DARE UK Privacy Risk Assessment Methodology Project (PRiAM) Project: D2 Report v1.1

A Privacy Risk Assessment Framework for Safe Collaborative Research

Risk Tiers for a consistent and transparent use of the five safes framework

UK Research and Innovation

HDRUK Health Data Research UK

ADRUK Data-driven change

DARE UK

Document Details

Date	16/09/22
Deliverable lead	Privitar Ltd
Version	1.1
Authors	Murakonda, S., Weller, S. (Privitar Ltd) Boniface, M., Carmichael, L., Hall, W., McMahon, J., Pickering, B., Surridge, M., Taylor, S., Baker, K. (University of Southampton) Atmaca, U-I., Epiphaniou, G., Maple, C. (University of Warwick)
Contact	suzanne.weller@privitar.com
Dissemination level	Public

Funding statement

This work was funded by UK Research & Innovation [Grant Number MC_PC_21030] as part of Phase 1 of the DARE UK (Data and Analytics Research Environments UK) programme, delivered in partnership with Health Data Research UK (HDR UK) and ADR UK (Administrative Data Research UK).

Disclaimer

This document reflects only the authors' views — the DARE UK programme, HDR UK and ADR UK are not responsible for any use that may be made of the information it contains.

Publication Acknowledgement

This report is independent research supported by the National Institute for Health and Care Research ARC Wessex. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

Executive Summary

Sharing data for research, when carried out responsibly, can have huge public benefits. However, without appropriate protections in place, institutions risk losing the trust of individuals. Hence, privacy risk assessment should be baked into the decision-making processes for sharing or providing access to data. The current approaches for assessing privacy risk are ad hoc, manual, opaque, and inconsistent across different organisations or even different individuals in the same organisation. In this report, we propose a new privacy risk assessment framework that can **improve consistency and transparency** in data sharing decisions. Our intention is to support shared subjectivity in decision-making among various stakeholders and enforce the subjective decisions consistently.

Our privacy risk assessment framework is built on top of the Five Safes, which is widely used across different public institutions in the UK. In the first PRiAM report (D1), we explored how various organisations using the Five Safes framework interpret it differently. It is impossible to assess if the framework is being used effectively, unless more details regarding how each of these safes were accounted for are available. **The proposed privacy risk assessment framework aims to facilitate better usage of the Five Safes**. The key idea is to enable data custodians to explicitly list the criteria they consider for assessing privacy risk, thereby enhancing transparency. These criteria are then used to categorise different data sharing scenarios into discrete tiers of risk that can further be tied to decisions around data sharing, therefore providing consistency in decision-making. Creating discrete levels of risk encourages **comparison-based reasoning** about risk in different scenarios as well as provides a starting point for the **creation of standard benchmarks**.

The privacy risk assessment framework can reduce the burden on data custodians by removing the need for bespoke handling of every single data access request. It enables a level of automation for certain (easy) decisions because the information required for assessing risk and its appropriate usage is already encoded in the framework and these can be mapped onto a standard set of decisions. Moreover, explicitly listing the criteria and decision-making within the framework supports collection of the information necessary in case of a future audit. The framework can also help the researchers by reducing their wait time to get access to data as well as recommending actions to reduce risk. Organisations can use the framework to openly publish the factors they consider when making decisions. By making the measures they take to protect data clearly visible, they can raise public trust.

During the project, we conducted multiple interviews with information governance practitioners developing or running secure research facilities, legal professionals, and academic experts. Insights from these interviews helped us learn the challenges with existing risk assessment frameworks and informed the design of our framework. The two main challenges we repeatedly heard during the interviews are around interoperability and lack of sufficient regulatory guidance. Due to the inherently subjective nature of risk assessment and differences in risk appetite of different organisations, exacerbated by different levels of maturity in technical infrastructure, collaboration across organisations for linking and sharing data can be painstakingly hard. This process can become even more protracted due to the regulatory uncertainties around who becomes the controller for a joined-up dataset. In fact, just determining the legal status of this new data can be difficult. The proposed framework cannot solve any of these challenges on its own. Any risk assessment framework would not be sufficient to address interoperability concerns and legal uncertainties. However, the framework enables a systematic approach to conversations for resolving the differences and building a safe collaboration for linking and sharing data.

1 Table of Contents

List of Figures

Abbreviations

D	Deliverable
DARE UK	Data and Analytics Research Environments UK
DARE UK PRiAM	DARE UK Privacy Risk Assessment Methodology
DPIA	Data Privacy Impact Assessment
ICO	Information Commissioner's Office
NIST	National Institute for Standards and Technology (USA)
PETs	Privacy-Enhancing Technologies
PROCED	PROactive, Collaborative and Efficient complex Discharge
PRAF	Privacy Risk Assessment Forum
TRE	Trusted Research Environment
WP	Work Package

2 Introduction

2.1 Purpose

This report is Deliverable 2 (D2) *"A Privacy Risk Assessment Framework for Safe Collaborative Research: Risk Tiers for a consistent and transparent use of the five safes framework"* of the DARE UK PRiAM project. The report is one in a series of four project reports, which together focus on working towards standardisation of privacy risk assessment for cross-domain access and re-use of sensitive data for research purposes.

2.2 About the DARE UK PRiAM project

The 'Privacy Risk Assessment Methodology' ("DARE UK PRiAM project") project was one of nine projects funded by UK Research and Innovation (UKRI), as part of its DARE UK (Data Analytics and Research Environments UK) Sprint Exemplar Project programme. The eight-month project commenced in January 2022 and completed in August 2022. This research project involved three partner organisations — University of Southampton, University of Warwick and Privitar Ltd — and brought together an interdisciplinary team of data governance, health data science, privacy, public patient and involvement, and security experts from ethics, law, technology and innovation, web science and digital health.

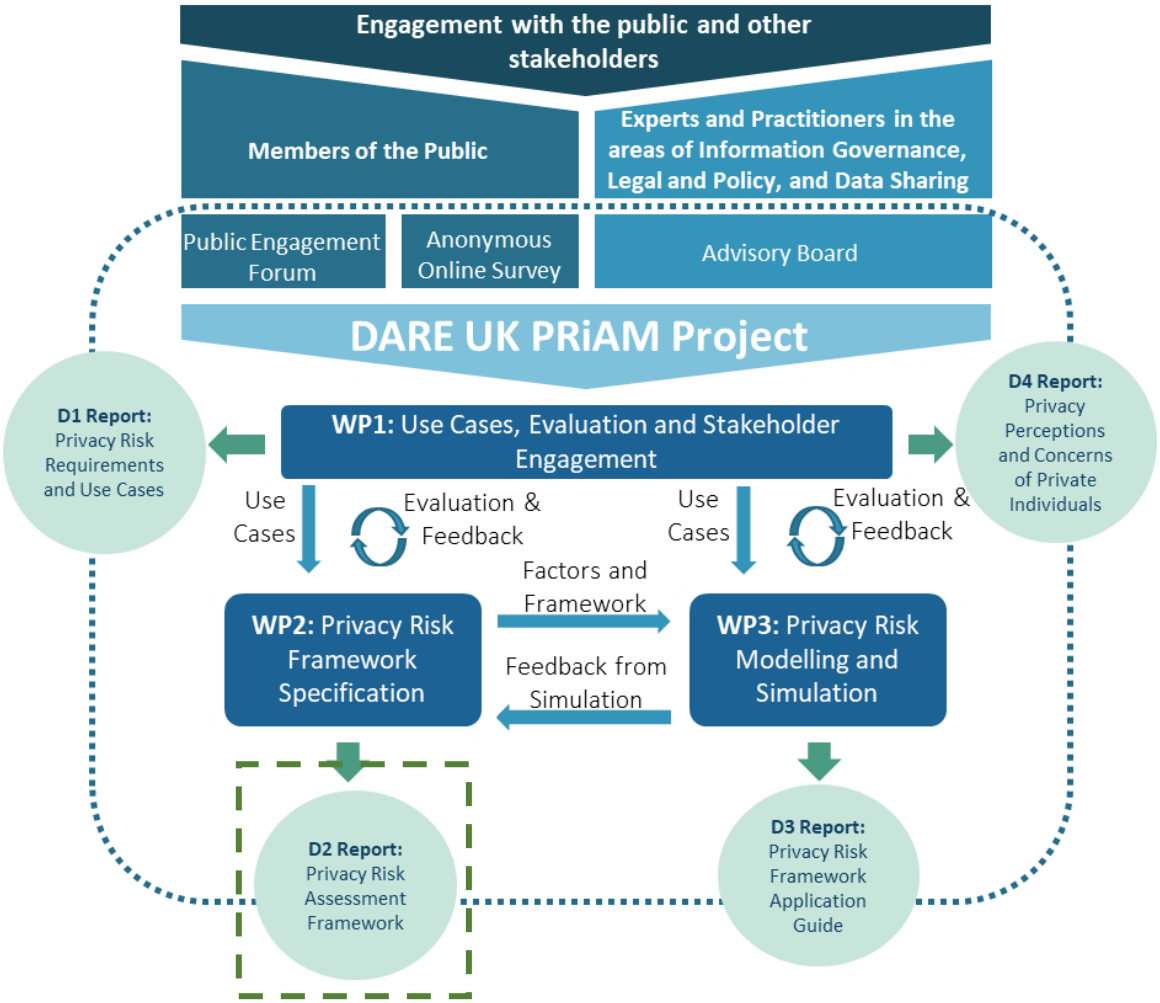


Figure 1: Overview of the DARE UK PRiAM Project: Deliverables, Stakeholder Engagement and Work Packages

2.2.1 Motivation

Trustworthy and collaborative data sharing and re-usage for approved research purposes can help to advance public health and patient care. Data and analytics systems are changing and new ways to share and access data are emerging, including the potential for greater federation[1] of resources and services. Health and social care research often require combinations of data from multiple sources, including data from electronic health records, digital health applications and wearable technologies (e.g., Sharon & Lucivero, 2019). These changes are bringing about new and evolving risks. Organisations responsible for carrying out and facilitating such research activities must ensure that reasonable and acceptable levels of privacy protection are in place so that individuals, groups of people and wider society are not put at risk of undue harm.[2]

2.2.2 Project objectives

This report is one in a series of four project reports, which together focus on working towards standardisation of privacy risk assessment for cross-domain access and re-use of sensitive data for research purposes. Our project objectives are as follows:

- Objective 1: Analyse **driver use cases** in public health prevention and integrated care.
- Objective 2: Identify **key factors contributing to privacy risks** within the Five Safes.
- Objective 3: Define a **risk tier classification framework** to provide a consistent framework for privacy risk assessment.
- Objective 4: Assess privacy risks for use cases using a cyber security **risk modelling and simulation** platform, focusing on privacy risk (re-identification), threats (linking), adversarial conditions (motivations, capabilities and opportunity), controls (homomorphic encryption, parquet encryption).
- Objective 5: Evaluate the framework, modelling and simulation through **engagement with multidisciplinary stakeholders** (e.g., members of the public, research councils, information owners, regulators).

