prioritise the most significant. Seek out critical friends, who are basically supportive but who will constructively point out both the good and the bad points of an approach.

5.3.2 Be clear about your aims and objectives in talking to your stakeholders

This will help you ensure that your messages are clear and consistent between themselves and also with other messages coming from different units in your organisation. You may have multiple aims, such as: (i) to promote trust in your organisation's handling of data, (ii) to build relations with relevant specialist groups, and (iii) to promote awareness about your reuse of data for public benefit. An important distinction is whether you are being transparent about what you hope to do in order to forestall problems going forward, or whether you are weighing up the risks of sharing and are concerned to understand potential objections to your plans. In the former case, then you should concentrate on broadcasting your message. In the latter case, you need more of a confidential two-way conversation, and be prepared to listen.

Your stakeholders' trust in you will be shaped by their interests and expectations, but that does not mean that you have to do everything they want! Trustworthiness involves being true to your word, not being all things to all people. It involves being conscientious and thinking of the needs of your stakeholders when you take your decisions, and ensuring that the business case for processing extends beyond your own bottom line.

The importance of your various stakeholders to your operations will help determine the level of engagement that is appropriate for each of them.

- **High level (Interaction):** At this level, you are likely to include the stakeholder in some of your decision-making.
- **Medium level (Consultation):** Here, you will take note of the opinions of the stakeholder, and factor them into your decision-making (you may ultimately end up discounting them, but not without some consideration).

- **Low level (Notification):** At the lowest level, the stakeholder's opinion is not highly salient, although you are prepared to make the effort to keep them informed.

5.3.3 Establish your key messages

This is critical to the effectiveness of your communications. Your key messages need to be clear and concise and address the concerns of your stakeholders. Of course they need to be consistent with the other messages of your organisation, and if your organisation is large enough, it makes a lot of sense to coordinate messaging with your comms team (which will also contain a good deal of expertise about how to reach target audiences). It will not help build trust if your messages contradict your organisation's wider marketing pitch, for example.

It is common sense to adapt your messaging to different audiences, in terms of the medium of communication, the language used, and the complexity of the messages themselves. A deposition detailed enough to satisfy a regulator would in all likelihood bore or baffle the average data subject who was merely concerned that his or her personal data was not being misused (even if the deposition addressed those concerns).

While tailoring messages, be careful not to be inconsistent. Though your messages of necessity can only paint partial pictures of your activities, they should at least all be pictures of the same thing. They should not mislead, or be contradictory. There are obvious ethical arguments against misleading your stakeholders, but they are bolstered by the pragmatic observation that someone who sends out untrustworthy messages, even in error, is highly unlikely to sustain trust.

5.3.4 About communication and public engagement activities

Your communication and public engagement activities should have clear timescales and goals to allow you to evaluate their effectiveness. The resources, both financial and human, devoted to them should be sufficient for the purpose. And you need the approval of the rest of your organisation.

Examples of communications and engagement activities might include:

- **Press releases:** A concise press release can help you reach a large audience with little financial outlay.
- **Social media:** Regular and committed use of social networking, such as Facebook or Twitter, allows you to communicate in real time.
- **Actively maintain a website:** This will allow you to provide consistent messages over time, accessible to all (or most of) your stakeholders.[35]
- **Involvement and consultation activities:** Going out and meeting your stakeholders, by holding focus groups, meetings, briefings and discussion forums etc., allows personal and face-to-face contacts to develop, which in

[35] Examples where a lot of thought has been given to the key issues of public benefit and trust can be found at www.adrn.ac.uk and www.datasaveslives.eu.

many circumstances is more supportive of trust than a purely corporate outward face.

The best methods for engagement will of course depend on the stakeholders themselves. Summarised in Table 4 are some of the properties of a few means of engagement. We do not pretend this is a comprehensive list.

Method	**Communication type**	**Reach**	**Audience**	**Resource level**
Press release	Sending	Depends on whether media carry the message	Large for mass media; informed for specialist media	Low
Website or page on the existing corporate website	Largely sending, but usually includes a 'contact us' link	Universal, especially if widely linked to, thereby increasing its search rank	Potentially large, but generally found via search, so requires an interested audience	Initial costs are front-loaded in the construction of the site or page, but don't underestimate maintenance costs
Generic email address	Receiving and replying, answering queries and fielding complaints	Universal	Usually found by browsing through a website's 'contact us' page; audience therefore typically concerned	Low to set up, may need more resource to respond if it is well-used
Social media	Sending messages, receiving replies, relinquishing control of the messages as others respond, retweet, etc.	Almost universal. Major social media events are often covered in mass media	Can use hashtags to allow interested audience to self-select	Low, but you need to keep messages coming. Potential for a crisis requiring a swift hands-on response if a message goes viral for the wrong reasons
FAQs	Sending (although hopefully informed by a genuine assessment of which questions	Probably on the website, so universal, but also depending how easily the reader can	Interested audience first must find the website and then navigate to the FAQs	Low, after the page is written. Updates will be needed every so often

Method	Communication type	Reach	Audience	Resource level
	are frequently asked)	access the FAQ page		
Privacy or Data Protection Impact Assessment	Sending	Narrow, although you may wish to publish in the interests of transparency.	Specialist	High, but DPIAs are required under some legal regimes (GDPR in particular), and so may not require extra resource
Focus groups	Receiving	Depends on how representative the group is	Representative	High
Advertising	Sending	Potentially universal, depending on medium. Could be highly targeted for online advertising	A range of audience types can be targeted	Medium to high
Public hearings	Sending and receiving	Local – usually involves physical presence in a room, although can also be run online	Specialised, interested	High, although the expense can be offset by running a series of similar events in different locations
Attending or presenting at conferences and events	Sending specific messages, and receiving information about the wider context	Narrow	Specialised	Fairly high
Polling	Receiving	Depends on poll sample	Depends on poll sample	Fairly high but usually outsourced
Citizens' juries	Receiving more than sending	Narrow for direct reach, but the output can be published	Interested but non-specialist	High

Method	Communication type	Reach	Audience	Resource level
Face to face meeting, business lunch	Sending and receiving equally	Directly restricted to a few selected individuals, but messaging may ripple through social networks	Restricted to participants, but highly interested and specialist	Low (depending on the quality of wine chosen)
Testimonials	Sending	Wide, if placed on website	The most receptive audience will be those to whom the testimonial providers are known	Low

Table 4: Means of engaging stakeholders

One final and very general point about communication is that by promoting trust, building relations and promoting good works you will be helping to associate a positive view with your organisation's use of data. This is important because you are operating in a complex global data environment over which you have limited control and bad news stories about data breaches and data security mishaps are all too frequent. If there has been a recent, widely-publicised data breach elsewhere in your sector, it may be that you, even though blameless, will be closely scrutinised by the media, or political campaigners.

Component 6: Evaluate the Data Situation

Overview: *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? To do this you can use the tool which can be found in the ADF Companion document* Data Situation Evaluation Tool*.*

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 the new obligation 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 Protection36 and the ICO.37 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

---

[36] http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236.
[37] https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias-1-0.pdf.

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.

* * *

To complete the data situation audit you need to evaluate it. Your top-level goal is to answer a two-part question: is there an anonymisation problem to solve and if so what is it? Recalling our introductory goal was to answer three primary questions, we now already have answers to PQ1 and PQ2. We are now left with the requirement to answer PQ3, addressing the issue of the sensitivity of your data situation, and therefore the likely impact of an incorrect decision.

We have made available on our website a spreadsheet tool (in Excel©) which does all of the calculations for you and arrives at the overall assessment but it is perfectly possible to calculate these manually and we also provide a template document for those who prefer that. Here we talk through the questions and (simple) logic of how the tool is constructed so that it is a black box. As with the ADF the tool itself will be subject to refinement and update.

6.1 Determining data situation sensitivity

Data situation sensitivity is comprised of three components: consent, expectations and data. We will look at each of these in turn before considering how they may be collated to provide an overall assessment.

6.1.1 Consent sensitivity

There are numerous reasons why data subjects might object to their data being reused. This brings us to the thorny issue of consent. To clear up one issue before we start, we are not thinking about consent in its technical legal sense (under GDPR or other legal instruments). We also recognise that, in law, consent is only one of many legal bases that might be used to justify processing as lawful and fair. However, our consideration here is different to that. The question we are addressing is primarily an ethical one which reflects on the importance of *informational autonomy* on the sensitivity of your data situation.

In principle, consent is a straightforward idea: you ask the data subjects 'can I do X with your data?' and they say yes or no. However, in practice the situation is much more complicated than this. Firstly, consent is layered. Secondly, the notion of consent is interlaced with the notion of awareness. This produces what we refer to as a scale of information autonomy. Consider the following questions:

1. Are the data subjects aware that their data have been collected in the first place?
2. Have the data subjects consented to the collection of their data?
3. Were the data subjects completely free to give consent to the collection of their data or have they agreed to collection because they want something

(a good or service) which they cannot obtain without handing over some data?
4. Are the data subjects aware of the original use of their data?
5. Have the data subjects consented to the original use of their data?
6. Have the data subjects consented in general to the sharing of an anonymised version of their data?
7. Are the data subjects aware of the specific organisations that you are sharing a functionally anonymised version of their data with?
8. Have they consented to your sharing their data with those organisations?
9. Are the data subjects aware of the particular use to which their anonymised data are being put?
10. Have they consented to those uses?

The more noes that you receive for the above questions, the less autonomy the data subjects have. What does this mean in practice? Put simply, the less autonomous data subjects are, the less able they are to take responsibility for what happens to their data and therefore the greater your own responsibility.[38]

The astute reader will have noted that not all (and possibly none) of the questions have straight yes or no answers. Awareness is a nuanced concept. For example, take question 1; I might be generally aware that I am being caught on CCTV every day but not know about every (or even any) specific instance of it. Or I might be aware that I have been caught but not know what happens to the film next and so on. Similarly, I may have *de facto* consented to a particular piece of data processing but not understood what I have consented to. Am I, in that case, even aware that I have consented? So awareness and consent interact.

What does this complex autonomy soup actually mean? You might be expecting us to say at this point that you should be seeking informed consent if at all possible but we are not going to do that. Given the current state of the information society it is both impractical and undesirable. Obtaining consent of any sort is complex. Obtaining real informed consent would – just as a starting point – require massive re-education and awareness campaigns, and even then giving consent for every piece of processing of every piece of data is something that very few people are going to welcome (if you have never ticked the box to agree to T&Cs on a website without having first read them, please get in touch with us, as we would like to know what that is like). This is not to say that well thought out consent processes do not have their place – they most certainly do – but they are not a panacea.

Ok so what is the point here? It is simply this: if you pose the questions above and the answers are mostly in the negative then your data situation is more sensitive. The notion of sensitivity is key; it is a connecting concept which enables clearer thinking about ethics and the reuse of (anonymised) data. We will come on to what you need to do about sensitive data situations shortly but is there anything else that heightens sensitivity?

[38] If you have answered no to <u>any</u> of these questions and you are using consent as your legal basis for processing under GDPR then you may need to reconsider that legal basis.

Now go back to the question and count the number of noes for your focal data situation in the above ten questions. That is the starting score for the data situation sensitivity, which you are going to add to in the next two sections.

6.1.2 Expectation sensitivity

Beyond consent, the question of whether a particular share or release conforms to the data subjects' reasonable expectations is also important. Nissenbaum's description of privacy is useful here. She describes the right to privacy not as a right to secrecy nor as a right to control "but a right to *appropriate* flow of personal information" (2010:127, her emphasis). To help tease out the appropriate flow, and what your stakeholders' expectations may be, we draw (loosely) on Nissenbaum's concept of *contextual integrity*. Contextual integrity is a philosophical approach for understanding expectations in relation to the flow of personal information and can usefully be applied to shed light on why some flows of personal data cause moral outrage. This approach uses the notions of context, roles and data (to be transmitted) as a framing tool for evaluating whether a flow of data is likely to be considered within expectations or outside of (i.e. violating) them.

We argue that the principles of contextual integrity can usefully be applied to the flow of anonymised data for the purpose of helping practitioners to make well thought out and ethically sound decisions about how they reuse data.

To untangle this complex notion for practical use you will need to think in terms of the roles and relationships between you and the proposed receiver of your anonymised data, and the purpose of the share/release. The complexity of the questions you will have to ask yourself will depend on the complexity of your data situation. But here is how they might look for a simple site-to-site share of data.

1. Do you (the sending organisation) have a relationship with the data subjects?
2. Does the receiving organisation have a relationship with the data subjects?
3. Do you and the receiving organisation work in different sectors?
4. Is your organisation's area of work one where trust is operationally important (e.g. health or education)?
5. Is there an actual or likely perceived imbalance of benefit arising from the proposed share or release?

Here the more questions you answer yes to, the more sensitive your data situation is. For each yes add 2 points to your data situation sensitivity score.

6.1.3 Data sensitivity

Finally, the data themselves can have properties that make the data situation more or less sensitive. Five questions capture the main points here.