2.2.3 Project structure

Three work packages (WPs) address user needs, privacy risk framework and implementation:

- **WP1 "Use Cases, Evaluation & Stakeholder Engagement"** analyses use cases, requirements, conducts evaluation and captures/disseminates lessons learnt to maximise impact.
- **WP2 "Privacy Risk Framework Specification"** identifies privacy risks factors and develops the risk tier classification framework.
- **WP3 "Privacy Risk Modelling & Simulation"** models risk factors and assesses use cases using the ISO/IEC 27005 information security risk management methodology.

2.2.4 Engagement with the public and other stakeholders

The project has engaged domain experts and members of the public to ensure a broad range of stakeholder interests and opinions are considered. A **Public Engagement Forum** was established with 10 members of the public to explore privacy risk perceptions through a series of four workshops. The Forum discussions were thematically analysed to produce a **survey** for quantitative validation of opinion expressed. This survey was distributed across the UK, with participation from 500 respondents. The outcomes from the Forum and survey are reported in D4 "Privacy Risk Perceptions and Concerns of Private Individuals".

[1] As an example federative approach see the "open, federated and interoperable technology stack for trusted research environments" and "Federated Data Analytics Infrastructure - Capability Maturity Model" outlined by Health Data Research UK (HDR UK, 2021b).

[2] The objective of risk management is "not to eliminate risk, but to reduce the risk as fully as practical" by identifying "'appropriate' responses" that balance benefits and risks effectively and appropriately (Kuner et al., 2015). In other words, those responsible for research taking place as part of safe research collaborations can only offer "reasonable, not absolute, protection" (Shaw & Barrett, 2006) to individuals, communities and wider society.

An **Advisory Board** was established consisting of 21 domain experts, including information governance practitioners, practitioners running or developing secure research facilities, legal professionals, oversight bodies, and academic experts. Using semi-structured interviews, the Advisory Board helped identify and understand the risk factors, controls and decisions related to privacy risk assessment. The outcomes of the Advisory Board are reported in D2 "Privacy Risk Assessment Framework".

2.3 Scope of the D2 report

As succinctly stated in Desai et al., frameworks such as the Five Safes do not solve any problem by themselves. They are just organisational procedures that facilitate discussions among all the stakeholders.

> ***"The Five Safes framework does not solve all problems; indeed, it could be argued that solves no problem, as it is simply a structure and an ethos, helping to frame discussion."***

A lack of transparency around how these discussions and decision-making processes are structured makes it difficult for external stakeholders, e.g., members of the public, to assess if the framework is being used effectively. To make the usage of the framework transparent, we need more details about the factors[3] that are considered under each of these "Safes" and how they are used in the decision-making process.

The goal of this **deliverable (D2) is to systematically capture the factors considered under the Five Safes dimensions (Safe Projects, Safe People, Safe Settings, Safe Data and Safe Outputs) for risk assessment and thus facilitate a better usage of the Five Safes framework**. This report describes our design for a standard framework for risk assessment in the context of providing researchers with access to data. We aim to capture best practices for data governance in a transparent framework, enabling data custodians to make consistent assessments of the risks involved. The framework builds upon the Five Safes, facilitating better and more consistent usage.

In Section 2 of the report, we outline the co-design methodology with the project Advisory Board. This is followed by Section 3 describing the key feedback in terms of requirements, challenges and recommendations from the Advisory Board. These findings contribute to Objective 5 to evaluate the framework with multidisciplinary stakeholders.

> **Project objective 5 of 5:** Evaluate the framework, modelling and simulation through engagement (advisory board, public) with multidisciplinary stakeholders including research councils, information owners, regulators, and public. (WP1, Outcome: evaluation and engaged network of info gov, legal &policy, practitioners and public stakeholders)

Section 4 of the report outlines a set of the common factors that are considered under the Five Safes, this addresses the project Objective 2 to identify these key factors including use cases were datasets are linked via federated

[3] For the purposes of the project, we use the definition of 'risk factor' provided by the U.S. National Institute for Standards and Technology (NIST, 2012): "A characteristic used in a risk model as an input to determine the level of risk in a risk assessment". Reference: National Standards Institute for Technology (NIST), U.S. Department of Commerce. (2012, September). Guide for Conducting Risk Assessments. NIST Special Publication 800-30; Revision 1. Joint Task Force Transformation Initiative: Information Security. Retrieved from: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-30r1.pdf.

networks of Trusted Research Environments, providing access for research programmes or collaborations. We elaborate on these risk factors in the form of a questionnaire in Appendix C.

> **Project objective 2 of 5:** Identify key factors contributing to privacy risks within the Five Safes when datasets are linked as part of research collaborations (including federated research networks). (WP2, Outcome: understanding of privacy risk factors and consequences).

Section 5 of the report helps meet Objective 3, describing the privacy risk assessment framework we have developed in order to use risk factors to make consistent assessments of the level of privacy risk when linking and analysing data. The report describes the framework, how it can be initially configured by data custodians and how it can be used by researchers and data custodians when requesting access to data.

> **Project objective 3 of 5:** Define a risk tier classification framework for consistent assessment considering impact (severity of compromise) and likelihood of privacy risk when data is linked and analysed. (WP2, Outcome: methodology for privacy risk assessment).

Our first deliverable (D1) examines the Five Safes Plus One[4] as a way of categorising privacy risk factors in the context of data-intensive, interdisciplinary health and social care research. Section 4 of D1 discusses various methodologies for privacy risk assessment, types of risk factors, scope of risk assessment, and different privacy protection goals. Hence, **D1 covers different approaches to identifying, organising, and using the factors for risk assessment**, mostly from academic and other existing literature. In this deliverable (D2), we categorised risk factors under the Five Safes to reflect the elements to be included in privacy risk assessment. We engaged with experts to list the factors explicitly, and developed a framework aimed at better usage of the Five Safes for decision making.

In D1, we discuss the impact of federation and emerging data processing patterns in relation to the Five Safes dimensions. We emphasise the importance of highlighting **different types of research collaborations** as part of the **Safe Projects dimension**, draw attention to a **wider range of stakeholders** in the **Safe People dimension**, and emphasise the **complexity of data flows** via the **Safe Settings dimension**. Note that discussions in D1 are at the abstract requirements level whereas in D2, we focus on identifying the detailed factors common to many scenarios.

Our third deliverable **(D3) explores and demonstrates the feasibility of encoding the risk factors identified in D1 and D2 in a systems security modelling tool**[5] [Phillips et al 2020], which supports the ISO/IEC 27005 methodology for information security risk management. D3 specifically focuses on converting and representing the privacy

[4] We refer to the 'Five Safe Plus One' principles to include 'Safe Return' an addition by the HDRA UK (Hubbard et al., 2020). The principle of 'Safe Return' refers to the special case of sending "individual analysis results back to the clinical setting that originated the data and where identities are known" if appropriate, such as for the purposes of "individual clinical care" and "invitations to participate in trials and other research projects" (Hubbard et al., 2020). Reference: Hubbard, T., Reilly, G., Varma, S., & Seymour, D. (2020). Trusted Research Environments (TRE) Green Paper (2.0.0), UK Health Research Data Alliance. Available at Zenodo. https://doi.org/10.5281/zenodo.4594704. Reference: Hubbard, T., Reilly, G., Varma, S., & Seymour, D. (2020). Trusted Research Environments (TRE) Green Paper (2.0.0), UK Health Research Data Alliance. Available at Zenodo. https://doi.org/10.5281/zenodo.4594704.

[5] Phillips, S., Taylor, S., Pickering, J.B., Modafferi, S., Boniface, M., & Surridge, M. (2022, June 20). System Security Modeller. Zenodo. https://doi.org/10.5281/zenodo.6656063.

concepts (risk factors and harms) from D2 in the format of ISO 27005 concepts (assets, controls, risks, threats, vulnerabilities, and consequences). D2 demonstrates a proof of concept for risk modelling and automated risk assessment for a typical research project accessing a trusted research environment. As we explain in section 5.2, the capability to recommend the actions to reduce risk is valuable to the researchers requesting access to data.

Our fourth deliverable **(D4) provides a summary of our findings from public engagement** about people's perception and understanding of privacy risk**.**

3 Codesign of risk assessment framework with the Advisory Board

The privacy risk assessment framework was developed using a user-centred design approach. To understand the current risk assessment processes, requirements and challenges, we identified a wide range of stakeholders involved in data sharing decisions - legal and policy experts, information governance practitioners, as well as academic experts. These stakeholders formed an Advisory Board allowing us to gather insights from different domains and collect feedback on design ideas.

Over the duration of the project, we conducted semi-structured interviews with the advisors in two phases. The first to gather information about the current practices and challenges identified by our experts. The second focused more on validating aspects of our design and gathering feedback.

The interviews were designed to help us understand the current risk assessment process. Specifically, we wanted to learn the risk factors that are currently considered, how they relate to the Five Safes, how stakeholders rank the likelihood and severity of harm associated with these factors, and the main personas involved in assessing risk for a data access request. The questions were kept broad in scope to allow space in the discussion for individuals to bring their perspectives. We also tailored our discussions with each expert using specific questions unique to their background and expertise.

The information we gathered from the interviews was summarised into notes for each insight. We then conducted an affinity mapping exercise to draw out common themes, areas of agreement and areas where there was disagreement or uncertainty. Key themes emerging from the affinity mapping, supplemented by the insights from existing literature discussed in the first PRiAM deliverable, informed the design our risk assessment framework. Our design aims to capture the recommendations and requirements that we learned from the Advisory Board.

4 Key findings from the Advisory Board

Our interviews focused on the scenario of linking datasets from different organisations and sharing them with researchers in a trusted research environment (TRE). This is representative of scenarios within collaborative research networks, such as PRiAM Use Case A: Complex hospital discharge — "PROactive, Collaborative and Efficient complex Discharge" (PROCED) project described in D1. The two key challenges that we consistently heard during the interviews were: 1) Interoperability and 2) Lack of clear guidance around data protection regulations. Any risk assessment framework would not be practical if it fails to account for these challenges. But it should be noted that a framework cannot solve these challenges by itself. All it can do is simplify the conversations necessary to resolve the differences and promote collaboration between organisations. We summarise the challenges and our recommendations in the next sections.