1. Are some of the variables sensitive?
2. Are the data about a vulnerable population? The term 'vulnerable' usually implies being 'at risk' of harm. A group is generally considered vulnerable

if the individuals in the group may have special difficulty giving free and informed consent to being data subjects.
3. Are the data about a sensitive topic? The topic area might be considered sensitive rather than, or as well as, the variables within the anonymised dataset because, for example, it involves particular public interest issues, or ethically challenging issues.
4. Is the use of the data likely to be considered sensitive?
5. Do you have reason to believe that the intended use of the data might lead to discrimination against the data subjects or a group of which they are members?

A simple litmus test is: would a reasonable person regard these data as sensitive? **Again, the more questions you answer yes to, the more sensitive your data situation. For each yes add 2 points to your data situation sensitivity score.**

We should also note that the three components of data situation sensitivity we have discussed, consent, expectations and data, interrelate. So trust questions (expectation sensitivity) will be more significant where the data are about a vulnerable population (data sensitivity).

Underlying this notion of sensitivity is one of potential harm. The notion of harm is commonly measured in quantitative/economic terms such as financial loss, but it is also recognised that it can be felt in subjective ways such as loss of reputation, embarrassment or loss of dignity. Harm might also occur at the individual, organisational or societal level. The latter two might arise because of knock-on consequences of a reuse of data that violates expectations (whether it is formally a confidentiality breach or not) and leads, for example, to the shutdown of data access and societal benefit not accruing because people become less likely to respond to surveys, provide accurate data etc. You should not underestimate harm at these levels – it means that all organisations who deal with data have a collective interest in everyone getting reuse right.

6.1.4 Desensitising factors

Now as discussed in component 5 you are not passive here – there are things you can do to reduce your data situation sensitivity. These revolve around engagement and transparency and public benefit. The following questions aim to capture this desensitisation.

1. Will there be some public benefit arising from the downstream use of the data?

   (Yes = -3, No =0)

2. Have you carried out consultations with groups of stakeholders (particularly the general public and/or data subjects)?

   (Yes = -3, No =0)

3. If yes to 2 have you implemented any recommendations arising?

   (Yes = -10, No = +3)

4. Does your communication plan engender trustworthiness through transparency (sufficient to offset adverse responses in the expectation sensitivity section)?

   (Yes = -5, No =0)

Now add up the answers and then add the total to the data sensitivity. That gives you your total score for data situation sensitivity.

Ok, hopefully you understand the rationale of the data situation sensitivity score. Now you will need to convert it into an actionable category. If your score is 4 or less, data situation sensitivity is *Low*. If it is between 5 and 10 it is *Moderate* and above 10 it is *High*. Use this classification in combination with the summary likelihood onto which we will now move.

6.2 Determining the summary risk score

At this stage you will not be carrying out a full risk analysis but using some broad brush tools to determine whether such an analysis is necessary.

We described in component 4 how to undertake a top-level assessment of disclosure risk by identifying those features of your data that can potentially increase or mitigate risk. Let us remind ourselves of those features.

Feature	**Effect on risk**
(Poor) data quality	May offer some data protection
Age of data	Older data are less risky
Hierarchical data	Increases risk
Longitudinal data	Increases risk
Population data	Increases risk. Conversely, sample data offers some protection
Sensitive data	Potentially increases the risk and impact of a disclosure
Key variables	The core of the re-identification problem
Microdata	Re-identification disclosure is a particular problem
Aggregate data	Attribution disclosure and disclosure by differencing are particular problems

Table 5: Effects of data features

Thinking about the focal data environment, now consider the following questions:

1. Are the data of high quality?
   a. Yes, the data are clean, and contain no or minimal errors and no or minimal missing data (2 points)
   b. The data contained errors but have been cleaned (1 point)
   c. The data contain some errors and/or missing data (1 point)

   d. The data are dirty, and contain many errors with missing data issues (0 points)
2. How old are the data?
   a. Less than 1 year (5 points)
   b. 1-5 years (4 points)
   c. 5-10 years (3 points)
   d. 10-20 years (2 points)
   e. More than 20 years old (0 points)
3. Do the data constitute a whole population or a sample?
   a. Population (5 points)
   b. Sub-population (4 points)
   c. Sample (0 points)
4. How many variables are there that fall within the standard key variable sets?
   a. 0 (0 points)
   b. 1-4 (1 point)
   c. 5-9 (4 points)
   d. 10+ (5 points)
5. Which of the following best describes the data?
   a. A single aggregate output (0 points)
   b. A set of aggregate outputs which do not overlap (1 point)
   c. A set of aggregate outputs which do overlap (4 points)
   d. Flat microdata (4 points)
   e. Hierarchical but not longitudinal microdata (7 points)
   f. Longitudinal but not hierarchical microdata (7 points)
   g. Hierarchical and longitudinal microdata (10 points)
6. Do the data include any data types that present particular re-identifiability challenges (e.g. genomics data, photographs, significant text narratives, timestamped location data or other timestamped sequences)?
   a. No (0 points)
   b. Yes (+10 points)
7. Which of the following best describes the focal environment?
   a. It is a secure facility with on-site access with limited personnel being able to access the data, which is housed within the data controller's infrastructure. (-25 points)
   b. It is a remote analysis server where users may submit code for analysis but are not able to access the data directly. Code and output are checked before the outputs are released to the user. (-20 points)
   c. It is a remote access server with controls on the who and how of access. Users will be able to interact with the data but do not have a copy themselves (so are prevented from linking to other datasets). How users access and work with the data is pre-specified. (-20 points)
   d. It is a secure facility owned by the user. (-15 points)
   e. It is a point-to-point data share based on a bespoke data sharing agreement(s) with purpose limitations, data minimisation, and specific named users. Some auditing for compliance is in place. (-5 points)

   f. It is a licensing environment. Users sign a licence agreement to access the data and then are able to download them. Restrictions and policing of secondary use are limited. (-2 points).
   g. The environment is open or quasi-open (with minimal sign up conditions). (0 points)
8. Are there data in the focal environment which could be used to re-identify any data subjects in the data?
   a. No (0 points)
   b. Yes (+10 points)

To arrive at your summary risk score add the scores from questions 1 to 8 and then add one for each yes you gave to the five <u>data sensitivity</u> questions in section 6.1.3 (data sensitivity also affects risk directly). Note that the resulting number (which is between -25 and +50) is not a direct measurement of risk. It is an indicator of risk magnitude which will be used to make a decision about whether (and how much) formal risk assessment is needed.

If your score is 0 or less then the summary risk is *Negligible*; between 1 and 5 it is *Moderate*; and over 6 it is *High*.

6.3 Overall assessment

The overall assessment comprises two elements: *the summary risk* (response to PQ2 at the beginning of component 6: "is there within your remit of responsibility a non-negligible disclosure risk that needs to be addressed?"); and *data situation sensitivity* (response to PQ3: "how sensitive is your data situation?"). To carry out your overall assessment you need to consider the relationship between these two elements and from that to classify the anonymisation task as:

1. **Unnecessary:** The summary risk is negligible and the data situation sensitivity is low and therefore you can close this off now.
2. **Essential:** The summary risk and/or the data situation sensitivity are elevated and therefore risk assessment and control processes are going to be needed.
3. **Borderline:** It is unclear at this stage whether controls will be necessary, so a more detailed assessment is required.

		Data situation sensitivity		
		Low	Medium	High
Summary risk	High	Essential	Essential	Essential
	Medium	Borderline	Essential	Essential
	Negligible	Unnecessary	Borderline	Borderline

Table 6: Overall assessment of the data situation

Note that you will do that for each element of the data situation that you are responsible for (as determined by your answer to PQ1). To arrive at this classification you can use the grid shown in Table 6 as a guide.

6.4 Closing remarks

So now you have evaluated your data situation and come to a decision about whether further risk assessment and control is necessary. If not then you can skip component 7. If your data situation sensitivity is medium or high you may want to strengthen some of your desensitising measures. Even if your data situation is in the negligible/low box you should still review components 8-10 as data situations can and do change.

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.

Disclosure risk assessment and control consists of ADF component 7:

7. *Select and implement the processes you will use to assess and control disclosure risk*

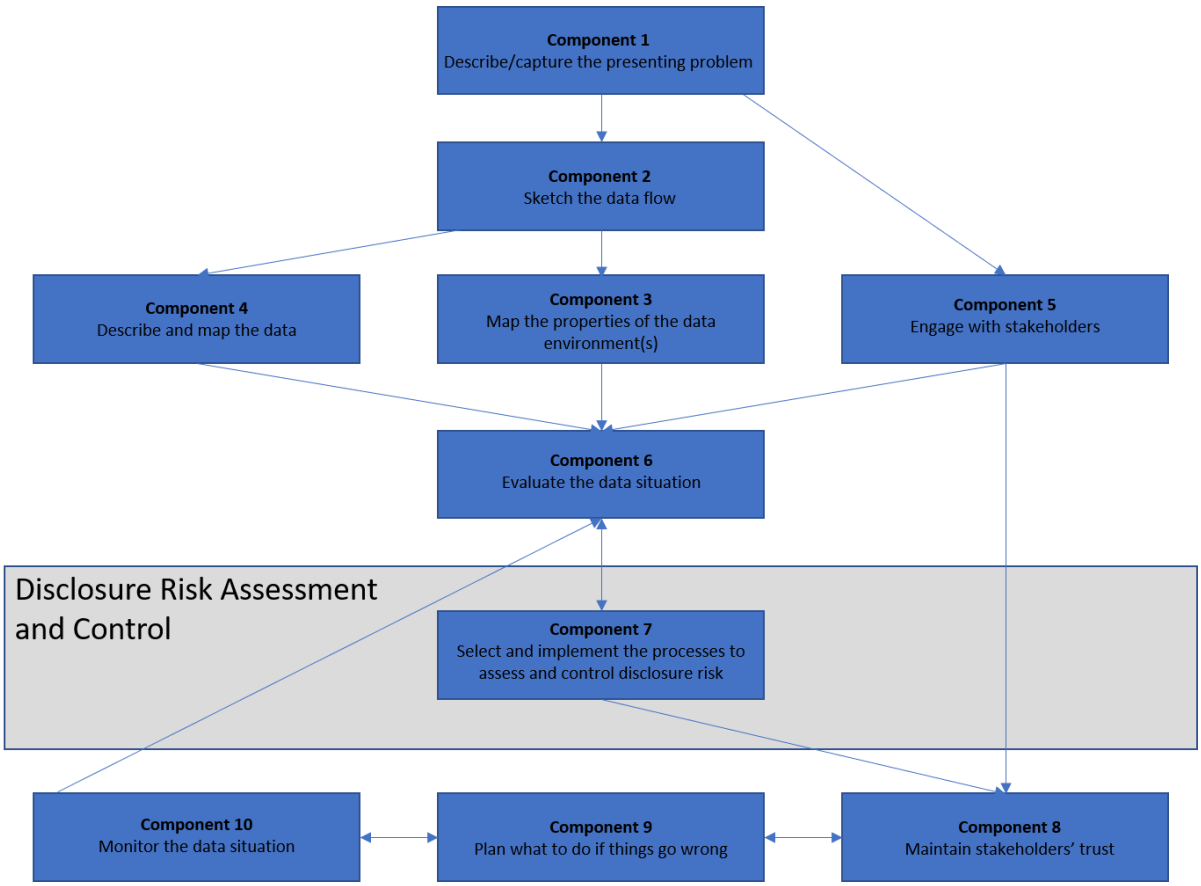


Figure 9: Disclosure risk assessment and control

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

Overview: *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.

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."[39]

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.

[39] https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf.

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.

* * *

Risk assessment is a crucial step in the process of producing safe useful data, helping you to determine:

- whether your data should be shared or released at all;
- whether and to what extent data controls should be applied; and
- the optimum means for sharing or releasing your data, i.e. into what environment and with what controls.

In practice, this can be very complex and risk assessment is probably the most difficult stage of the anonymisation process, requiring judgement and expertise on the part of the data practitioner. The complexity is partly because it is not evident what additional relevant information might be taken into account and how different factors might affect risk. As such factors include the motivation of the intruders, the efforts to which they might go, and the techniques they might use, it is clear that such factors can never be definitively specified. It is also unknowable what information might become publicly or privately available in the future, from sources other than yourself, which might be used to link with the data you wish to release to reveal identity. Notwithstanding these inherent limitations, there are steps you can take to assess the disclosure risk associated with your data and the share/release environment.

We introduce a four-part process for assessing disclosure risk. The first two procedures are always necessary, while the third and fourth may or may not be required depending on the conclusions drawn after conducting the first two.

1. **Incorporation of your data situation audit to produce an initial data specification (section 7.1).** You used this analysis in component 6 to decide whether further action was necessary but here we will employ it to make top level changes to the specification.
2. **An analysis to establish relevant plausible scenarios for your data situation (section 7.2).** When you undertake a scenario analysis, you are essentially considering the how, who and why of a potential confidentiality breach.
3. **Data analytical approaches (section 7.3).** You will use data analytical methods to estimate risk given the scenarios that you have developed under procedure 2.