Challenges

C1. Fragmented policies and incompatible infrastructure hinder interoperability

Different organisations have different risk appetites and levels of maturity in technical infrastructure. Often, the infrastructure of one organisation is not compatible with that of others. For collaborations that involve linking and

sharing data distributed across organisations, these differences in risk behaviour and infrastructure become a huge barrier to interoperability. Making the technical infrastructure compatible is more feasible and it's even a necessary condition to build federation of TREs. Unfortunately, there is no easy fix to the former – it will remain a problem for the foreseeable future due to the inherent subjectivity in risk assessment and differences in risk tolerance.

C2. Lack of clear regulatory guidance complicates an already arduous task of linking datasets

Identifying the data controller and determining the legal status of data after linkage can be challenging (e.g., consider the case of linking consented data with unconsented data). Even the simpler case of determining the conditions for "reasonably likely" to assess whether the data has been anonymised requires multiple subjective decisions. The law is purposefully vague to leave space for contextual interpretation. Hence, any attempt towards standardising risk assessment, especially when linking datasets, would inevitably face these issues due to the inherent uncertainty in requirements for compliance with data protection regulations.

Recommendations

R1. Risk assessment should be ongoing and continuous

Risk assessment should not be treated as a one-off activity done at the point of data transfer. It should rather be a continuous process of evaluation and re-evaluation as situations evolve. With changing contexts and data subject expectations over time, risk should be monitored continuously and must be followed up with appropriate actions to minimise or mitigate it. For example, with advances in technology, data collected in the past might now be used to make novel inferences. If the data were collected based on consent, a renewed consent informing the data subjects about the new possibilities might be needed. Hence, there should be a procedure in place for scheduled and surprise audits of data sharing practices. Measures should also put in place to manage the potential impact of harm on data subjects.

R2. Risk assessment frameworks should help communicate the processes around data sharing to non-experts

The primary function of a privacy risk assessment framework is to help the data custodian make informed decisions about data access requests. Equally important is the ability to provide transparency into the decision-making process and communicating it to non-experts. Failing to convince the general public that their data is being handled responsibly and with integrity can damage the trust in institutions. Taking the public through a journey about data sharing practices and the controls in place to protect their privacy can increase trust. Every project need not do this separately. In fact, it would be inefficient if the same work is done by multiple entities. Hence, there is need for a coordinated nation-wide public relations effort to improve the awareness of general public about trustworthy sharing of data for research in TREs.

R3. Risk assessment frameworks should support data custodians to gather the information required for future audits

Risk-based regulations often make it hard for information governance teams to know if they have done enough to manage privacy risk. During audits from external authorities, data custodians should be able to present the measures they have put in place to reduce the potential harms to individuals in order to demonstrate compliance with data protection regulations. An ideal risk assessment framework should support data custodians in gathering this information at the time of approval of data access requests, making it easy to provide the required evidence for risk mitigation measures, in case of an audit or unexpected incident.

R4. Risk assessment frameworks should provide guidance to small and medium sized organisations looking for advice regarding best practices

Any standard risk assessment process should not become an unsurmountable burden on small and medium scale organisations that are already under-resourced. In fact, to promote standardisation, such a framework should support organisations with less mature information governance teams by providing guidance on best practices and the creation of defaults that could be considered benchmarks within an industry.

R5. Risk assessment frameworks should help data custodians clearly express non-negotiable requirements that must be met for access to data

The seemingly simple ability to explicitly capture the non-negotiable requirements from the data custodians can go a long way in building the trust of data subjects. Moreover, this helps the custodian avoid trivial data access requests that should be rejected and provides clarity to the researchers on what actions they need to take in order to obtain access to data.

R6. Risk assessment frameworks should support both Eyes-off and Eyes-on models of TREs

TREs that follow the eyes-off model provide the researcher with a synthetic dataset, which has similar format as the real dataset, to prepare the code that must be run against the real data. The TRE operators then run the computation on behalf of the researcher and share only the outputs with the researcher. This model offers better privacy due to the limited exposure of raw data to researchers and promotes good scientific practices such as formulating the hypothesis to test before looking at the data. Running an eyes-off TRE requires clean data that has already been pre-processed. Most of the current TREs follow the eyes-on model, where researchers get to inspect the original data during their analyses. This model can be helpful when data is noisy, has missing values, or other intricacies that are common in real-world data.

Due to the different purposes these two models serve, we do not expect any one model of TRE to completely replace the other. Hence, risk assessment frameworks should support both these models of TREs if they are to drive standardisation efforts.

R7. Risk assessment should begin by focusing on the purpose of data processing rather than the data itself

When assessing privacy risk, ask *why* the data is being requested before delving into the details of *what* data is being requested. Risks in the project category are often the most subtle, important, and unfortunately tempting to ignore. But thinking about the purpose first helps alleviate some of the non-negotiable concerns due to processing of data, such as aligning with expectations of the public. Although some organisations prefer looking at details of the data requested first, the Advisory Board expressed strong support for approaching the project category risks first. Starting with the project category to understand the purpose of processing, then go through the data flows and ensure that the controls on Safe Settings, Safe People and Safe Outputs that can be offered by every environment is sufficient to mitigate the risk of data entering it.

R8. Risk assessment should be performed, ideally within the same framework, every time data flows from one environment to the other – be it sharing *inputs* to the researcher workspace or *outputs* for publication

Data flows in various forms and formats through different environments during a project. It must be ensured that the required checks are performed every time before data is shared. A risk assessment framework should support assessing risk during all the phases of data movement in a project. For example, in a typical scenario, organisations share data with a trusted research environment, which then provides access to a modified version of this data to researchers in a secure workspace, who get to take out some aggregate statistics for publication. The risk assessment framework should help in all these data sharing decisions starting from how organisations should share data with trusted research environments to what statistics can be allowed for public sharing. This blurs the distinction between input and output from the framework perspective – it is all about data flow and the controls relating to the environment into which it is entering.

Although our Advisory Board generally agreed with this thinking, we observed a strong recommendation for explicitly stating the output controls to give transparency for the processes by which data can leave an environment.

R9. Organisations should actively seek to understand the reasonable expectations of data subjects even if the processing of data has a legal basis

The Advisory Board expressed a strong emphasis on the need to learn public perception before sharing data, even if the sharing has a clear legal basis. It was pointed out that most of the recent projects that suffered a public backlash had a clear legal basis to share the data. Given the potentially huge consequences of not understanding the public expectations, we recommend organisations to actively work with the general public starting from the initial phases of the project. Having a way to clearly communicate about the protection measures in place with non-experts is a must for this process to be successful.

R10. Role of ethics committees in privacy risk assessment should be more clearly defined

Most research projects go through an ethics committee approval process before getting access to data. We learned that the role of these committees is often poorly defined and some of them tend to focus on details that are best assessed elsewhere (e.g., rigour of the research methodology). This can leave less room for assessing the unique aspects of privacy risk (e.g., expectations of the data subjects). Moreover, a lack of clarity on what to anticipate from these committees can be exasperating for researchers. Hence, it is important to scope out the role of ethics committees and providing a clear guidance on what to expect and what *not* to expect from them.

R11. Procedures for researchers to access to data from different TREs should be simplified by standardising the training requirements and governance approaches

As of the time of writing, there is no standard training that is accepted by all the existing TREs. Researchers shouldn't have to go through multiple trainings that are very similar to get access to data from different TREs. We are aware of the proposal to create "research passports" based on an agreed standard of training. We found strong support for this idea during our interviews with the Advisory Board. Along with training, standardising the whole governance procedure at all the TREs would massively reduce the burden on researchers to understand and navigate custom approval processes at each TRE.

R12. A minimal set of standard agreed requirements (technical and procedural) must be met for an environment to be considered a Trusted Research Environment (TRE)

Multiple organisations have setup their own TREs with varied specifications. In order to have an interoperable infrastructure and a standard risk assessment methodology to support it, there must some clear definition of what constitutes a TRE. By this we don't mean that all the TREs should be the same. In fact, given the different purposes they might serve, it may be helpful to have TREs with different capabilities. A minimum baseline of protection measures that are necessary to consider a facility a TRE are essential to ensure interoperability across facilities as well as to create a standard risk assessment approach.

R13. Data sharing agreements/legal contracts should be standardised or at least come with default templates

From our discussions, we learned that often the most time-consuming step in data sharing processes is preparing the legal contracts/data sharing agreements. Although there have been some efforts to create common template that can be shared and repeatedly utilised across use cases, it remains as a major source of friction in the data sharing processes. Part of this can be attributed to the lack of clear guidance on how to interpret data protection regulations in different contexts. We understand that a single template would not work for all data sharing scenarios but the extent to which standardisation efforts in this space can simplify the governance burdens cannot

be over-stated. Common templates used across all TREs that cover typical scenarios would be helpful to speed up this process.

R14. Data custodians should worry about typical risks but also be prepared for worst-case scenarios

Given the potential damage a single bad event could do, it is understandable that most data custodians exhibit a risk-averse behaviour. But they should also consider the harms due to not providing data for research. A common message we have heard across the Advisory Board is to not let worst case theoretical attacks become an impediment to data sharing for research. Instead, data custodians should focus on keeping appropriate measures in place to handle typical risks as well as being aware of and managing worst-case scenarios.

R15. Data custodians should consider concerns beyond re-identification and focus on the harms to data subjects

Assessing the risk of re-identification is a huge part of any privacy risk assessment but identifying individuals is not always necessary to cause harm. Information that applies to groups of individuals or to only one individual who cannot be identified can still inflict reputational harm or be used to discriminate. Hence, data custodians should move beyond discussions of identifiability of data and focus on the potential harms to data subjects (as individuals or groups of individuals) due to the processing of data.

Other findings

Beyond the recommendations stated above, we also identified some risk factors that were commonly considered across organisations when assessing privacy risk. In the following sections, we will describe these risk factors and how they can be used as part of a risk assessment framework designed to address the above recommendations.

5 Identification of privacy risk factors

In PRiAM D1 report, we explored how the Five Safes are used and interpreted differently across different organisations. We also describe a selected number of other approaches to assessing privacy risk. Although different interpretations and approaches exist the PRiAM Advisory Board indicated that despite this, the underlying factors considered when assessing privacy risk are generally very similar.

Objective 2 of our project aimed to identify key factors contributing to privacy risks within the Five Safes. Through the Advisory Board we aimed to uncover the latent factors that organisations are using to assess risk and to list them explicitly to achieve transparency and drive standardisation. We sought to categorise the risk factors we discovered under the dimensions defined by the Five Safes, that is risk factors associated with Projects, Settings, Outputs, People and Data. The schematic diagram in Figure 2 captures an overview of the general risk factors that emerged from our interviews with the Advisory Board. These factors are also described in more detail in Appendix A using an example questionnaire which can be enhanced and further developed by the stakeholders involved in provisioning access to data.

Our use of the term risk factors is aligned with the NIST definition of risk factor as “[a] characteristic used in a risk model as an input to determining the level of risk in a risk assessment”. In PRiAM D1 report, we explored different types of risk factors including assets, consequences, controls, threats and vulnerabilities as well as the relationship between them, under the ISO 27005 framework, that further elaborated through our risk modelling work described in PRiAM D3 report. The risk factors emerging from our interviews covered most of these different types, with the majority can most easily be described as vulnerabilities or controls. The risk factors identified are not exhaustive and further factors will be identified through analysis of current and proposed governance protocols, data processes and changing infrastructures. Our goal here is to show how such factors can be identified, communicated and considered within a risk assessment framework, along with their integration with risk modelling tools (See PRiAM Report D3).

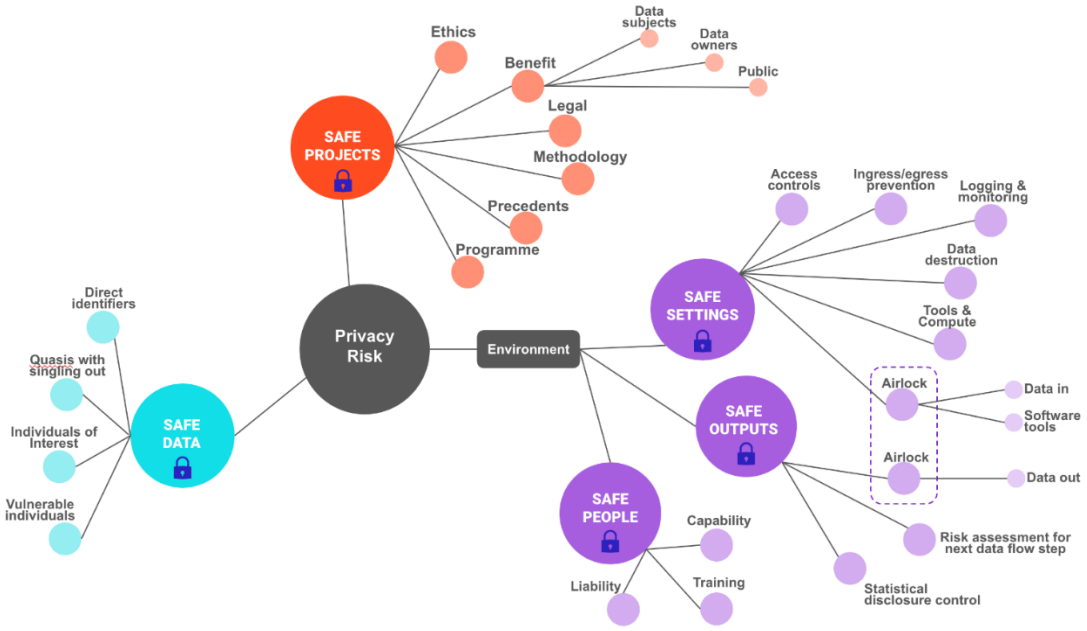


Figure 2: Schematic diagram illustrating some of the risk factors considered under each of the Five Safes. Risk factors are the assets, consequences, controls, threats and vulnerabilities that are used to determine the level of risk in a risk assessment process.

Safe Projects

The risk factors identified under Safe Projects are related to determining whether the use of data is appropriate. For example:

- **Ethics**: Is the proposed usage of data ethical? Has this been determined by an appropriate ethics review board?
- **Benefits**: Does the proposed usage of data demonstrate potential benefit to a variety of stakeholders including data subjects, data owners and society in general?
- **Legal**: Is there a lawful basis for processing the data under appropriate data protection regulations?
- **Methodology**: Does the project have a clear methodology describing how the data will be used to produce the intended project outcomes?
- **Precedents**: Are there previous scenarios where similar projects have been granted access to similar data under the same conditions?
- **Programme**: Is the project part of a wider programme? Does this have any impact on how the data or outputs from the project could be used?

Generally, if the risks indicated for factors included in Safe Projects have not been sufficiently mitigated, it tends to be a showstopper for allowing access to the data. These factors are also the most subjective and hardest to assess, often relying on expert committees whose roles and responsibilities are not always clearly defined.

Safe Settings

The risk factors identified under Safe Settings were generally described in terms of the technical controls in place to protect the data accessed by the researchers. They capture the measures available to mitigate the risk of unauthorised access to data and to ensure that the usage of data is appropriate. For example:

- **Access controls**: Are there controls such as multi-factor authentication mechanisms and role-based access controls to authorise and authenticate researchers?

- **Ingress/egress prevention**: Is the network on which the research takes place isolated? Is there scanning for malware?
- **Logging & monitoring**: Is data about the researcher's activity logged when they use the data? Is there active monitoring and alerting in place?
- **Data destruction**: What happens to data and access for the researcher at the end of the project? How is this implemented?
- **Tools & compute**: Are appropriate data science tools available on the machine where the research takes place? Is there sufficient compute to manage the size of the dataset and types of analysis involved?
- **Airlocks**: What processes are in place to check and approve data that is allowed to enter the researcher workspace, software that is permitted and data that is allowed to leave the environment.

Assessing the protections that are in place can assist data custodians in making decisions about the types of data they are comfortable sharing in various contexts.

Safe Outputs

Similar to Safe Settings, the factors identified under Safe Outputs generally describe the controls in place to ensure that the output of the research does not contain data that could re-identify or disclose information about individuals. For example:

- **Airlock**: is there an airlock where project outputs may be manually inspected to ensure they contain no personal information and are of the type and format expected from the project proposal?
- **Risk assessment process for next data flow step**: for complex data flows, the output from one secure environment may become the input to another data flow step in another environment. A similar risk assessment process should take place before the data is released.
- **Statistical disclosure controls**: have controls such as aggregation with small cell suppression, rounding and differential privacy been applied to ensure the output is not disclosive?

Safe People

A selection of risk factors identified under Safe People are:

- **Capability**: Does the researcher have a track record of successful outcomes? Are administrators of a secure environment, such as a TRE, skilled professionals with suitable qualifications?
- **Training**: Has the researcher recently undergone appropriate training in handling sensitive data and using the tools available in the TRE?
- **Liability**: Are controls such as Data Sharing Agreements in place that clarify the researcher's responsibilities, their liability and that of their institution?

Although traditionally it's only the researchers that are assessed under Safe People, in the context of the emerging complex data flows exemplified by the use cases discussed in D1, the definition of Safe People should be considered to include all stakeholders who may have access to the data.

Safe Data

Risk factors identified under Safe Data are considered when estimating the potential harm (a combination of likelihood and impact) to individuals due to processing of the data. These factors should ideally be a combination of quantitative information to determine the likelihood of re-identification and qualitative information to assess the impact of re-identification. For example:

- **Direct identifiers**: Does the data contain direct identifiers such as passport numbers which would allow individuals to be easily discovered in the dataset?
- **Quasi-identifiers that permit singling out**: Are there combinations of attributes such as date of birth, gender and postcode which when taken together single out individuals and allow them to be re-identified by linking with other data sources?
- **Individuals of interest**: Are there individuals in the dataset who may be a target for a motivated attacker, for example a political figure?
- **Vulnerable individuals**: Does the dataset contain vulnerable groups of people who may be impacted more severely if re-identified from the dataset?

Note that the factors listed in Figure 2 are only reasonable proxies that organisations use for assessing the risk from a dataset. We advocate a rigorous approach, especially the use of risk quantification techniques to assess the likelihood of re-identification and other privacy attacks, such as approaches described in PRiAM Report D3

Note on Privacy enhancing technologies (PETs): PETs such as de-identification, differential privacy and homomorphic encryption can be applied to a dataset and any processing carried out on it. These can be used to reduce the inherent level of risk of a dataset. We have chosen not to include these controls directly as risk factors because their application and the extent to which they reduce risk is highly contextual, moreover the range of technologies and their technical maturity is a fast-moving area. In order to make the framework as robust and future-proof as possible, we propose instead to state factors relating to vulnerabilities in the data. It remains that PETs can be used to control for these vulnerabilities and reduce risk from the data.

Note on linking datasets: When linking datasets from different organisations, we observed that the factors being considered when assessing privacy risk remain the same as above. What changes is how they are applied. Specifically, more attention is paid to the Environment in which linkage happens, the risk from the joined-up dataset, and the legal status as well as controllers of the newly formed dataset.

Note on standardising TREs: At the time of writing, there is no accepted definition of a TRE although specifications related to secure environments do exist[6]. Multiple organisations have developed their own versions of TREs by following different principles and with different sets of controls in place. Explicitly capturing the controls available under the Environment category will provide greater transparency into the processes and thus drive the efforts towards standardisation. This becomes especially important when federating dataflows across TREs.

Note on interdependencies between the Safes: When performing risk assessments, it is useful to consider the notion of an "Environment" that describes the context into which the data will be shared.[7] An environment defines the technical and legal controls, along with the organisational procedures in place in the scope of where the data is shared. For the complex dataflows surrounding TREs, data may pass through multiple Environments such as a main TRE data repository, a workspace specifically configured for a researcher and finally the safe outputs from a research project may enter the public domain. At each dataflow step, considering data entering a new environment allows data custodians to identify exactly which Setting, Output and People need to be considered under each of Safe Settings, Safe Outputs and Safe People.

The concept of an Environment is also useful when considering which combination of factors within Safe Settings, Safe Outputs and Safe People make sense when taken together. For example, absence of security controls to

[6] UK Statistics Authority 2017 defines the Digital Economy Act Accredited Processing Environments (https://uksa.statisticsauthority.gov.uk/digitaleconomyact-research-statistics/better-access-to-data-for-research-information-for-processors/) whilst the NHS England Transformation Directorate has published (Aug 2022) the NHS Secure Data Environment specification for consultation.

[7] A similar reasoning was also advocated in the Anonymisation Decision-Making Framework by UKAN, where the environment into which data is shared is a necessary component of risk assessment. Reference: https://msrbcel.files.wordpress.com/2020/11/adf-2nd-edition-1.pdf

prevent data entering or leaving an environment cannot be compensated for by providing detailed training to individual researchers. Instead, the combination of strong authorisation and authentication controls, together with training for a researcher are required to mitigate both the risk of unauthorised access and the risk of an accidental data breach through poor procedures.

6 Privacy risk assessment framework

Assessing privacy risk, due to its contextual nature, has historically been a challenging and arduous process. Risk-based data protection rules can leave legal, compliance and governance teams unsure about whether they've done enough, as privacy risk assessment is subjective. For example, to determine whether data falls inside the scope of data protection law in the UK, a risk assessment must be carried out to establish whether the risk of re-identification of an individual is remote. Interpretation of "remote" and the factors that need to be considered when making this assessment is left to the subjective judgement of information governance experts within organisations. This uncertainty leads to over-reliance on manual processes and layers of approval, leaving governance teams overloaded in processes and paperwork.