4. **Penetration testing (section 7.4).** This involves validating assumptions made in 2 by simulating attacks using 'friendly' intruders. The ICO recommends carrying out a motivated intruder test[40] as part of a practical assessment of a dataset's risk. This can be both informative and good practice but takes skill and expertise as well as time and resources.

7.1 Incorporating your top level assessment

The top-level analysis that you carried out in component 6 also enables you to identify where you need to focus your attention in the technical analysis that follows. At this stage, you can also simplify the dataset. Anything that you can do to reduce the complexity of the data will in turn reduce the complexity of the technical analysis that you have to conduct at the next stage.

It might be now that you identify a sensitive variable that is not needed for the use case – if so, take it out. Your default assumption should be that if it is not needed then it should be deleted. If the data are hierarchical, is the preservation of that property required for the use case? Being hierarchical will often magnify the risk markedly, and you may have to compensate with some stringent controls elsewhere in the data. Could the data be simplified to a non-hierarchical structure?

Are there any variables with a lot of detail? If so, is that much detail really necessary for the use case? Frequency tables and descriptive statistics should also be examined. Are there variables whose distribution is highly skewed – say with one category which contains most of the cases and a dozen small categories? Can the small categories be merged? Are there any continuous variables on the dataset that might be rounded or banded?

Such brutality to the data may seem blunt or even draconian but remember that whatever you do you will not eliminate the risk entirely. The more data you release, the riskier it will be, so if a risk is unnecessary for the use case, do not take it and place your pseudonymised data (without the key of course!) in a controlled access environment instead.[41] Initially, removing low-utility/high-risk features will not affect the overall utility. Eventually though, you will hit a point of diminishing returns, where utility reduction will start to become evident and then it is necessary to move on to the second procedure.

7.2 Scenario analysis

The purpose of scenario analysis is to ground your assessment of risk in a framework of plausible events. If you use the Elliot and Dale framework outlined in the ADF companion document *Building Disclosure Scenarios*, then you will run

[40] ICO Anonymisation Code of Practice: 19, https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf.

[41] One factor to bear in mind here is the nature of sharing/dissemination you are working towards. If you are in a data situation where you will be dealing with a series of multiple bespoke data requests, then performing a risk analysis and editing process with each individual request could be quite onerous. It may be simpler to produce a single dataset that meets the negligible risk criterion in any conceivable use case. However, the disadvantage of this approach is that it will inevitably be a lowest common denominator dataset and some users may not be able to access the data they want. As ever there is a balance to be struck here.

through a series of considerations using simple logic to arrive at a set of key variables. In constructing these, you need to consider all the sources of the data to which the would-be intruder might have access.[42] Below are examples of other sources of data that may be relevant when developing your scenarios of disclosure.

- **Public sources of data:** including public registers, professional registers, electoral registers, land registry, estate agents' lists, newspaper reports, archived reports and announcements, parish records and vital statistics such as birth, death and marriage records, or fundamental commitments such as that to 'open justice' in courts.
- **Social media and other forms of found data:** including data generated by the data subjects themselves in online interaction and transaction. This runs from deliberate self-publication (CVs, personal websites), to material where the goal is primarily interactive (social networking sites), to expectations placed on particular targets (such as celebrities publicly thanking the National Health Service for its care following a health issue). Needless to say, this is a growing source of publicly available information and it has been demonstrated that it is plausible to attack an open dataset using a combination of social media data and other publicly available sources (Elliot et al 2016b).
- **Other similar data releases:** including releases from your partner organisations and other organisations in your industry or sector.[43]
- **Official data releases:** including data releases from the Office for National Statistics (in the UK), national government departments and agencies, and local authorities.
- **Restricted access data sources:** including the resources of any organisation collecting data. At first, it may seem difficult to imagine how you would know what is in such data sources but they can often be the easiest to find out about. Why? Because although the data are hidden, the data collection instruments are often public. They include the forms that people have to complete in order to access a service, join an organisation or buy a product. If you can access the forms used to gather data, then you can make a good guess about what data are sitting on the database that is

[42] In considering your wider data protection obligations you should also think about the possibility of a user interacting with the data in a way not (reasonably) foreseen with harmful effects. In respect of re-identification issues this really only concerns the spontaneous recognition of data units. In reality, instances of this very rare and should almost always be dealt via governance measures (e.g. making clear to a user what they should do in such an event rather than by restricting the data).

[43] Ideally, if multiple organisations were releasing open data on the same population then they would co-ordinate their anonymisation processes. However, in most cases in practice, such an undertaking would be very difficult. The importance of other releases will be greater if the data generation processes are similar, the time of collection is similar and if there is partial overlap of variables. This set of circumstances will usually only arise when the organisations are closely related. This, in principle, allows key extension; as both datasets are anonymised we are not here talking about direct re-identification, but the fusion of two anonymised datasets could make both more vulnerable.

fed by the form. For the task of generating key variables that should be sufficient.

It is easy to become overwhelmed by the feeling that there is too much data out there – where do I even start? Certainly doing a full scenario analysis is very time-consuming and beyond the resources that many organisations are likely to have available. Fortunately, for many data situations a full analysis will be disproportionate – and you only have to do so much before your threshold of risk is exceeded and an alternative method of sharing your data, which keeps these risks at bay, should be found. This will be particularly true where you are working in a tried and tested area. If other organisations have been releasing similar data for a while without any apparent problems then your resources that you need to devote to this element can be more modest.

One tool that can cut down the amount of time required at this stage is a standard key set. Standard keys are generated by organisations carrying out ongoing data environment analysis (scanning the data environment for new data sources). You should be aware that standard keys are generic and are set up primarily for use with licence-based dissemination of official statistics and will not be relevant to every data situation. However, standard keys can be useful because, if your data are not safe relative to these standards, that in itself indicates that you may have a problem, even before you consider non-standard keys. A set of standard keys can be found in the ADF companion document *Standard Key Variables*.

So what in practice can you do if you are not carrying out full scale data environment and scenario analyses? The simplest approach is to carry out thought experiments that put flesh on the bones of the imagined adversary.

For example, imagine that you work for a local authority wanting to release a dataset of social care service users as open data. Suppose the dataset contains seven variables: age (banded), sex, ethnic group, ward (Local Authority subdivision), service accessed, the year that service was first received, and type of housing.

Now imagine a data intruder who draws on publicly available information to attack a dataset that you have released as open data. Run this scenario through the Elliot and Dale framework. In particular, think of a plausible motivation and check that this passes the 'goal not achievable more easily by other means' test. In this example, you might end up with inputs that look something like this:

- **Motivation:** What is the intruder trying to achieve? The intruder is a disgruntled former employee who aims to discredit us and, in particular, our attempts to release open data.
- **Means:** What resources (including other data) and skills do they have? They have publicly available data. Imagine that they are unemployed, and have unlimited time, but do not have access to sophisticated software or expertise for matching.
- **Opportunity:** How do they access the data? It is open data so there is no obstacle.
- **Target variables:** The service(s) individuals are using.

- **Goals achievable by other means?** Is there a better way for the intruders to get what they want than attacking your dataset? Possibly, but given their motivation they may also want to discredit our open data policy, in which case the answer is 'no'.
- **Effect of data divergence:** We believe our dataset to be reasonably accurate. However, we are only publishing data that is at least one-year-old; the intruder's data will be less reliable. This will create uncertainty for the intruder but not enough to rely on.

Once you have a plausible scenario, then look through the standard key set to see if any of those correspond meaningfully to the total information set that the intruder might have. In this case, the standard key B4.2 from the ADF companion document *Standard Key Variables* looks relevant. If we cross-reference the list of variables under that key with the list that we are considering releasing, that gives us the following intermediate outputs.

- **Attack type: what is the technical aspect of statistical or computational method used to attack the data?** Linkage of data about individuals living within our local authority derived from publicly available information, to records in the open data set.
- **Key variables:**
    - Ward
    - Ethnic group
    - Age (banded)
    - Sex

These key variables can then be used as a starting point for the technical disclosure risk assessment. If you are taking this approach then it is wise to construct more than one scenario. The number that you will need will depend on the totality of the data situation, and specifically who will have access to the data, and the complexity of the data in question. With this situation, we are talking about open access but relatively simple data. With open data, we will often want to also assess the nosy neighbour scenario (Elliot et al 2016b for a rationale), which would suggest adding *type of housing* to the list of keys, but would also mean that we were simulating an attack by an unsophisticated intruder who was just trying to find a single specific individual (rather than any high-certainty match from a host of possibilities).

Of course, if your data does not fit nicely into the format of the standard keys then you are going to have to do some work to populate this framework yourself. You should avoid focusing too closely on apparent vulnerabilities in the data.[44] Uniqueness – and particularly data uniqueness – does not in itself re-identify anybody. It does indicate vulnerability, but if there is no well-formed scenario through which uniqueness can be exploited, then re-identification is very unlikely to happen in practice. On the other hand, a sophisticated intruder might focus on

[44] For a good analysis of the pitfalls of doing this, see, for example, Sánchez et al's (2016) critique of de Montjoye et al's (2015) account of the uniqueness (called 'unicity' by de Montjoye et al) of small strings of credit card purchases.

those vulnerabilities to carry out a fishing attack.[45] It comes down to whether there is a well-formed and feasible scenario where they would be motivated to do that.

So create your scenarios, generate your key variables and then carry them through to your risk assessment.

7.3 Data analytical risk assessment (DARA)

Having gathered the low hanging fruit of reducing the detail in the data and generated your sets of keys now you are ready to move on to carry out a *data analytical risk assessment (*DARA). We would always recommend that you get expert advice at this stage even if only to ratify what you have done. However, much can be done without external help, and the more that is done in-house, the richer the conversation that you can have with independent experts, including communicating your specification of the problem to them, and interpreting their findings and recommendations. In this section, we will set out a process that could be performed in-house, without (i.e. before) consulting anonymisation experts.

7.3.1 File-level risk metrics

The first step in the DARA is to obtain a file-level measure of the risk. There are quite a few of these and selecting the right one can be a bit of a Chinese puzzle in itself. There are three key questions whose answers will guide you.

1. Is your data a sample or a population?
2. Does your scenario assume response knowledge and if so at what level? Response knowledge is the very useful knowledge (to an intruder) that a record corresponding to a particular known population unit is present in the microdata.
3. If it is a sample then is it (approximately) a random sample?

By 'population' here we do not just mean the UK population (although that would be one example). For the purposes of statistics, a population is a complete set of objects or elements that share a particular characteristic of interest. For example, if your data are about all members of Anytown Cycle Club or all claimants of a certain benefit then these are a population (the word 'all' is the indicator here).

Response knowledge is a simple idea but it can be complex to apply. In some scenarios, you may want to assume an intruder with *ad hoc* but full knowledge about a particular individual. For others you may want to consider a situation where the super-population (i.e. a larger set from which the population is drawn) is constrained. Perhaps I know that Anytown Cycle Club members all have to live in Anytown and own a bicycle; if I also know that you have those characteristics then I know that the probability of you being in the sample is considerably higher than I would estimate if I did not have that knowledge.

[45] I.e. finding unusual looking records in the dataset and attempting to find the corresponding individuals. Note that this is distinct from a *Phishing attack* – a cyber-attack which uses a communication (usually email) to induce the recipient to click on a link or open an attachment with adverse consequences.

7.3.2 Response knowledge scenarios with microdata

First, consider the situation where your intruder has full response knowledge. This is the simplest case to assess. You need to identify how many unique combinations of the key variables you have in your dataset. This can be done simply using a spreadsheet or statistical package.

With the companion documents (on the UKAN website), you will find a two CSV files that accompany this book, which can be opened in a spreadsheet system or statistical software, *Bassteon.csv* and *Basseton Sample.csv*. These files contain example synthetic data that have been generated using some simple models, but which look like census data to give you something to practice on. The ADF companion document *Instructions for Calculating the Number of Uniques in a File* gives instructions for calculating the number of uniques. You can adapt either of these to your own data. In the example in the companion document, the key variables that we have used are age, sex, marital status, ethnic group, type of housing, tenure, number of cars, and whether the house has central heating. This corresponds to the types of things that neighbours might routinely know about each other. You might want to play around with other combinations of variables.

In the example data, using the 'nosy neighbour' key we find that over 17% of the records are unique. What would such a result imply? Simply put, if our intruder has response knowledge for any of the individuals whose records are unique on those characteristics then the latter are at high risk of being re-identified – even decades later, everyone remembers the characteristics of the kids they went to school with. The only protection that these records have is the unreliable possibility of data divergence. Given that nearly 1 in 5 of the records in our dataset have this status we might decide that that is too high.