Ad-hoc and manual procedures for assessing privacy risk lead to inconsistency in decision making. This can be inconsistency across decisions made by different organisations or across different people within the same organisation, or even the same person at different points in time. Existing processes can be too lengthy, opaque, and time consuming leaving the researchers frustrated or encouraging them to explore shortcuts. Hence, the objectives we want to achieve with our privacy risk assessment framework are: **ensure consistency in decision making, provide transparency into the process, and automate, at least parts of the process.**

In order to work towards a standard privacy risk assessment framework, we need to identify 1) what information is considered for assessing risk and 2) how this information is being used. We address 1) by explicitly listing the risk factors (see Section 4 and Appendices). To address 2) the factors are grouped together, ranked and assigned to a risk tier. The goal is to assign data sharing scenarios to discrete tiers of risk, and then tie decisions about whether to and how to provide access to data for requests falling into each tier. This approach supports the goals of consistency and transparency as follows:

Consistency

For each data access request, the framework explicitly captures and documents:

- The factors that were used to assess the privacy risk within the Five Safes
- An overall tier representing the level of risk that was determined
- The decision that was made based on this risk tier

The factors considered for any single data access request are encoded in the framework to be used repeatedly for future requests. This means that for each risk assessment, the same factors are considered, therefore scenarios with similar properties will be assigned to the same risk tier leading to the same decision. Moreover, consistent decision-making can happen regardless of who is operating the framework and at what point in time.

Transparency

The framework requires that data custodians explicitly state the factors they consider when making a risk assessment, how this relates to their perception of the overall risk of the data sharing scenario and the decisions that will be made based on this overall risk.

This transparency of the process can be helpful when it comes to providing evidence of responsible data stewardship either during an audit by a regulatory body, or during risk communication with the public.

Transparency is also provided to the researchers requesting access to data. They can clearly understand how their data access request will be evaluated and the factors they need to address in order to ensure a successful outcome. This transparency has the potential to expediate the process by providing researchers with a clear set of requirements for the information they need to provide and the approvals they need to gather in advance.

The framework is designed with the subjective nature of risk assessment in mind; while achieving the goals of consistency and transparency, we want the framework to support a shared subjective view of risk. Hence, the approach is to **have a framework that allows for comparisons and benchmarks to be made** that can be adjusted if needed in exception. The framework supports this by allowing data custodians to rank and group data sharing scenarios of similar risk into discrete tiers. We understand that different organisations have different risk appetites which can be supported by encoding their preferences into the factors and how these are mapped to risk tiers. We note that difference in risk appetites introduces the challenge of interoperability. We do not expect any framework to resolve this issue in full. All we can expect from a risk assessment framework is a process to systematically approach the discussions to achieve consensus and hence interoperability. Our approach supports interoperability conversations by providing the common language of risk factors and risk tiers.

In the following sections, we first discuss the steps that a data custodian needs to follow to set up the framework in their organisation. We explain how researchers can interact with the framework to request access to datasets. We conclude by discussing aspects of customising the framework for a particular organisation and the challenges in achieving interoperability across organisations.

6.1 Setting up the privacy risk assessment framework

Setting up the framework in an organisation is a one-time, initial effort to establish agreed principles on data sharing. Hence, the initial conversations need to involve all the relevant stakeholders. The discussions and decisions made during this setup process will provide transparency and consistency in decision making for all future data requests. Although this is an investment up-front, we expect the costs to normalise over future data access requests. The risk tiers themselves can also be refined over time as more data requests are processed. This could be needed if organisations become custodians of new types data or if new processes are introduced such as review from a new ethics board, introduction of a public involvement and engagement panel or a development in the technologies available to a TRE.

There are three main steps in setting up the framework:

- **Step 1:** Listing risk factors as questions
- **Step 2:** Creating discrete levels along Project, Data, and Environment categories and
- **Step 3:** Mapping combinations of levels to an overall risk tier (and optionally a decision about data sharing). Below, we explain each step in more detail and piece them together to show an example of what an instantiation of the framework might look like in practice.

6.1.1 Step 1: Express risk factors as a set of questions

The first step in setting up the framework is to explicitly list all the criteria i.e., the risk factors that go into decision making about data access. This information can be captured in the form of a questionnaire; aligned with how most DPIAs are currently performed. Questions should be phrased so they have objective answers, even if answering them requires a subjective assessment. This helps in creating a decision-making process that can be consistent across different data access requests.

The framework comes with a default set of risk factors in the form of questions and possible answers (provided in Appendix A). Organisations explicitly list risk factors can enhance transparency in their decision-making processes.

6.1.2 Step 2: Create discrete levels along project, environment, and data categories

In the second step, data custodians map various combinations of answers to the questions created in the first step to different levels of protection from the Environment (including the Setting, People and Outputs) and different levels of risk from Projects and Data. Discrete levels can help in comparative reasoning i.e., although it is hard to reason about the absolute risk in a particular scenario, comparing different scenarios for relative risk and ordering should be easier. As part of this process, the minimum set of criteria that must be met for a data access request to proceed must be stated. Explicitly stating the minimum criteria can help researchers in identifying obvious red flags, if any, in their data access requests.

6.1.3 Step 3: Map combinations of levels to an overall risk tier

Once the levels of protection and risk have been established, the final step is to map these levels along the three categories into an overall risk tier and optionally, tie these tiers to decisions about data access. As a default, it can be helpful to start by considering the Project risk level. If the risks associated a Project are sufficiently low, then the next step is to assess the protection that can be offered by the Environment and decide the level of Data you are comfortable sharing into such an Environment. This is basically treating data as a residual risk. An example formulation of this mapping is shown in the next section.

We understand that certain organisations prefer starting with Data and then handling the rest. However, the Advisory Board strongly supported the approach of considering risk factors under the Project category first and with utmost importance.

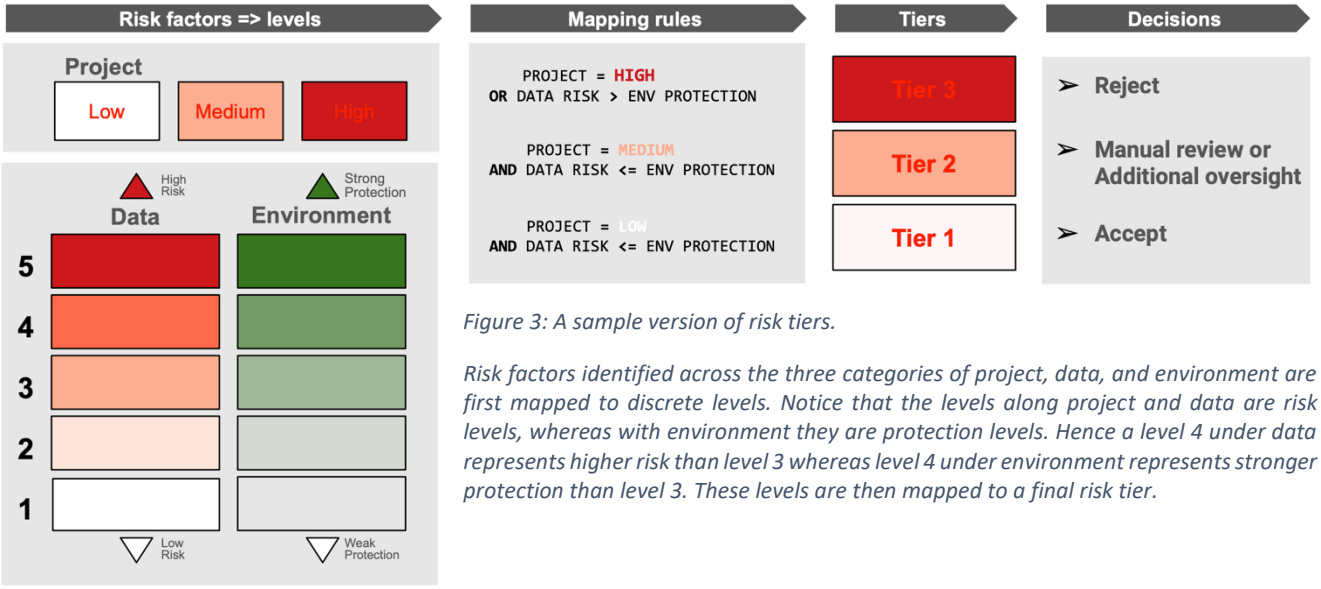


Figure 3: A sample version of risk tiers.

Risk factors identified across the three categories of project, data, and environment are first mapped to discrete levels. Notice that the levels along project and data are risk levels, whereas with environment they are protection levels. Hence a level 4 under data represents higher risk than level 3 whereas level 4 under environment represents stronger protection than level 3. These levels are then mapped to a final risk tier.

A summary of the steps for setting up the privacy risk assessment framework is shown in Figure 4. As explained in the previous subsections, the procedure is to first list the risk factors explicitly, then create discrete levels along the project, data, and environment axes, and finally to map these levels to an overall risk tier.

6.2 Using the privacy risk assessment framework

Once set up by the data custodian, the framework can be used by researchers for making data access requests. They firstly answer the questionnaire which captures the factors that apply for their scenario. Then based on their responses, the risk tier of their request and an associated interim decision will be provided to the researcher. In the best-case scenario, the researcher will get a fast-tracked decision assuming that the information provided can be independently verified by the data custodian. This decision can be either to provide access to the data if the request

is classified as of low risk or to reject access if the request is classified as of higher risk. When the scenario is of high risk, it is also possible to automatically recommend the actions that the researcher can take for moving to a lower risk tier. From a researcher perspective, this capability can help to clarify the actions that are required to obtain access to the dataset. Some example actions to move between risk levels and risk tiers are discussed below.

Moving between risk levels and risk tiers: Different controls can be applied to reduce the risk from data usage. Consider the case of providing access to data for PRiAM Use Case A: Complex hospital discharge — "PROactive, Collaborative and Efficient complex Discharge" (PROCED) project (see PRiAM Report D1). This is a project deemed to be of public benefit and a legal basis to process the dataset. But the source dataset will contain sensitive medical information along with direct identifiers. In order to reduce the risk in this scenario, the organisations involved can choose to either increase the protections from the environment or reduce the risk from data or a combination of both. Techniques for increasing the protection from the environment include controls such as multi-factor authentication, researcher training, activity logging, strict data ingress and egress controls, enforcing contractual controls on researchers, etc. Techniques for decreasing the risk from data include applying redaction, generalisation, tokenization, homomorphic encryption, etc to the data being analysed and statistical disclosure techniques or differential privacy to the outputs of the analysis.

Moving from privacy vs utility to risk vs protection: When transforming data to provide it to the researcher, discussions are generally phrased as privacy vs utility: redacting too much information to protect privacy will reduce the utility of the data for analysis. We suggest moving away from this thinking and start reasoning about the protections that should be in place for sharing a certain dataset. There is no data that is too sensitive that an appropriate environment cannot be created for providing access to it for the right purposes. This need not mean demanding the strictest possible controls for every scenario. Simple transformations to the data, even if insufficient by themselves, when coupled with the right set of environmental controls go a long way in reducing privacy risk.

Note on requesting exceptions: In case the researcher feels their case is an exception (i.e., less risky than what's shown), they can always request a manual review by making a case to the data custodian. The data custodians can keep track of the exception requests and update their default risk levels and risk tiers accordingly.

Note on questions answered by the researcher: the researcher need not answer every single question set by the custodian. For example, they don't need to answer any questions under the Data category. Researchers can just state the data they require and answers to those questions can be automatically inferred by scanning the data. It is also possible to group answers to multiple questions as a certification. This is to simplify answering the questionnaire and exposing a simple interface to the researcher. Instead of answering all the detailed questions, one can associate a particular certification as some chosen answers for a particular set of questions. For example, the ISO 27001 certification can be used to directly answer multiple questions that fall under the Environment category. Similarly, a particular Environment in an organisation can be named and this named Environment can be used to directly answer a set of questions with the appropriate answers.

6.3 Adapting the privacy risk assessment framework to a particular organisation

The framework can be customised to specific needs and/or to encode the preferences of any organisation. It can also be iteratively refined as data access decisions are made. This customisation can be done through a combination of techniques such as:

- Modifying the questionnaire to reflect the organisation's perceived indicators of risk, by adding or removing risk factors to the default list (e.g., removing training requirement for researchers).
- Modifying the possible answers to a question to account for new capabilites available at the organisation (e.g., a more secure environment or a different type of training).

- Customising the number of levels and the rules to assign levels to reflect the breadth of use cases and perception of risk in different scenarios (e.g., having more levels than the defaults or having the same number of levels but assigning them differently).
- Customising the mapping from levels to overall risk tiers or the final decision about the data access request, to better reflect the risk appetite of the organisation.
- Adding sub-questions to provide guidance on how to answer the main question in the framework. Although different organisations care about the same main question (ethical projects), they might use different principles to answer the question. Sub-questions are a way to embed guidance that can help answer the main question. They let organisations encode their principles into the framework.

6.4 Interopertability and challenges in a federated setting

The adoption of a standard framework for assessing risk can solve many interoperability problems. However, multiple organisations may consider risk differently, and therefore interoperability becomes a challenge, particularly for complex dataflows where networks of organisations, environments and research programmes are involved. We believe that our risk assessment framework can aid in conversations around interoperability by *providing visibility* into the decision-making processes and *driving standardisation* efforts.

In the following sections, we describe how the privacy risk assessment framework can support interoperability through visibility and standardisation by using an example use case we detailed in PRiAM Deliverable D1.

6.4.1 Visibility enables interoperability

Consider a simplified version of the dataflow in the PRiAM Use Case A: Complex hospital discharge — "PROactive, Collaborative and Efficient complex Discharge" (PROCED) to provision data from multiple sources into a TRE workspace and provide access to researchers (described in detail in section 2.1 of Deliverable D1). This represents a system with multiple organisations and multiple environments.

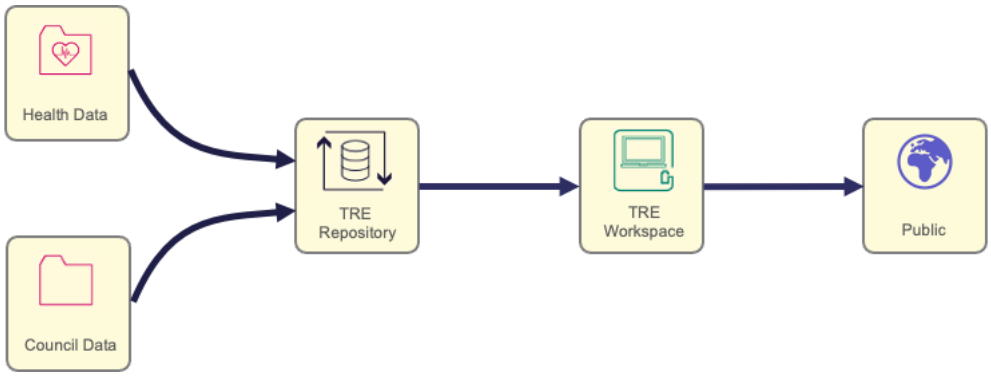


Figure 4: Data flow for the PRiAM Use Case A, where health data and council data are linked to study complex patient discharge scenarios.

Risk assessments take place at each point where data leaves an environment. Providers of Health Data and Council data need to ensure good governance practices when sharing their data with a TRE. TRE Managers need to ensure that only the necessary data is shared into a workspace environment for trained researchers. Finally, when leaving the TRE, researchers and TRE managers need to ensure that public outputs from the research do not allow personal information to be disclosed.

Our framework helps facilitate interoperability in this dataflow by providing visibility into the decision-making processes. Data custodians in the system can openly publish the factors they use to make data access decisions. Organisations managing TREs can publish the sets of controls that are available in their environment. Each entity is now making their interface to the system transparent. This gives interoperability benefits such as:

- Upstream data controllers now have visibility into the controls and risk assessment processes available at dataflow steps further downstream. This helps them to make decisions about whether to share data based on the governance processes that will be in place for further use of the data.
- Upstream data controllers can also use the framework to specify the criteria that are used for downstream decisions, encoding their decision-making process, but allowing it to be administered by someone else. For example, the Health Data Provider in the PROCED use case can specify to the TRE manager the specific instantiation of the framework in terms of the factors, levels, tiers and decisions that must be used when provisioning data to researchers in TRE workspaces.

6.4.2 Standardisation enables interoperability

Consider a scenario where a researcher must run a joint analysis on data held by two different TREs. In order to obtain access to both the datasets, the access request must pass the criteria set by data controllers at each of the TREs. Traditionally, this can be a tedious task as both the TREs might use different and incompatible risk assessment procedures. The framework can help in this scenario as questionnaires from each TRE can be merged, ensuring a single place where the researcher needs to provide information. If we assume a largely similar approach from each TRE, the data access request can be considered and approved by both parties simultaneously.

Standardisation of the controls provided by the TRE as well as training and vetting schemes for the researcher will make it easier to consolidate the risk factors and tiering used by different organisations. For example, it would be a burdensome process for a researcher to undergo a largely similar training programme from three different providers in order to access data in three different TREs. Standardisation of the content and methodology of the training courses should allow the training from a single, accredited programme to be accepted by all the TREs.

By making explicit the common factors used to assess risk, which can then be adopted by a standards body, use of the framework can be a starting point to drive standardisation by reinforcing a common approach.

6.4.3 Interoperability challenges remain

Full interoperability would require all organisations to perceive and treat privacy risk the same way. However, we understand that the context of data sharing may vary significantly between organisations.

If we accept that different organisations must retain the ability to configure the framework with their own risk factors and decision logic, then the requirement exists for a method to consolidate instances of the framework from different organisations and to resolve conflicts. There is no simple answer for this and some of the most difficult scenarios could be where the factors and ranking of factors is largely different between the participating organisations, or where a joined-up dataset has multiple data controllers who do not agree on a data access decision.

For these scenarios the framework can only provide a structured approach, allowing the parties to negotiate and encode their joint view of risk, creating a new instantiation of the framework to support their joint decisions.

7 Possible extensions and future work

The report has outlined the design and use of a privacy risk assessment framework to enable consistent and transparent assessment of privacy risk. As a next step, we recommend the following future work items:

Build and evaluate prototype with network of TREs

The design of the framework was informed by the requirements elaborated in D1 and recommendations from the project Advisory Board. The next stage would be to build a software prototype to enable data custodians to use the

framework in practice. Iterative development of the prototype with users will allow new requirements to be uncovered and usability to be enhanced. Running the prototype with a federated network of TREs will enable interoperability features to be understood and developed such as the ability to merge risk factors from multiple data custodians.

Continue our codesign and framework development with public engagement

Analysis of our public involvement and engagement through the Privacy Risk Assessment Forum (PRAF) has shown that the privacy risk framework provides a mechanism to demonstrate trustworthiness. However, we also found that in order to motivate the public's engagement we would need to ensure they do not feel overwhelmed by the information and effort required (see D4).

Further work is required in cooperation with PRAF to provide lay summaries the risk factors and decision-making aspects of the framework to ensure they provide sufficient transparency to the public.

Forge links to standardisation bodies, industry, independent authorities

Interoperability of the framework will be greatly improved by the standardisation of privacy risk factors within industries and across similar contexts. Discussions with standards bodies and authorities will help establish a path to standardisation of risk factors identified in the project and the framework for assessment.

Appendix A: Example Risk Factors Questionnaire

This questionnaire is only the first step towards curating the factors that must be considered for privacy risk assessment. Although we expect it to be good starting point, it is not comprehensive, and we expect stakeholders to build on top of this to include risk factors that are important in their context.

Safe Projects

1. Does the project have a clear and demonstrable public benefit?
    - ☐ Yes
    - ☐ No
2. Does the project have a legal basis for processing the data?
    - ☐ Yes. Please specify what the legal basis is _____
    - ☐ No
3. Does the project require an ethical committee approval? If yes, is there one?
    - ☐ The project doesn't need an ethics committee approval
    - ☐ The project needs an ethics committee approval but doesn't have one
    - ☐ The project has an ethics committee approval from HRA
    - ☐ Not sure
4. Does the project have processes in place to understand the public perception and their expectations around the usage of this data?
    - ☐ Yes
    - ☐ No
5. Does the project involve novel processing or usage of data that hasn't been encountered before?
    - ☐ Yes, and a DPIA was performed
    - ☐ Yes, but no DPIA was performed
    - ☐ No
6. Does the project involve linking data sets from different institutions?
    - ☐ Yes
    - ☐ No
7. Does the project involve transfer of data across legal jurisdictions?
    - ☐ Yes, but only to jurisdictions that are determined to have an adequate level of data protection
    - ☐ Yes, and a transfer impact assessment was performed

- ☐ Yes, but not no transfer impact assessment was performed
- ☐ No

8. How long would the project access to the data?
    - ☐ Less than six months
    - ☐ Six months – One year
    - ☐ More than one year