Faced with unique patterns in your population dataset what are your options? Essentially, you have three choices: (i) give up now, do not release the data and set up a more controlled share instead; (ii) proceed to apply disclosure control; or (iii) if you still want to persist with your proposed release/share then carry out penetration (or pen) testing (which we discuss in section 7.4). If you go for option (ii) and decide to apply data-focused, rather than environment-focused, disclosure control, then you will need to revisit this step once your data-focused control mechanisms have been applied, in order to reassess the risk. You may also need to iterate between options (ii) and (iii) depending on the outcome of the pen test.

7.3.3 Scenarios involving microdata without response knowledge

What if my file is not a population and my scenario analysis does not suggest response knowledge? Here you have a sample. There is a simple method known as *data intrusion simulation* (DIS)[46] which can help here. DIS provides a statistic that is straightforward to understand: *the probability of a correct match given a unique match*. In other words, it tells you how likely it is that a match of auxiliary

[46] For a full technical description of the DIS method, (Skinner & Elliot 2002). A brief explanation and some examples showing how it works are shown in the ADF Companion Document.

information against a record that is unique within the sample dataset is correct.[47] However, if you are looking at a strongly non-random sample then this step is a little trickier and you should consult an expert. Using the same data as the uniques test above, the ADF Companion Document *Instructions for Calculating the DIS Score* contains the instructions you need for implementing DIS in Excel.

The output of the DIS process is a measure of risk taken as the probability that a match against a unique in your dataset (on your selected key variable set) is correct. Essentially this takes account of the possibility that a unique record in a sample dataset may have a statistical twin in the population that is not represented in the sample.

For scenario keys of any complexity, the outcome will not be zero risk. You knew this already of course, but quantifying risk immediately raises the question about how small a probability you should be aiming for. We cannot give you a single threshold because unfortunately there is no straightforward answer. Here instead in Table 7 is a set of rough guidelines for helping you think about your output. We have mapped particular quantitative outputs of the DIS process onto qualitative categories on the assumption that the data are non-sensitive, and indicated the type of environmental solution that might be suitable. If the data are sensitive, then we need to shift down the table by one or even two categories, so that a DIS output of 0.03 should be treated as signalling a moderate risk (or even a high risk if the data are very sensitive), instead of the low risk it would signal on non-sensitive data. As ever in this field, context is all.

DIS output	Risk signalled	Indicated sharing decisions
<0.001	Very low	
0.001-0.005	Low	Open data maximum
0.005-0.05	Low	End user licensed data maximum
0.05-0.1	Moderate	Restricted user licensed data maximum
0.1-0.2	High	Online remote access solutions maximum
>0.2	Very high	Highly controlled data centre solutions only

Table 7: Classification of output from the DIS algorithm

To stress, these are only for ballpark guidance but (assuming that your scenario analysis has been thorough) they should serve to indicate whether your overall level of risk is proportionate to your proposed solution. A final point here is that we have not used our preferred threshold term 'negligible' in this table. The reason for this is that the term will only be applied to your assessment across the entire data situation, not to individual analyses.

If you have an unfavourable result at this stage, and you are out of your risk comfort zone (given the receiving data environment), what do you now do? The simplest solution at this stage is to apply aggregations to some of your key variables and/or sub-sample your data. Both of these will reduce the probabilities you have generated. Another alternative is to change the environment to one in a

[47] A technical point: this method does assume that your sample is random, although in fact it is robust with respect of some degree of variance from random.

lower risk category. You may also revisit your use case. What constraints on data and environment are consistent with the needs of the use case?

One question that might arise is whether the above categorisations of DIS are a little on the conservative side. If the intruder only has a 1 in 5 chance of being correct, does that represent sufficient uncertainty regardless of the environment? If the risk was spread evenly across the file or if it were impossible for the intruder to pick out unusual records then that might be true but unfortunately, as noted earlier, some records are visibly more risky than others, vulnerable to fishing attacks or spontaneous recognition. This will be true even if you are broadly in your comfort zone.

7.3.4 Record-level risk metrics

Conceptually, understanding disclosure risk at the record level is very simple: unusual combinations of values are high risk. Unfortunately, identifying all the risky combinations in a dataset is not straightforward and deciding what to do about them perhaps even less so. It might be at this point in the proceedings that you decide to call in the expert, but if you carry on there are some things that you can do that will at least have the happy side-effect of familiarising you with the data and their properties.

How does one define a risky record? There are many answers to this question and it is still an active research area. Yet focusing on the concept of uniqueness reveals two simple pragmatic principles:[48]

1. The more information you need to make a record unique (or to 'single it out'), the less unusual it is.
2. The more information you need to make a record unique (or to 'single it out'), the more likely that any match against it is prone to data divergence, increasing the likelihood of both false positive and false negative matches.

The first principle is primarily relevant to scenarios where there is sample data and no response knowledge. The second is relevant when either your data are population data or your scenario assumes response knowledge.

Start with principle 1. The basic idea is as follows. You have performed scenario analysis and generated a set of key variables. Say, for argument's sake, that you have eight of them.[49] Principle 1 says that if a record is unique in your data on, say, three of those variables it is more unusual than if you need values for all eight variables to make it unique. One way to think of this is that each time you add another variable to a key you divide the population or sample into smaller groups.

[48] For those who wish to dig a little deeper, there is some science underpinning this approach (Elliot et al 2002, Haglin et al 2009). A related approach is (Fienberg & Makov 1998).

[49] The astute reader may have noted that it is not just the number of variables that matters, but also the properties of those variables, most notably the number of categories, the skewness of the variable and correlations with other variables. However, these basic principles are sound and even this simplification will improve decision-making. Recall that at this stage we are considering steps that can be taken in-house, prior to calling in an expert consultant, and at some point it will be sensible to leave the complexities to them.

Eventually everyone will be unique, so uniqueness itself is not such a big deal; the unusual people are those who are unique on a small number of categories.

The second proposition is in some ways simpler. Essentially each piece of information you have in your set of key variables carries with it the possibility of divergence. So each variable that you add to a key increases the probability of divergence for any match against that key. For sample data in scenarios without response knowledge, that has to be weighed against the informational gain from the additional variable. For population data or response knowledge scenarios the informational gain is irrelevant – a unique is a unique – but the impact of divergence is important and a sophisticated intruder will focus on individuals who are unique on a small number of variables.

Given these two principles, we can set out a rough and ready way of picking out unusual records. There is software available called SUDA (Elliot & Manning 2003) that can produce a more sophisticated version of this but understanding the principles first is still useful.

On the assumption of an eight variable key, you need to search for uniques on small subsets of those eight variables. We assume also that you have access to a statistics package.

1. First, run all of the two-variable cross tabulations. Do you have any uniques? If you do then identify the records that they belong to (filtering will do the trick here). These records are unusual enough to be noteworthy.
2. Now find the smallest non-unique cell in all of the two-way cross tabulations that you have run. Filter your datasets on that combination of values. Then run frequency tables on the remaining six variables. Are there any uniques in those frequency tables? If there are, then they will also be candidates for being interestingly unusual records (although they will probably not be as unusual as the ones you identify in stage 1).
3. Repeat step 2 with next biggest non-unique cell in the two-way cross tables and continue repeating until you have reached a threshold (a cell size of 10 is a good rule of thumb).
4. Repeat steps 1-3 for each key variable set that you have.

This takes you as far as covering the 3-way interactions. In principle, you can repeat the exercise with the 4-way interactions but that can involve a lot of output. Nevertheless, if your sample size in a response knowledge scenario is reasonably large then it might be important to do this.

Now you have a list of unusual records. What can you do with that? Well firstly, you can do a subjective assessment of the combination of values – do any of the combinations look unusual? Knowledge of the general structure of the population, which you are likely to have as you have a professional interest in the data, will undoubtedly help here. Perhaps present them to colleagues to get a sanity check. Such subjective analysis is obviously not perfect and subject to all sorts of biases but it can be informative (everyone would tend to agree sixteen year old widowers are rare, for example). If you have records that definitely appear unusual then you almost certainly need to take further action.

It is also important to consider how many records you have marked out as unusual. Is it a large portion of the size of your data file? If you have a relatively small number of records (relative to the file size), say less than 1%, then it might be possible to deal with them by techniques that involve distorting the data, which we consider below. If the proportion is larger than that, then a more sensible approach is to carry out further aggregation and rerun the above analysis.

However, a cautionary tale will explain why it is also important to avoid knee-jerk reactions. A few years ago one of the authors was carrying out some work on behalf of a statistical agency identifying risky records within a longitudinal dataset, using the risk assessment software SUDA. The analysis threw up some odd patterns with some apparently very high risk records. A bit of exploratory analysis revealed that these were records where the individuals had changed sexes several times in the space of a year or had reversed the ageing process! In other words they were the result of errors in the data. Arbitrary data errors will often lead to unusual looking records so not all unusual records are actually a risk. More importantly, this sort of noise in the data generation processes does itself provide a handy side benefit of 'natural protection' against intruders using fishing attacks.

7.4 Penetration tests (pen tests)

You might conduct a pen test[50] at any point in your disclosure control process. You might perform this after you have applied the disclosure control as a test as to whether your controls are sufficient. Or you might do it early in the process to guide your decisions on how to control risk.

There are essentially four stages to a pen test: (i) data gathering; (ii) data preparation and harmonisation; (iii) the attack itself; and (iv) verification. The first stage tends to be the most resource intensive and the second and third require the most expertise. In general, external expert involvement will be helpful, even if you have the expertise yourself, to bring the perspective of an independent attacker.

7.4.1 Data gathering

Data gathering involves going out into the world and gathering information on particular individuals. This phase can be resource-intensive. Exactly what it will look like will depend on the nature of the scenario that you are testing, but would typically involve at least some searching of the Internet.[51]

A key point in this process is to decide whether one is assuming that the intruder has response knowledge or not, which will have been indicated by the scenario analysis. If so, then the data holder will provide the matcher with a small sample

[50] The term 'penetration' or 'pen' test is used in cybersecurity contexts to describe a simulated attack to break through a security system. Here we are using it to simulate an intruder who is attempting to attack anonymised data – we will use the term 'pen test' throughout to mean this. Note that in other places the term 'intruder testing' is also used.

[51] The intruder test reported by Elliot et al (2016b) gathered information on 100 individuals, taking about three person-months of effort. That test also included a second augmented attack using data purchased from the commercial data broker CACI.

of random formal identifiers (usually name and residential address), drawn from the dataset. If not then the simulated matcher will usually adopt the stance of finding unusual looking records in the dataset and attempting to match to corresponding individuals in other sources (a fishing attack).

7.4.2 Data preparation and harmonisation

Once the data gathering phase is complete then the data have to be harmonised with the target dataset. This will require work both across the data, and at the level of individual records, as in all likelihood there will be several issues to address to achieve this. Gathered data will often be coded differently to the target data; for example you might have gathered information about somebody's job from social media, but how exactly would that be coded on the target dataset? There will be *data divergence* with the gathered information; for example the gathered and target data are unlikely to refer to the same set of time points so how likely is it that a given characteristic will have changed in the time differences and if so is that an important consideration? How confident are you in a piece of gathered information? For example, Google Street View may show a motorcycle parked in the driveway of a target address. If you have a variable in your dataset indicating motorcycle ownership, this is very tempting to adopt as a key piece of information, as it will be a highly skewed variable (most people do not own a motorcycle). But it may have belonged to a visitor, or the house might have changed ownership between the time of the Google visit and when the target dataset was created, or the bike might have been bought or sold in the interim. So, when constructing your keys on a record-by-record basis, though you need to take into account all the information that you have gathered about a particular identity, some of it should be flagged as less reliable at this preparatory stage so that it can be treated more cautiously at the attack stage.

Some scenarios simulate linkage between an identification dataset and a target dataset, rather than between gathered data and a target dataset. Here no data gathering is necessary but data harmonisation will still usually be necessary and issues of data divergence still critical, although the focus here will tend to be on the dataset as a whole rather than upon individual records.

7.4.3 The attack

The details of the attack stage will also depend on the nature of the data and the scenario. But typically it will involve attempting to link the information that you have gathered at stage 1 to your dataset. Usually this will involve a mixture of automated and manual processes. In essence you try to establish negative and positive evidence for links between your attack information and records in the dataset.

When you carry out the linkage you will quickly become aware that this is an inexact science and the task is rarely as simple as dividing the potential matches into two piles. There is the matter of your confidence in the matches. This could simply be a subjective estimate of how likely you think it is that a match is a true match or it could involve a more quantitative approach. This will partly depend on what type of data intruder you are simulating. Is this an expert carrying out a demonstrative attack or simply the next door neighbour being nosy? Table 8 shows what an output from this process might look like.

Name	Address	Record number	Confidence	Effective confidence
Johnny Blue	10, Canterbury Gardens	10985	95%	95%
Jamie Green	68, York Walk	45678	95%	95%
William Pink	53, Winchester Lane	42356	90%	60%
Fred Purple	39, Southwell Drive			30%
Archibald Black	68, Christ Church Avenue	671	85%	85%
Jane Indigo	23, Westminster Close	37	80%	40%
		9985		40%
Patricia Vermilion	20, Bath & Wells Street	70637	60%	60%
Wilma White	53, Exeter Road	68920	50%	50%
Gertrude Gold	57, Chichester Broadway	35549	40%	40%
Brittany Magnolia	12, Lincoln Row	22008	30%	30%
Petra Puce	75, Norwich Terrace	68680	30%	30%
Stephanie Red	11, Ely Place	81994	30%	30%
Simon Violet	136, Peterborough Way	91293	20%	20%
Estimated number of correct matches				7.05

Table 8: An example of output from a penetration test

We see from Table 8 that there are two individuals matched against record 42356 and that the individual 'Jane Indigo' is matched against two records. Here the matcher has been unable to distinguish cleanly between two possible matches against a record but is fairly confident that one of them is correct (this is captured in the effective confidence column). It may be important to record these, because a real intruder may (again depending on the nature of the scenario) have options for secondary differentiation which are not available in the simulation. In other words, he or she may take close matches and engage using a different approach from the original data collection activity (for example, actually visiting a matched address and capturing further data by direct observation).A second point to note is that no match has 100% confidence associated with it. This reflects the reality that we can never be completely certain that we are correct. There is always a possibility that (i) the dataset contains data for a person who is highly similar to our target – their statistical twin – or that (ii) the assumption that our target is in the data is incorrect.

7.4.4 Verification

Finally, once you have selected the matches, they need to be verified. Ideally, this would be carried out by a different person or organisation than the person doing

the matching. If the matcher is carrying it out – at the risk of stating the obvious – they should only do it once they have decided upon their final list of matches.

In interpreting the results of a penetration test one needs to exercise some caution. Although the simulation will be a more direct analogue of what an actual intruder might do than with data analytical approaches, there are still differences which will impact on the results. Elliot et al (2016b) list the following:

1. ***Ethical and legal constraints.*** Penetration tests are constrained ethically and legally; a real attack may not be.
2. ***Expertise variance.*** Typically, the matcher will be an expert, or at least skilled and knowledgeable about data. Even if they 'dumb down' their matching process in an effort to simulate a 'naive' intruder they will not be able to switch off their knowledge. This will particularly affect the estimation of match confidences.
3. ***Time available for data gathering.*** In order to get a picture of the risk across the whole dataset, pen tests usually consider multiple individuals. Resource constraints mean that the amount of time spent gathering information on each of those individuals will be limited. A real data intruder may be able to achieve their goal with just a single correct match and therefore may be able to focus attention on a specific individual.
4. ***Dataset specific results.*** Be careful about generalising any results to your data products and data situations in general.
5. ***Difficulties in simulating real response knowledge.*** A real data intruder with response knowledge might have *ad hoc* knowledge with respect to their target that is hard to simulate through gathered data. If one wants to simulate such an attack, one would need to co-opt data subjects and members of their social network into the study. This is an interesting possibility, but to our knowledge no such study has ever been carried out and realistically would be too resource intensive for practical risk assessment.
6. ***Pen tests only give snapshots.*** The data environment is constantly changing and more specifically the availability of data that could be used to re-identify individuals is increasing. A pen test if done well may tell you a great deal about your risk *now* but that risk can and indeed will change.
7. ***Arbitrary variation of data divergence.*** Typically in these exercises one is gathering current data to carry out the simulated attack whereas the target data are past data. Temporal data divergence can markedly reduce the accuracy of matches so the degree of divergence between the data collection for the target dataset and the data gathering for the simulated attack will impact on the results.

Taking these considerations into account, what sort of level of successful matching would one consider problematic? It is difficult to generalise this. If you have produced a table like Table 8 and you see most of the high confidence matches are true matches then you have a problem and you need to rethink your data situation. But what if you have, say, a single correct match? The false positives are important here – are some of these high confidence matches? If so then the single correct match is swamped by false positives, in which case how could an

intruder determine that that match was a correct one? Remember they will not have the advantage of being able to verify!

This reminds us to think about risk from the intruder's perspective – could claiming a match that turns out to be incorrect backfire on them? If so then they might well be cautious before making a claim. Another aspect to bring to the table in your thinking at this stage is the sensitivity of the data. If you think the impact of a correct match is high then your tolerance for a single correct match will be lower than if the expected impact is low.

Related to this is the importance of cross-checking the correct match rate achieved against the rate estimated by the matcher. To derive the former, simply sum up the confidences (converted to proportions). So, you can see in Table 8 the expected number of correct matches is 7.05. That is on the basis of the matcher's confidence in their matches, they should expect 7.05 correct matches.

So you will now have two figures: the actual number of correct matches that the matcher has made, and their expected number given the confidence. Both of these figures are relevant. If the matcher has been unsuccessful in matching then any confidence that they have is false – they will be claiming matches which are false positives. On the other hand if they are unconfident about their matches then they will not be reasonably be able to make a claim. The high risk situation arises where the matcher is confident and their confidence is well founded because they have made matches. There are two indications that we are in this situation: (i) the expected number of matches being close to the actual number, and (ii) the high confidence matches being mostly correct (so in the example if three of the top four matches were good, that would be cause for concern).

Of course, a real data intruder might hit on a match that by chance happens to be correct, and they may not care or even know about nuances such as confidence levels. Although you have to think about such eventualities, you cannot build your data sharing practices around them – the correct place to deal with them is in your breaches policy, which we discuss in component 9.

A final question is what we assume the intruder knows about the disclosure control applied to the data. Nothing? The methods employed? The methods plus the parameters used? This will partly depend on the moment in the anonymisation process in which the penetration test is run, and the type of disclosure control that has been applied. If you have merely aggregated and deleted variables then we can assume that the intruder simply observes the effects of the control process. However, if data distortion has been applied then a sophisticated intruder will be able to use knowledge of the details of this if they are published.

7.5 Disclosure control

Disclosure control processes essentially attend to either or both of the two elements of your data situation: the data and their environments. If your risk analysis in component 6 suggests that you need stronger controls then you have two (non-exclusive) choices.

1. Change the data (specification)
2. Reconfigure the data environment

In some cases the environment is a fixed point of reference within the data situation: "we want to release an open version of this dataset" or "we want to share these data with organisation X for purpose Y". In this case your anonymisation solutions will have to be data-focused and the environment will be fixed. We will discuss these cases in sections 7.5.1-7.5.4. In other cases it is possible to achieve anonymisation, at least in part, through reconfiguring the environment, which we will discuss in section 7.5.5.

7.5.1 Changing the data

Usually one starts from a fairly fixed proposal of what the release/share environment will be, defined in components 1 and 4. It may be that this fixed idea has to change but initially one has to work on changing the data. The most common place to start is aggregation.

1. Keeping the use case in mind, can you lose detail on your key variables to reduce the measurable risk?
2. If your data situation is sensitive, can you remove or reduce detail on sensitive variables?

Often the answer is yes. You will lose some utility but not to the extent that the data lose most of their value.

Variables that tend to be a focus here are spatial and temporal ones – typically place of residence and age. The latter is particularly important if the data is about multi-member households. Other variables that can be considered are those with skewed distributions (where minority categories can be merged together). However, any variable that appears in your scenario keys should be considered.

At this point, you should also consider producing a sample rather than releasing all of the data. Any level of sampling will reduce the risk, but sample fractions that one would normally consider range from between 1% and 50%. Most microdata products from censuses and social surveys are released as samples at the bottom end of this range and these are generally regarded as high utility products, so it is worth giving some serious consideration to this possibility for release use cases.

One overarching advantage of metadata controls, such as aggregating scenario keys and sampling, is that you can easily rerun your risk measurements in order to see what impact a particular aggregation has on the overall level of risk. Doing this with data distortion controls is more difficult.

7.5.2 Data distortion

In general, if it is possible to reduce the risk to an appropriate level through aggregation, variable deletion and sampling, then that should be the preferred approach. Applying data distortion controls affects the data utility in an unpredictable and generally non-transparent manner and leaves you with the difficult question about whether or not to release information about the distortion.

However, if you have done all that you think you might be able to do with metadata-level controls and the risk is still too high, then you will have to move on to data distortions or reconfigure the environment. If the latter is not possible

because the use dictates a particular environment, then distortion of the data remains as the only possibility.

Now you have to decide whether the distortion should be random or targeted. Random distortion in fact has relatively low impact on the risk – you will have to do quite a lot of distorting before you get a significant impact. Random distortion works by reducing the baseline confidence in any match. Targeted distortion potentially has a big impact on the disclosure risk. The point of targeting is to focus on the high risk records (those identified by your record-level risk metrics). So if you turn a sixteen year old widower into a sixteen year old single person then you have merged him into the crowd and the risk goes away. However, the big cost is that you alter variability and introduce bias. Therefore, our guidance is to do this only very sparingly.

The second issue is that once you have distorted the data then the standard risk metrics will no longer work. There are techniques for measuring post-distortion risk, but these are experimental and complicated to implement. Therefore, there are two options.

1. Add in distortion to pick up a small amount of residual targeted risk when you are quite close to your acceptable level anyway.
2. Carry out a pen test.

In general, we would not advise using data distortion controls if you can avoid them, and if you do you should consult an expert first.

7.5.3 Using a confidentiality model

There are two types of confidentiality model[52] in common use: *k-anonymity* (and its derivatives) and *differential privacy*.

K-anonymity is based on extrapolation from the ideas of uniqueness and statistical twins discussed above. K represents an integer, and the principle is that you protect individuals within databases by ensuring that for each record in a dataset, there are at least k-1 other individuals who are identical on a given set of key variables, so that the intruder does not know which record belongs to their target individual. K-anonymity is achieved by reduction of detail in the variables in your data either across the whole data set (sometimes called *domain generalisation* or *global recoding*) or within some parts of it (sometimes called *microaggregation* or *local recoding*). K-anonymity is not without flaws and specifically it may still be possible to make inferences about a non-controlled target variable. To use a simple example; if all records in a given set of k records have the same value for the target variable, then an intruder will be able to infer that value for the target even without re-identifying the record. For this reason, various extensions to k-anonymity have been proposed; the two most commonly used ones being *l-diversity* and *t-closeness*.

[52] These are often referred to as 'privacy models'. We do not use that term here as it as a misnomer; the models are very little to do with privacy as such. For a thorough introduction into using these techniques, (Domingo-Ferrer et al 2016). For good summaries of differential privacy, (Desfontaines 2019, Sartor 2019).

Differential privacy is a guarantee, not a risk assessment model. The guarantee is a limit on the amount of information specific to any individual that is revealed by an analysis. Although it is not an algorithm, a set of algorithms for implementing differential privacy is available to support tasks like releasing aggregate statistics, creating synthetic data, and training machine learning models.

These algorithms typically rely on introducing a controlled, typically small, amount of noise into the statistics or models before releasing them. This noise is calibrated to be of the same magnitude as an individual's possible contribution to the release, effectively masking the contribution of any one individual. In (Elliot et al 2016a) we took a critical view of differential privacy but our position now is that it has a role to play in the anonymiser's tool box. A longer description of differential privacy, written by Guy Cohen and Dr Hector Page of Privitar, appears in the ADF companion document *Anonymisation with Differential Privacy*.

7.5.4 Synthetic data

Synthetic data takes the notion of changing the data to its extreme. Rather than manipulate the existing data, it replaces the dataset completely with a synthetic version which has been generated using some model of the original data. Data synthesis is a live research area and, at present, it will not be a viable solution for most data situations. However, it is being seriously looked at as an alternative to disclosure control data for open data releases by statistical agencies.[53]

7.5.5 Reconfiguring the environment

Reconfiguring the environment, instead of changing the data, essentially involves controlling who has access, how they access the data and for what purposes. Options to consider are:

1. Allowing access only within your own secure environment.
2. Specifying the requisite level of security for the data.
3. Specifying that all analytical outputs must be checked and sanctioned by you before they are published.
4. Specifying the people who may access the data (and for what purposes).

Placing or tightening controls on the environment will tend to have quite significant effects on the risk, often ruling out particular forms of attack, for example, and so if the data are sensitive they are certainly worth considering.

One concept worth mentioning here is the *Five Safes* (Ritchie 2017, Arbuckle & Ritchie 2019). This is a framework comprising five risk (or access) 'dimensions' underpinning data sharing decisions: safe project, safe people, safe data, safe setting, and safe output. These dimensions map reasonably well onto our own description of the data environment and although it is not an empirical risk assessment model, it can be a useful adjunct to the ADF if you are intending to control risk primarily by using environmental controls.

[53] For a technical overview of data synthesis for confidentiality protection, (Drechsler 2011). For a recent discussion about how to measure disclosure risk in synthetic data, (Taub et al 2018).