Safe Settings

1. Does the environment have the analytics tools and compute capacity required for the project?
    - ☐ Yes
    - ☐ No
2. Physical Location: where can the researcher access the environment from?
    - ☐ Secure analytics site
    - ☐ No control on the physical location
3. Access control: how are researchers authorised to access data/environment and authenticated?
    - ☐ Multi-factor authentication mechanisms
    - ☐ Clearly defined access authorisation procedures
    - ☐ Clearly defined roles and Role Based Access Control (RBAC)
    - ☐ Identity verification when onboarding new users
4. On what devices can the researcher access the data/research environment?
    - ☐ Local machines
    - ☐ Local machines with controls on screen record/in an environment to monitor potential recording
    - ☐ Virtual Desktop Interface (VDI)
    - ☐ Devices in a secure private room
5. Is the plain text data visible to researchers and available for inspection inside the environment?
    - ☐ Yes
    - ☐ No
6. Network: what are the machines running the analytics tools connected to?
    - ☐ No internet connection
    - ☐ Private Virtual network + a separate subnet for each researcher/project
    - ☐ Public internet

7. Data ingress: what are the controls around data ingress to the environment?
    - ☐ Restrictions on “paste”
    - ☐ Restrictions on size of data that can enter the environment
    - ☐ Logging what data gets ingested
    - ☐ Notification mechanism when researcher ingests data
    - ☐ Approval process for every new data to be brought in
    - ☐ Scanning data for potential personal identifiable information (PII)
    - ☐ Malware scanning
8. Data linkage: how is linking with other datasets controlled?
    - ☐ Isolated workspaces for different projects with no linkage allowed
    - ☐ Separate workspaces but permit ingress after approval
    - ☐ Legal contracts to prevent re-id/reuse/linkage
    - ☐ No restrictions on linkage
9. Software: what is the method for bringing in software inside the environment?
    - ☐ Public internet access
    - ☐ Package mirror
    - ☐ Only restricted tools and software
    - ☐ Researchers can install but notifying mechanism in place
    - ☐ Researchers can install anything
10. Activity logging: what information about the researcher's activity is logged when they access data/environment?
    - ☐ Log on/log off to workspace - time, IP address, location etc
    - ☐ Data brought in, if any
    - ☐ Data taken out
    - ☐ Compute deployed
    - ☐ Changes to environment configuration
    - ☐ Edit security settings/access settings
    - ☐ All code that is run against the data
    - ☐ All keystrokes and mouse activity
11. End of project procedure: how is the data/access to data handled after the project expires?
    - ☐ Remove users from the workspace

- ☐ Close the workspace
- ☐ Close the workspace and destroy data

Safe Outputs

12. Data egress: what are the controls and processes for taking data outside the environment?
    - ☐ No restrictions
    - ☐ Only certain users can take the data out
    - ☐ Manual approvals
    - ☐ Scanning data for potential personal identifiable information (PII)

Safe People

13. Researcher training: what are accreditations of the researcher getting access to the data?
    - ☐ Completed the Safe Research Training by the UK Data Service
    - ☐ Completed the NHS Data Security Awareness training course from Health Education England
    - ☐ ONS SRS Provisional researcher
    - ☐ ONS SRS Full researcher
    - ☐ Accredited under the Digital Economy Act
14. Researcher capabilities: what is the skillset of the researcher getting access to the data?
    - ☐ Experience working with noisy real-world data
    - ☐ Experience working with similar data/domain expertise
15. Researcher liability: in what capacity are you requesting access to the data?
    - ☐ As an independent, individual researcher
    - ☐ As member of a researcher group
    - ☐ As an individual researcher affiliated with an institution
16. Contractual controls: what clauses are in place with the researcher or the affiliated institutions?
    - ☐ No attempts to actively re-identify individual
    - ☐ No linking with other datasets
    - ☐ No usage for projects beyond the approved one
    - ☐ No forward sharing
    - ☐ No retention beyond project deadline
    - ☐ Agreed consequences on breach of contract

Safe Data

1. To assess what's permitted: which of the following properties does the requested data satisfy?

    - ☐ Has direct identifiers
    - ☐ Has quasi-identifiers that permit singling out
    - ☐ Is historically shown to be vulnerable to privacy attacks (e.g., location data, payment histories, genome sequences, data from fitness devices, etc.)
    - ☐ Has individuals of interest
    - ☐ Has vulnerable individuals
    - ☐ Has information about a more than a threshold number of individuals (custom value)
    - ☐ Is considered special category data by data protection regulations
    - ☐ Is considered sensitive of some type (IP, commercial, etc)

2. To assess what's required: which of the following properties does the requested data satisfy?

    - ☐ Might have information about more individuals than necessary for the project
    - ☐ Might have more information about individuals than necessary for the project
    - ☐ Might have information about fewer individuals than necessary for the project
    - ☐ Might have less information about individuals than necessary for the project

Appendix B: Example Risk Levels

Note that these example risk levels, based on the risk factors in Appendix A, are only for illustration. **They should neither be treated as final recommendations nor as a comprehensive list of scenarios mapped to various levels.**

Safe Projects - Risk Levels

Questions	Answers	Low	Medium	High	High
Does the project have a clear and demonstrable public benefit?	Yes	✓	✓		✓
	No			✓	
Does the project have a legal basis for processing the data?	Yes	✓	✓	✓	✓
	No				
Does the project require an ethical committee approval? If yes, is there one?	The project doesn't need an ethics committee approval				
	The project has an ethics committee approval from HRA	✓	✓	✓	✓
	The project needs an ethics committee approval but doesn't have one				
	Not sure				
Does the project have processes in place to understand the public perception and their expectations around the usage of this data?	Yes	✓	✓	✓	
	No				✓
Does the project involve novel processing or usage of data that hasn't been encountered before?	Yes, and a DPIA was performed		✓		
	Yes, but no DPIA was performed				
	No	✓		✓	✓
Does the project involve linking data sets from different institutions?	Yes		✓		
	No	✓		✓	✓
Does the project involve transfer of data across legal jurisdictions?	Yes, but only to jurisdictions that are determined to have an adequate level of data protection				
	Yes, and a transfer impact assessment was performed		✓		
	Yes, but not no transfer impact assessment was performed				
	No	✓	✓	✓	✓
How long would the project access to the data?	Less than six months	✓			
	Six months – One year		✓		
	More than one year			✓	✓

Environment: Safe Settings, Safe Outputs, Safe People - Protection Levels

Questions	Answers	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	LEVEL 4	LEVEL 5
Does the environment have the analytics tools and compute capacity required for the project?	YES	✓	✓	✓	✓	✓	✓
	NO						
Physical Location: where can the researcher access the environment from?	Secure analytics site						✓
	No control on the physical location	✓	✓	✓	✓	✓	
Access control: how are researchers authorised to access data/environment and authenticated?	Multi-factor authentication mechanisms	✓	✓	✓	✓	✓	✓
	Clearly defined access authorisation procedures	✓	✓	✓	✓	✓	✓
	Clearly defined roles and Role Based Access Control (RBAC)	✓	✓	✓	✓	✓	✓
	Identity verification when onboarding new users	✓	✓	✓	✓	✓	✓
On what devices can the researcher access the data/research environment?	Local machines	✓				✓	
	Local machines with controls on screen record/in an environment to monitor potential recording			✓			
	Virtual Desktop Interface (VDI)		✓		✓		
	Devices in a secure private room						✓
Is the plain text data visible to researchers and available for inspection inside the environment?	YES				✓		
	NO					✓	
Network: what are the machines running the analytics tools connected to?	No internet connection						✓
	Private Virtual network + a separate subnet for each researcher/project	✓	✓	✓	✓	✓	
	Public internet						
Data ingress: what are the controls around data ingress to the environment?	Restrictions on “paste”	✓	✓	✓	✓		✓
	Restrictions on size of data that can enter the environment	✓	✓	✓	✓		✓
	Logging what data gets ingested	✓	✓	✓	✓		✓
	Notification mechanism when researcher ingests data				✓		✓
	Approval process for every new data to be brought in				✓		✓
	Scanning data for potential personal identifiable information (PII)				✓		✓

	Malware scanning						✓
Data linkage: how is linking with other datasets controlled?	Isolated workspaces for different projects with no linkage allowed				✓	✓	✓
	Separate workspaces but permit ingress after approval			✓			
	Legal contracts to prevent re-id/reuse/linkage	✓	✓		✓		✓
	No restrictions on linkage						
Software: what is the method for bringing in software inside the environment?	Public internet access						
	Package mirror	✓	✓			✓	
	Only restricted tools and software				✓		✓
	Researchers can install but notifying mechanism in place			✓			
	Researchers can install anything						
Data egress: what are the controls and processes for taking data outside the environment?	No restrictions						
	Only certain users can take the data out			✓			
	Manual approvals	✓	✓	✓	✓	✓	✓
	Scanning data for potential personal identifiable information (PII)				✓	✓	✓
Activity logging: what information about the researcher's activity is logged when they access data/environment?	Log on/log off to workspace - time, IP address, location etc				✓		✓
	Data brought in, if any			✓	✓	✓	✓
	Data taken out	✓	✓	✓	✓	✓	✓
	Compute deployed				✓	✓	✓
	Changes to environment configuration				✓		✓
	Edit security settings/access settings				✓		✓
	All code that is run against the data				✓	✓	✓
	All keystrokes and mouse activity						✓
End of project procedure: how is the data/access to data handled after the project expires?	Remove users from the workspace	✓	✓	✓	✓	✓	✓
	Close the workspace			✓	✓	✓	✓
	Close the workspace and destroy data				✓		✓
Researcher training: what are accreditations of the researcher getting access to the data?	Completed the Safe Research Training by the UK Data Service						
	Completed the NHS Data Security Awareness training course from Health Education England						
	ONS SRS Provisional researcher						
	ONS SRS Full researcher		✓	✓	✓		✓

	Accredited under the Digital Economy Act						
Researcher capabilities: what is the skillset of the researcher getting access to the data?	Experience working with noisy real-world data		✓	✓	✓	✓	✓
	Experience working with similar data/domain expertise				✓	✓	✓
Researcher liability: in what capacity are you requesting access to the data?	As an independent, individual researcher					✓	
	As member of a researcher group			✓			
	As an individual researcher affiliated with an institution			✓	✓		✓
Contractual controls: what clauses are in place with the researcher or the affiliated institutions?	No attempts to actively re-identify individual		✓	✓	✓	✓	✓
	No linking with other datasets		✓	✓	✓	✓	✓
	No usage for projects beyond the approved one		✓	✓	✓	✓	✓
	No forward sharing		✓	✓	✓	✓	✓
	No retention beyond project deadline		✓	✓	✓	✓	✓
	Agreed consequences on breach of contract		✓	✓	✓	✓	✓

Safe Data - Risk Levels

Questions	Answers	Level 1	Level 2	Level 3	Level 4	Level 5
To assess what's permitted: which of the following properties does the requested data satisfy?	Has direct identifiers					✓
	Has quasi-identifiers that permit singling out	✓	✓	✓	✓	✓
	Is historically shown to be vulnerable to privacy attacks (e.g., location data, payment histories, genome sequences, data from fitness devices, etc.)				✓	✓
	Has individuals of interest			✓	✓	✓
	Has vulnerable individuals			✓	✓	✓
	Has information about a more than a threshold number of individuals (custom value)			✓	✓	✓
	Is considered special category data by data protection regulations		✓	✓	✓	✓
	Is considered sensitive of some type (IP, commercial, etc)	✓	✓	✓	✓	✓