7.5.6 A note about outputs

To conclude this component we would like to say something about outputs. As mentioned in the Introduction, in ADF terms analytical outputs are data and so publication of outputs is the movement of data from one environment to another. We describe this in more detail in data situations 3 and 4 in component 2. However outputs are invariably a considerably more restricted form of data than inputs. Nevertheless you are publishing these – possibly into the open environment and therefore the environment is substantially more risky and there is evidence that simple tables of counts, summary statistics and models can be disclosive. Because of this your controls will be data focused using further restrictions to the outputs. Differential privacy as a technical solution is specifically designed to deal with this type of data flow. The Five Safes model mentioned above contains specific considerations about output. This usually involves the application of rules or principles by an output checker.[54] But whatever solution is employed, the overarching principle is the same as with any dynamic data situation: assess the risk and then control it until it is negligible.

[54] For a detailed discussion about the types of output disclosure control used in this model, (Ritchie & Elliot 2015, Griffiths et al 2018). For a discussion about how machine learning models can be disclosive, (Veale et al 2018).

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.

Impact management consists of ADF components 8-10:

8. *Maintain stakeholders' trust*
9. *Plan what to do if things go wrong*
10. *Monitor the evolving data situation*

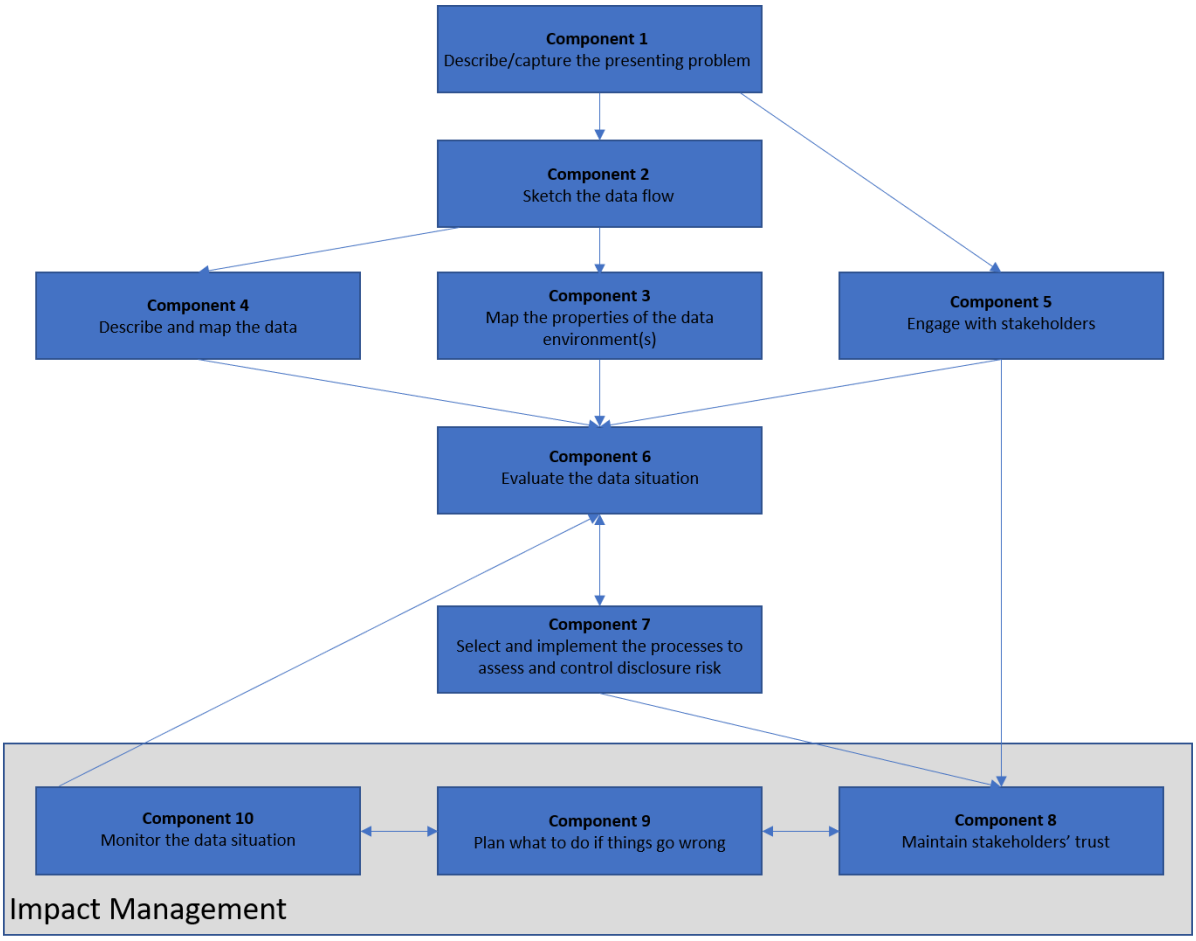


Figure 10: Impact management

Component 8: Maintain Stakeholders' Trust

Overview: *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.*

* * *

You should by now have a strong sense of the relative importance of your stakeholders, and how far their mistrust could complicate your data situation. This should help prioritise your engagement plans – for instance, if your processing is high profile or especially sensitive, then it is important that the media, interest groups and regulators are kept onside through honest communication and accommodation of their preferences wherever possible. If your stakeholders only hear from you in the event of a major problem, they will be far harder to reassure. Furthermore, if, in your data sharing, you are interacting with bodies over which you have little control, transparency on your part will help establish when the responsibility for a problem lies elsewhere.

The initial work of establishing trust has already been done in component 5, but having established trust you need to maintain it going forward. This will mean planning to ensure you keep in touch with stakeholders, and understand their interests as they evolve through time. It may be that a stakeholder changes (for example, a new government may change the political climate, or it may be that a new set of consumers are interested in your data), but in general maintaining trust does not require much more than ensuring lines of communication are kept open, and being attentive and reacting promptly when you are contacted.

- If you have established a particular website as the means for explaining your processing decisions, make sure that it is maintained and kept up to date. Don't let a URL lapse; if your stakeholders get a 404 error when they try to find you, that's hardly going to promote trust. Keep FAQ pages available, relevant and up to date.

- Make sure any email addresses for points of contact, such as info@yourname.com, stay live (if you change them, make sure the old emails are forwarded to the new address), and if answering emails immediately is impracticable, set up an automatic holding reply.

- If you use social media for interacting with stakeholders, ensure it is someone's responsibility to monitor the accounts, take note of problems or discontent as soon as they become evident, and be ready with a consistent and accurate message. Make sure you inform your social networks about any changes in policy. You could also give your opinion of recent events or

controversies via social media, and retweet interesting messages or ideas. Activities such as this can give a sense of the values that your organisation espouses.

- Major changes in policy or decisions can be the subject of a press release, ideally drafted with a specialist communications team (whether or not in-house).
- As part of your general networking activity, make sure you pick up on issues that colleagues or peers are having to deal with. A major event like the Cambridge Analytica scandal of 2018[55] can impact the trust of data managers across the profession whether or not they were involved.
- Where you feel you have to consider the trust of a large group – for example the data subjects in a big dataset, or even the general public – you should already be aware of key representative groups (Article 35(9) specifically mentions consulting data subjects' representatives). Maintain links with these proxies.
- If part of the narrative of your trustworthiness is based upon independent testimonials, then keep these updated, adding new ones where possible.
- If, as part of the process of establishing trust in component 5, you have produced and published a Data Protection Impact Assessment, then also publish any updates or revisions.
- If, as part of the process of establishing trust in component 5, you published a report on your data sharing activities, consider publishing annual or biennial supplements.
- Piggy-back on other events. If stakeholders are engaged in the course of other aspects of business (e.g. marketing, or managing supply chains), then maybe you can scrounge an invitation to a reception, or a half-hour speaking slot at an event.
- If in doubt, pick up the phone. If you have reason to believe a stakeholder is getting restive, and trust is beginning to fray (especially if you don't know why), then reach out, make contact and see if you can gauge their concerns.

This is a fairly long list of bullet points. It may look forbidding. But hopefully most of the work under component 8 is more a matter of keeping already-spinning plates spinning. You should already have built trust, and maintaining it is far easier than establishing it in the first place. It is rare that a completely new stakeholder emerges to be wooed at a later stage. If you can gather evidence of your conscientious practice and trustworthiness as you go, then you should be able to adjust your messaging gradually with the new evidence, occasionally updating the picture. The processes envisaged in component 8 should in the main be periodic adjustments, interventions and checks, rather than a constant drag on your time and resources.

---

55 https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal.

In all these methods of consultation and engagement, be prepared to receive and act upon negative feedback. That doesn't necessarily mean changing your data processing decisions and practice, but you clearly need to put effort into explaining and contextualising them. Where improvements are required, then the value of the feedback will be evident.

Component 9: Plan What to Do if Things Go Wrong

Overview: *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).[56]

* * *

Sometimes, even when you follow all the recommended advice, things can go wrong. As shown in component 2 it is important that you have effective governance policies and procedures in place which essentially identify who does what, when and how, and that you generally support a culture of transparency. A

[56] For more information about personal data breaches, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/.

natural extension of this is putting in place mechanisms that can help you deal with a disclosure in the rare event of one occurring.

9.1 Ensure you have a robust audit trail

Being able to provide a clear audit trail taking into account all relevant anonymisation activities and processes will be crucial for (i) demonstrating that you have followed all correct procedures in the event of a breach or a problem, and (ii) identifying where, if at all, in your processing activities you might need to make changes to prevent a similar occurrence happening in the future. In practice this means keeping clear and up-to-date records of all your processing activities, detailing who did what, when and how. Some of this information can itself increase disclosure risk and thus these records may – by default – be internally facing. Not being transparent about the anonymisation process may, however, affect data utility and for this reason you may wish to provide a top level public narrative about your anonymisation processes. A crucial point is that negligible risks do sometimes still happen (they are not zero) and being able to demonstrate that you carried out a valid risk assessment given the information available to you will be important for all your stakeholders – hence the need for good record keeping prior to the event![57]

9.2 Ensure you have a crisis management policy

A crisis management policy will ensure you deal effectively and efficiently with a data breach were one to occur. It should identify key roles and responsibilities and detail an action plan stating, step by step, the processes that should be followed in the event of a breach.

There are (at least) two key tasks within crisis management: managing the situation and communicating it to stakeholders. These tasks, if taken on by more than one person, require close cooperation from the beginning right through to the post-breach review.

9.2.1 Managing the situation

Set out a plan for managing the situation. The types of activities you will need to cover are outlined in steps 1 to 6 below. Establishing step-by-step what you will need to do will help you both better manage the situation and avoid having to make decisions in haste. In your plan you should identify the person who will take overall responsibility for managing the situation. You should also include a clear description of their responsibilities.

In the event of a data breach your staff will need to know their roles and responsibilities. Your plan should make these clear. For example, when a member of staff first becomes aware of a breach what should they do? Who should they contact and how? What should they do if the person identified as the first point of contact is not immediately available?

[57] This also aligns with Article 30's requirement for record keeping of processing activities, see https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/documentation/what-do-we-need-to-document-under-article-30-of-the-gdpr/.

9.2.2 Steps in a crisis management plan

Everyone in your organisation should know what your strategy is and their role in it. A plan for managing a data breach might include the following steps:[58]

1. **Respond swiftly:** Include in the plan the first series of actions for a range of possible relevant situations and how they might be undertaken. For example, in the event of a breach relating to datasets published on your website immediately take them down from the website. This may seem like closing the stable door after the horse has bolted, which indeed it is, but some of the datasets may not yet have been compromised and you will at least stop anyone else getting hold of them from you.
2. **Assess the impact:** Include in the plan how the potential impact might be assessed and recorded. The key questions here would be:
    a. Can you guesstimate the potential for other copies of the data being in existence – e.g. from knowledge of users, website traffic?
    b. What is the nature of the breach?
    c. Are the data sensitive?
    d. Is anyone likely to be affected by the breach, and if so how many?
    e. What is the nature of the harm likely to be experienced?
3. **Put measures in place to limit the impact:** Include a feedback loop so that once step 2 is completed you can reconsider if any further interim action can be taken. Think through the types of further action that might be required and plan how you would deliver them.
4. **Notify the appropriate people:** Include in the plan details about who should be notified about the breach, how and within what timeframe.[59]
5. **Penalties:** Include in the plan details about any penalties for those responsible for behaviours indirectly or directly leading to a breach. Make sure identified penalties are transparent, fair, consistent and enforceable.
6. **Review the breach and your handling of it:** The aim here is to learn lessons from the event and put procedures in place to prevent a further occurrence. You should stipulate who will undertake the review and within what time frame.

9.2.3 Communicating the situation

Within your crisis management plan you will need to detail a strategy for communicating with key stakeholders, especially those who may potentially be directly affected by the breach, the ICO, the media and other interested parties. You should identify a spokesperson to represent you/your organisation to ensure your messages about the breach and your responses to it are clear and consistent. Transparency is always preferable but you will probably need time to get all the key information together so you may need an initial holding response to stakeholders such as '*we are investigating the matter'*. Nevertheless, it is

[58] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/personal-data-breaches/.

[59] Notification of a personal data breach is mandatory unless the breach is unlikely to result in a risk to the rights and freedoms of natural persons in accordance with Article 33.

important that you are concrete and on the record about what you are doing as early as possible in the process.

9.3 Ensure you have adequately trained staff

You should ensure that all staff involved in your data processing activities are suitably skilled and experienced for the tasks they undertake and that they understand their responsibilities. You will in all likelihood need to conduct training to ensure staff are up-to-date with relevant anonymisation issues. This might take the form of:

- In-house training on the principles and procedures of your data processing activities.
- External training on core factors such as anonymisation issues and techniques, data security, data protection law etc.

Other ways to support the safe handling of data might include:

- Organising regular team meeting/briefings to look at anonymisation issues such as 'what are my responsibilities under a Data Sharing Agreement when processing data from another source?'
- Implementing a staff non-disclosure agreement to provide clear guidance to staff about their data confidentiality responsibilities inside and outside their workplace and when employment at your organisation ceases.

9.4 Ensure you undertake a periodic review of your processing activities

A review process is likely to be most effective if it is undertaken periodically and not just when a crisis occurs. You should stipulate who is responsible for the review, when and how it will be undertaken and within what time frame. For this, you might want to develop your own standardised form that captures your data processing activities and the criteria against which they will be assessed.[60]

[60] These are part of 'information asset registers' mandated by Article 30, which must be kept evergreen, along with the carrying out and maintaining of DPIAs required for high risk profiling (Article 35).

Component 10: Monitor the Evolving Data Situation

Overview: *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.*

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 discussed 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.

* * *

Having shared or released an anonymised dataset, do you need to do anything else in respect of those data? The simple answer is yes. It is our recommendation that you do not just release and forget about your data. Continuing advancements in IT capabilities, supporting ever-greater access to data and capacity for their analysis, and an ever increasing amount of available data, mean that there is always the potential for the data environment in which you have shared or released your data to change. So whilst your data may be considered safe at the time of its release this may not be the case in the medium term. This is a view also taken by the ICO.

> Means of identifying individuals that are feasible and cost-effective, and are therefore likely to be used, will change over time. If you decide that the data you hold does not allow the identification of individuals, you should review that decision regularly in light of new technology or security developments or changes to the public availability of certain records.[61]

[61] ICO guidance on *Determining What is Personal Data*: 9, https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf.

There are a number of measures you can take to monitor the data environment once you have shared or released your data. These measures should include (but are not limited to):[62]

1. Keeping a register of all the data you have shared or released.
2. Comparing proposed share and release activities to past shares and releases to take account of the possibility of linkage between releases leading to a disclosure.
3. Be aware of changes in the data environment and how these may impact on your data. This means:
    i. keeping abreast of developments in new technologies and security that may affect your data situation by, for example, reading technology journals/blogs, watching relevant podcasts and/or attending relevant events;
    ii. monitoring changes in the law or guidance on data sharing and dissemination by engaging with relevant organisations such as the ICO and UKAN; and
    iii. keeping track of current and new public data sources by, for example, reviewing the information available on the Internet and through more traditional sources such as public registers, local community records, estate agents' lists, professional registers, the library, etc.

If possible, you should also keep track of how your data is being used. If you are controlling access this is fairly straightforward. If you are releasing an open dataset then you might want to consider a process whereby users register their intended use before downloading. This type of information is invaluable later when you are considering the next release, developing its use case (component 1) and considering risk/utility trade-offs (component 7).

If your organisation is large enough you may wish to appoint a Chief Data Officer to oversee these activities. Certainly, you will need to ensure someone in your organisation takes responsibility for overseeing these measures.

[62] You should also periodically review data sharing agreements and update them if there are changes to law or to the particular data situation to which they pertain and indeed keep compliance under review.

CLOSING REMARKS

The 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. If you have read all the way through to here then hopefully you will have been convinced that effectively anonymising your data whilst still remaining compliant with GDPR is possible, given a suitable framework and set of tools. The regulation was always intended to facilitate proper and appropriate data sharing and reuse as well as protecting data subjects. The ADF provides a mechanism for realising both of these ambitions.

GLOSSARY

Anonymisation: A complex process to transform **identifiable data** into non-identifiable (anonymous) data. This usually requires that **identifiers** be removed, obscured, aggregated and/or altered in some way. It may also involve restrictions on the **data environment**.

Anonymisation Decision-Making Framework (ADF): The structured framework for rendering the risk of **re-identifying** individuals from a **dataset** negligible, developed by UKAN and the subject of this Guide.

Attribution: The process of associating a particular piece of data with a particular **population unit** (person, household business or other entity). Note that attribution can happen with **re-identification** (if for example all members of a group share a common attribute).

Confidence: A measure, often subjective, of the certainty with which an **intruder** (or matcher within a **penetration test**) would believe that a match between a **population unit** and a **data unit** is correct.

Confidentiality: A quality of information which the **subject** would not reasonably expect to be **disclosed** without their **informed consent**, or some other overriding justification. With **personal data** this concerns the disclosure of **identified** or **identifiable information**.

Data controller: An entity that makes decisions about the processing of some data. Note that being a data controller is not an individual role (in the manner of say a Caldicott guardian with a responsibility in the National Health Service to keep patient data secure) but a relationship between an entity and the data they control.

Data distortion controls: Any method of **disclosure control** that controls **disclosure risk** by manipulating the variable values at the level of individual **data units**.

Data divergence: This represents the differences at record-level between two **datasets** (data-data divergence) or between a single dataset and reality (data-world divergence). Sources of data divergence include: data ageing, response errors, mode of collection, coding or data entry errors, differences in coding and the effect of **disclosure control**.

Data environment: An explanatory concept in the **ADF**, best understood as a context for an item of data.

Data flow: The movement or transfer of data through a system, describing who has responsibility for and access to them, and the contexts in which it is held.

Data intruder: A **data user** who attempts to **disclose** information about a **data subject** through identification and/or **attribution**. Intruders may be motivated intruders or inadvertent intruders. Motivated intruders may be motivated by a wish to discredit or otherwise harm the organisation disseminating the data, to gain notoriety or publicity, or to gain profitable knowledge about particular data

subjects. Inadvertent intruders may spontaneously recognise individual cases within a **dataset**. Data intruders are sometimes referred to as *attackers*, *snoopers* or *adversaries*.

Data minimisation: A long-standing **data protection** principle enshrined in GDPR as a **data controller's** requirement to ensure that the **personal data** they hold should be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. In other words, one should not collect or process more, or more detailed, data than necessary for the purpose.

Data processor: An entity that processes **personal data** on behalf of a **data controller** but makes only non-essential decisions about means, not essential decisions about the ends of that processing.

Data protection: The set of laws, policies and procedures that regulate data processing, in order to, among other things, balance **data subjects' confidentiality** interests, and **data controllers'** interests in extracting value from their data.

Data release: Any process of data dissemination where the **data controller** no longer directly controls who has access to the data. This ranges from general **licensing** arrangements, such as end user licensing where access is available to certain classes of people for certain purposes, through to fully **open data** where access is unrestricted.

Data share: A **dynamic data situation** where the **data controller** has made a decision to allow a fixed set of entities access to a given **dataset**.

Data situation: The relationship between some data and their **environments**, seen as a total system, defined in the **ADF**.

Data situation audit: The initial stage of the **ADF** that clarifies the nature of the **data situation** and the elements that require further analysis.

Data subject: An identified or identifiable living individual (natural person) to whom a particular piece of data relates.

Data unit: A case in a **dataset**; a set of data about a single **population unit**.

Data user: An entity (person or organisation) that processes data. In the context of **anonymisation** it is usually employed to mean that the data are non-personal and therefore users are not **data controllers** or **data processors**.

Data utility: The value of a given **data release** as an analytical resource – the key issue being whether, and how well, the data represent whatever it is they are supposed to represent. **Disclosure control methods** can have an adverse effect on data utility. Ideally, the goal of any disclosure control regime should be to maximise data utility whilst minimising **disclosure risk**, and in practice disclosure control decisions trade off these two parameters.

Dataset: Any collection of data about a defined set of entities. Normally employed to mean data where **data units** are distinguishable (i.e. not summary statistics).

De-identification: The removal or masking of direct **identifiers** within a **dataset**.

Differencing: A **re-identification** attack whereby two different and overlapping codings for a variable (often geography but could be anything) are overlain leading to intersecting categories which contain small numbers of cases.

Differential privacy: A system which limits the amount of information specific to any individual that can be revealed by an analysis.

Direct identifier: Any data item that, on its own, could uniquely identify an individual case. Examples include a **data subject's** name, address and unique reference numbers, e.g. their social security number or National Health Service number.

Disclosive data: Data that allow **data subjects** to be identified (either directly or indirectly) and/or reveal information about data subjects. Data can be disclosive without any actual disclosures having happened.

Disclosure control methods: Methods for reducing the **disclosure risk**, usually based on restricting the amount of, or modifying the, **data released**.

Disclosure risk: The probability that an **intruder** identifies and/or reveals new information about at least one **data subject** in disseminated data. Because **anonymisation** is difficult and has to be balanced against **data utility**, the risk that a disclosure will happen will never be zero. In other words there will be a risk of disclosure present in all useful anonymised data.

Dynamic data situation: A **data situation** where data is being moved from one **data environment** to another.

False negative: Failure to match two related **data units** or a data unit and a **population unit**.

False positive: An incorrect match between two **data units** or between a data unit and a **population unit**.

Formal anonymisation: Any process that removes or masks **direct identifiers** on a **dataset.**

Functional anonymisation: A holistic approach to **anonymisation** which asserts that data can only be determined as anonymised or otherwise in relation to its **environment**.

Harmonisation: The process of recoding a variable on a **dataset** so that it more directly corresponds to an equivalent variable on another dataset.

Identifiable data: Data that contain **indirect identifiers**.

Identifiable individual: An individual (natural person) who can be identified via **indirect identifiers**.

Identified data: Data that contain **direct identifiers**.

Identified individual: An individual (natural person) identified via **direct identifiers**.

Impact management: A process which, acknowledging that the risk of a **disclosure** from data that has been **released** or **shared** is not zero, puts in place strategies to reduce the negative impact of such a disclosure should it happen.

Indirect identifiers: These can in principle include any piece of information (or combination of pieces of information). For example, consider a combination of information for a 'sixteen year old' and 'widowed'. Whilst *age* and *marital status* are not immediately obvious **identifiers**, our implicit demographic knowledge tells us that this combination is rare. This means that such an individual could potentially be **re-identified** by, for example, someone spontaneously recognising that this record corresponded to someone they knew. Also sometimes known as quasi identifiers.

Informed consent: A person's formal agreement to allow **personal data** to be provided for research and statistical purposes, based on full exposure to the facts needed to make the decision intelligently, including awareness of any risks involved, of uses and **users** of the data, and of alternatives to providing the data. A basic ethical tenet of scientific research on human populations, as well as a legal standard under which identifiable **confidential** information can be disclosed.

Intruder testing: See **penetration test**.

K-anonymity: The criterion that there are at least k records within a **dataset** that have the same combination of **indirect identifiers**. Sometimes termed as using a threshold of k (typically k=3 or k=5).

Key variable: A variable common to two (or more) **datasets**, which may therefore be used for **record linkage** between them. More generally, in **scenario analysis**, a variable likely to be accessible to the **data intruder**.

Licence agreement: A permit, issued under certain conditions which enables a researcher to use **confidential** data for specific purposes and for specific periods of time. It includes contractual and ethical obligations, and penalties for improper **disclosure** or use of **identifiable data**.

Metadata-level controls: Disclosure control methods that work by restricting the data rather than **distorting** it. Examples are **sampling**, variable deletion and aggregation/recoding.

Microdata: Records containing information on individual **data units**. Each record may contain hundreds or even thousands of pieces of information.

Noise addition: The **distortion** of data through some random process.

Open data: Data **released** without any access restrictions, usually by publishing on the Internet.

Output statistical disclosure control: A process by which analytical or **tabular** outputs are manipulated so that they are non-personal. This is most relevant to

data centres where access is controlled but the data are highly detailed and would be personal if **released** as open data.

Penetration (pen) test: An approach to **disclosure risk** assessment where an authorised agent attempts to **re-identify** individuals within a **dataset** using other (possibly publicly available) information.

Personal data: Any information relating to a **data subject**.

Personal information: A term used under the Statistics and Registration Service Act (2007) – applying to data that are **released** by Office for National Statistics only – for information that either **directly identifies** an individual case or does so in conjunction with other information that is already in the public domain (published). Information for which identification requires privately-held information does not constitute personal information. Personal information in this definition includes information about the dead as well as the living.

Population: the set of **population units** that a **dataset** is drawn from. The dataset could be a sample and so not all units within the population will necessarily be in the dataset.

Population unique: A record within a **dataset** that is unique within the **population** on a given set of **key variables**.

Population unit: An entity in the world. It is usually employed to mean the sociophysical analogue of a corresponding **data unit** although in any given **dataset** a given population unit may not have a corresponding data unit.

Pseudonymisation: A term defined in GDPR as the processing of **personal data** in such a manner that the personal data can no longer be attributed to a specific **data subject** without the use of additional information, provided that such additional information is subject to technical and organisational measures to keep it separate.

Record linkage: A process by which records about the same **population units** in different **datasets** are combined to produce a single dataset

Re-identification: The discovery of the concealed identity of one or more individuals in a **dataset** by using additional relevant information.

Remote access: On-line access to protected **microdata**.

Remote access server: A system – often virtual – where **data users** do not access data directly but instead submit analytical requests which are run (usually automatically) and then the users provided with the analytical output. That output may be checked for **disclosiveness** or the system maybe set up so as to only allow a restricted range of requests known to produce only safe output.

Remote analysis server: An environment wherein **data users** may submit data analytical requests and receive the output of that analysis, but may not directly view or interrogate the data. There are different forms: some where the analytical requests are processed by humans, and others where they are processed automatically, perhaps through a user-friendly interface. These servers are

considered more secure than **remote access servers** but the usability is lower as exploring the data is more difficult

Respondent: A person who responds to a survey. A respondent might provide data about just themselves but sometimes about others (as well) and data could have been generated without the **data subjects'** knowledge. So a respondent is not necessarily a data subject.

Response knowledge: The knowledge that a given **population unit** is included in a **dataset**. This could be through private knowledge, e.g. that a friend or work colleague has mentioned that s/he **responded** to a particular survey or it could be through simple knowledge that a particular population unit is a member of the population and the data is a full dataset for that population (e.g. a census).

Restricted access: A **data protection** measure that limits who has access to a particular **dataset**. Approved **users** can either have: (i) access to a whole range of raw data in a safe setting and process it themselves or (ii) access to outputs, e.g. tables from the data.

Rounding: A method of **statistical disclosure control** where a figure is rounded off to a defined base; it is most commonly applied to tables of counts. Normally the base is 3, 5 or 10.

Safe setting: An **environment** such as a data lab whereby access to a disclosive **dataset** can be controlled.

Sample unique: A **data unit** within a **dataset** which is unique within that dataset on a given set of **key variables**.

Sample unit: A **data unit** in a **dataset** which is the sample of some **population**.

Sampling: Releasing only a proportion of the original data records on a **microdata** file. In the context of **disclosure control**, a **data intruder** could not have **response knowledge** as s/he could not be certain that any particular person was in the file.

Sampling fraction: The proportion of the **population** contained within a **dataset**. With simple random sampling, the sample fraction represents the proportion of **population units** that are selected in the sample. With more complex sampling methods, this is usually the ratio of the number of units in the sample to the number of units in the population from which the sample is selected. A low sampling fraction can provide some protection to a dataset where an **intruder** might not be able to infer that a **sample unique** is a **population unique**.

Scenario analysis: A framework for establishing the **key variables** that might be used by a data **intruder** to re-identify **data units**.

Secondary differentiation: A strategy adopted by a **data intruder** to distinguish between multiple candidate matches between **data units** and **population units**. For multiple data units matched to a single population unit this involves identifying variables where the two records differ and then targeting resources on establishing the value of that variable for the population unit. For a

single population unit matched against multiple population units this involves identifying which of the population units matches the data units on variables not included in the original match key.

Sensitive variables: Variables contained in a data record that belong to the private domain of **data subjects** who would not like them to be **disclosed**. There is no exact definition given for what is a 'sensitive variable'. The context is important; the distinction between sensitive and non-sensitive can depend on the circumstances. For example, one's religion might be considered as a sensitive variable in some countries and not so in others. GDPR (and other legislation) lists certain types of data as 'special category data', which need to be treated with greater care in order to be GDPR-compliant.

Special unique: A **sample unique** that has a high probability of being a **population unique**. This can be evaluated statistically and also through common sense knowledge. For example, intuitive knowledge of UK demographics will tell you that '16 year old widowers' are unusual. Therefore, if you have one such in your data for a particular geographical area then they may well be a population unique.

Statistical disclosure: A form of data **confidentiality** breach that occurs when, through statistical matching, an individual **data subject** is identified within an **anonymised dataset** and/or confidential information about them is revealed. A statistical disclosure may come about through: (i) the processes of **re-identification** and **attribution** or (ii) the process of attribution alone.

Statistical Disclosure Control (SDC): An umbrella term for the integrated processes of **disclosure risk** assessment, disclosure risk management and **data utility** assessment.

Suppression: A **disclosure control** process where parts of the data are made unavailable to the **user**. All metadata-level controls could be viewed as a form of suppression but the term is more usually used to describe more targeted approaches like cell suppression, the removal of outliers and local suppression of particular values within **microdata** records.

Synthetic data: Data that have been generated from one or more models of the original data, designed to be non-disclosive.

Tabular data: Aggregate information on entities presented in tables.

Target dataset: An **anonymised dataset** in which an **intruder** attempts to identify **data subjects**.

Target variable: Within a **scenario analysis**, information that an **intruder** would like to learn about a **population unit** or units.

REFERENCES

ARBUCKLE, L., & RITCHIE, F. (2019). The five safes of risk-based anonymization. *IEEE Security and Privacy*, 17(5), 84-89. https://doi.org/10.1109/MSEC.2019.2929282.

ARRINGTON, M. (2006). AOL proudly releases massive amounts of user search data. *TechCrunch*, https://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/.

BATESON, N. (1984). *Data Construction in Social Surveys*. London: George Allen and Unwin.

CNN MONEY (2010). *5 Data Breaches: From Embarrassing To Deadly*. https://money.cnn.com/galleries/2010/technology/1012/gallery.5_data_breaches/.

DE MONTJOYE, Y.-A., RADAELLI, L., SINGH, V.K., & PENTLAND, A. (2015). Unique in the shopping mall: on the reidentifiability of credit card metadata. *Science*, 347(6221), 536-539, https://doi.org/10.1126/science.1256297.

DESFONTAINES, D. (2019). *Differential Privacy in Practice*. https://desfontain.es/privacy/differential-privacy-in-practice.html.

DOMINGO-FERRER, J., SÁNCHEZ, D., & SORIA-COMAS, J. (2016). *Database Anonymization: Privacy Models, Data Utility, and Microaggregation-Based Inter-Model Connections*. Morgan & Claypool, https://doi.org/10.2200/S00690ED1V01Y201512SPT015.

DRECHSLER, J. (2011). *Synthetic Datasets for Statistical Disclosure Control: Theory and Implementation*. New York: Springer.

DUNCAN, G.T., ELLIOT, M.J., & SALAZAR-GONZÁLEZ, J.J. (2011). *Statistical Confidentiality*. New York: Springer.

DUNCAN, G. & LAMBERT, D. (1989). The risk of disclosure for microdata. *Journal of Business & Economic Statistics*, 7(2), 207-217, https://doi.org/10.1080/07350015.1989.10509729.

ELLIOT, M.J. & DALE, A. (1999). Scenarios of attack: the data intruder's perspective on statistical disclosure risk. *Netherlands Official Statistics*, 14, 6-10, https://www.researchgate.net/publication/343963431_Scenarios_of_attack_the_data_intruder's_perspective_on_statistical_disclosure_risk.

ELLIOT, M., MACKEY, E., O'HARA, K., & TUDOR, C. (2016a). *The Anonymisation Decision-Making Framework*, Manchester: UKAN, https://ukanon.net/ukan-resources/ukan-decision-making-framework/.

ELLIOT, M.J., MACKEY, E., O'SHEA S., TUDOR, C. & SPICER, K. (2016b). End user licence to open government data? A simulated penetration attack on two social survey datasets. *Journal of Official Statistics,* 32(2), 329-348, https://doi.org/10.1515/JOS-2016-0019.

ELLIOT, M. J. & MANNING, A. M., (2003) *SUDA: A Software Tool For Use With Statistical Disclosure Control For Microdata*. https://www.click2go.umip.com/i/software/suda.html.

ELLIOT, M.J., MANNING, A.M. & FORD, R.W. (2002). A computational algorithm for handling the special uniques problem. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 493-509, https://doi.org/10.1142/S0218488502001600.

ELLIOT, M., O'HARA, K., RAAB, C., O'KEEFE, C.M., MACKEY, E., DIBBEN, C., GOWANS, H., PURDAM, K. & MCCULLAGH, K. (2018). Functional anonymisation: personal data and the data environment. *Computer Law & Security Review*, 34(2), 204-221, https://doi.org/10.1016/j.clsr.2018.02.001.

FIENBERG, S.E., & MAKOV, U.E. (1998). Confidentiality, uniqueness, and disclosure limitation for categorical data. *Journal of Official Statistics*, 14(4), 385-397, https://search.proquest.com/docview/1266844133.

GRIFFITHS, E., GRECI, C., KOTROTSIOS, Y., PARKER, S., SCOTT, J., WELPTON, R., WOLTERS, A. & WOODS, C. (2018) *Handbook on Statistical Disclosure Control for Outputs*. Safe Data Access Professionals Group (SDAP), https://www.cancerresearchuk.org/sites/default/files/thf_datareport_aw_web.pdf.

HAGLIN, D.J., MAYES, K.R., MANNING, A.M., FEO, J., GURD, J.R., ELLIOT, M.J. & KEANE, J.A. (2009). Factors affecting the performance of parallel mining of minimal unique item sets on diverse architectures. *Concurrency and Computation: Practice and Experience*, 21(9), 1131-1158, https://doi.org/10.1002/cpe.1379.

JOBIN, A., IENCA, M. & VAYENA, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389-399, https://doi.org/10.1038/s42256-019-0088-2.

NISSENBAUM, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Palo Alto, CA: Stanford University Press.

OHM, P. (2010). Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review*, 57, 1701-1777, https://www.uclalawreview.org/pdf/57-6-3.pdf.

REITER, J.P. (2005). Estimating risks of identification disclosure in microdata. *Journal of the American Statistical* Association, 100(472), 1103-1112, https://doi.org/10.1198/016214505000000619.

RITCHIE, F. (2017). The 'Five Safes': a framework for planning, designing and evaluating data access solutions. *Data For Policy 2017,* London, http://dx.doi.org/10.5281/zenodo.897821.

RITCHIE, F., & ELLIOT, M. (2015). Principles-versus rules-based output statistical disclosure control in remote access environments. *IaSSIST Quarterly*, 39(2), 5-13, https://doi.org/10.29173/iq778.

RUBINSTEIN, I.S. (2013). Big data: the end of privacy or a new beginning? *International Data Privacy Law,* 3(2), 74-87, https://doi.org/10.1093/idpl/ips036.

SÁNCHEZ, D., MARTÍNEZ, S. & DOMINGO-FERRER, J. (2016). Comment on 'Unique in the shopping mall: on the reidentifiability of credit card metadata'. Science, 351(6279), 1274, https://doi.org/10.1126/science.aad9295.

SARTOR, N (2019) *Explaining Differential Privacy in 3 Levels of Difficulty*. https://aircloak.com/explaining-differential-privacy/.

SKINNER, C.J. & ELLIOT, M.J. (2002). A measure of disclosure risk for microdata. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 855-867, https://doi.org/10.1111/1467-9868.00365.

SMITH, D. & ELLIOT, M. (2008). A measure of disclosure risk for tables of counts. *Transactions on Data Privacy*, 1(1), 34-52, http://www.tdp.cat/issues/tdp.a003a08.pdf.

TAUB, J., ELLIOT, M., PAMPAKA, M., & SMITH, D. (2018). Differential correct attribution probability for synthetic data: an exploration. In Domingo-Ferrer, J. & Montes, F. (eds.) *Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference, PSD 2018*. Cham: Springer, 122-137, https://doi.org/10.1007/978-3-319-99771-1_9.

TROTTER, J.K. (2014) Public NYC taxicab database lets you see how celebrities tip. *Gawker*, https://gawker.com/the-public-nyc-taxicab-database-that-accidentally-track-1646724546.

TVERSKY, A. & KAHNEMAN, D. (1974). Judgment under uncertainty: heuristics and biases. *Science,* 185(4157), 1124-1131, https://doi.org/10.1126/science.185.4157.1124.

VEALE M., BINNS R. & EDWARDS L. (2018). Algorithms that remember: model inversion attacks and data protection law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), http://doi.org/10.1098/rsta.2018.0083.

The 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. Effectively anonymising data whilst still remaining compliant with GDPR is possible, given a suitable framework and set of tools. GDPR is intended to facilitate proper and appropriate data sharing and reuse as well as protecting data subjects. The ADF provides a mechanism for realising both of these ambitions.

Commendations of the 1st edition

"This authoritative and accessible decision-making framework will help the information professional to anonymise personal data effectively. The framework forms an excellent companion piece to the ICO's code of practice."

Elizabeth Denham – UK Information Commissioner

"It is my belief that this book will come to be seen as a gold-standard in the field: it is fundamentally rational, scientifically and technically rigorous, easily understandable and framed in a way that makes them useable. I intend for it to become mandatory reading across my research group."

Professor Paul Burton - Infrastructural Epidemiology at the University of Bristol

Cover Design by Media Services, The University of Manchester.