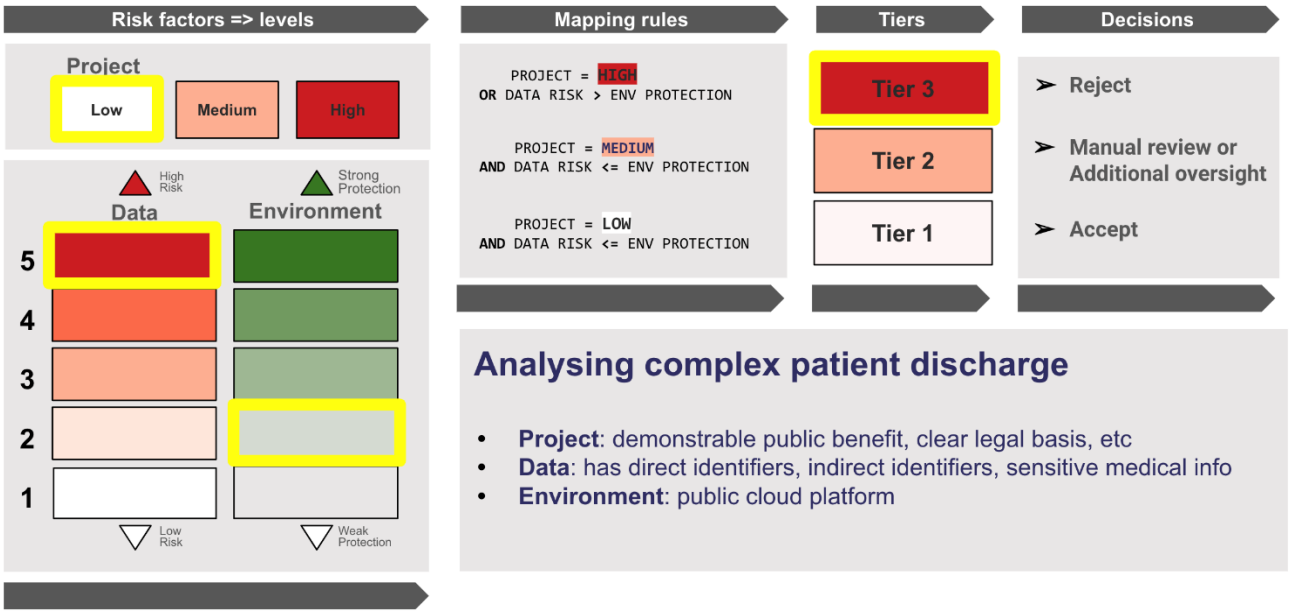
To assess what's required: which of the following properties does the requested data satisfy?	Might have information about more individuals than necessary for the project				✓	✓
	Might have more information about individuals than necessary for the project				✓	✓
	Might have information about fewer individuals than necessary for the project				✓	✓
	Might have less information about individuals than necessary for the project				✓	✓

Appendix C: Worked Example

Here we provide a simple illustrative example of how our framework can be used to assess risk and make decisions regarding data access for a scenario related to complex patient discharge. This is a simplied version of PRiAM Use Case A: Complex hospital discharge — "PROactive, Collaborative and Efficient complex Discharge" (PROCED) project described in PRiAM Report D1.

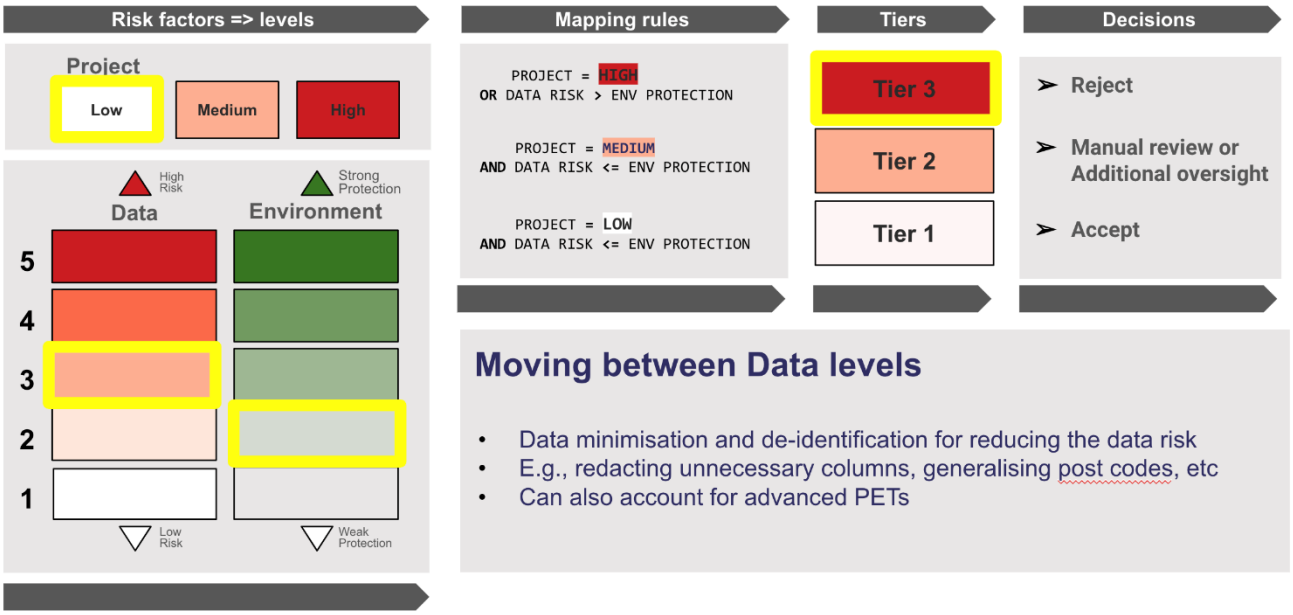
As a starting point for the example, we assume that the data custodians have already instantiated a version of the framework, including the risk factors they consider, grouping of the factors into risk levels, rules for mapping these to risk tiers and the decisions they will make based on the final tier.

Now we take a hypothetical scenario of a researcher requesting access to the dataset for analysis with the purpose of understanding the factors which affect patient discharges from hospital. For this request: the project is of potential public benefit, there is a appropriate legal basis and all other risk factors under Safe Projects are indicative of a low risk level. Let's consider an imaginary case where the data requested contains both direct identifiers – e.g. NHS numbers, indirect identifiers such as date of birth and postcode, and sensitive medical information about a patients diagnosis and treatment. This clearly falls into the highest level of risk from the data. Finally, say the proposed environment in which the research will be performed is a public cloud platform with no security measures. This falls into a low level of protection for Settings, People and Outputs. This set of factors maps onto the highest risk tier, and the data access request is automatically rejected.

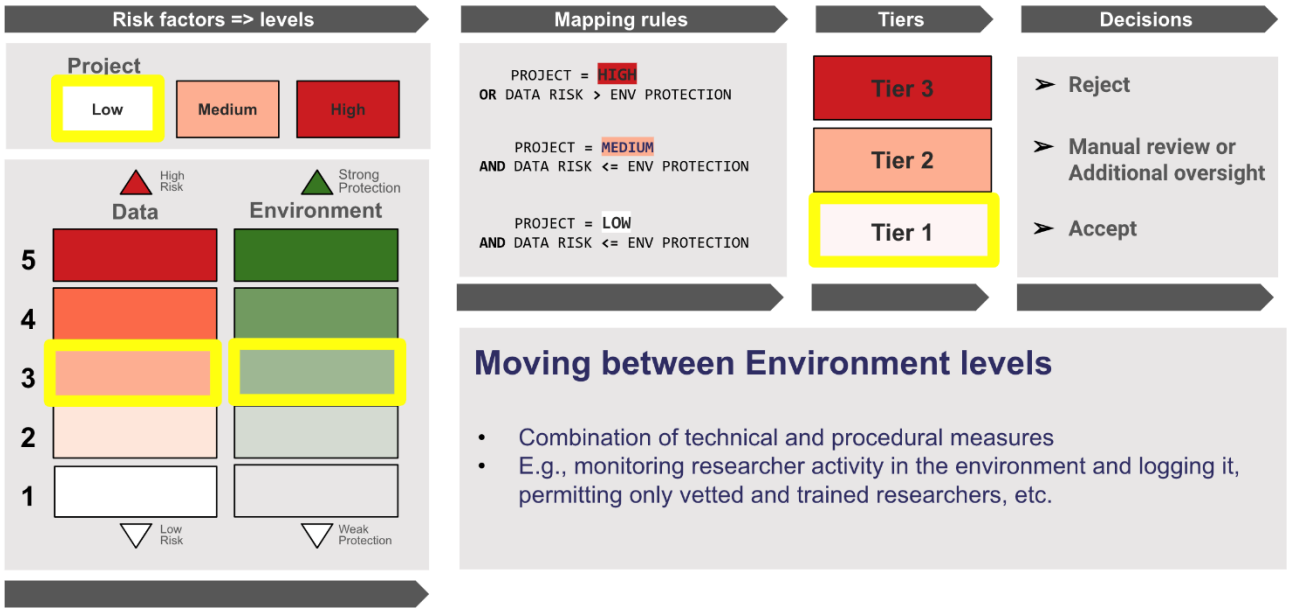


The framework supports the researcher in understanding what protections can be applied to reduce the privacy risk. As a first step, data minimisation and de-identification can be applied to tokenise direct identifiers, redact unnecessary columns and generalise fields such as postcode to a wider geographic area.

This action can reduce the level of risk from the data to say a medium risk level 3. However, the environment in which the data is processed still lacks the protections needed to mitigate against this level of data risk. Therefore, the request is still assessed as a high-risk tier and is rejected.



To improve the protection level of the environment, multiple controls can be considered relating to Safe Settings, Safe Outputs, and Safe People. For example, say the data will be provisioned into the secure workspace of a TRE with strict access controls and network isolation. The TRE might also have an airlock process which ensures that insights produced from the research do not allow any personal information to be inferred. Together with training of researcher in using the TRE and handling sensitive data, these measures can significantly increase the protection level of the environment. Now the combination of reduction in risk from the Data, and stronger protections on the Setting, Outputs and People mean that the overall privacy risk is low, and the data access request can be accepted.



Appendix D: Interview Questionnaire

7.1 Part One: Data sharing processes

In your experience what is the process followed when sharing data with research projects?

- How do researchers know what data is available in your organisation?
- How can researchers request data for research projects?
- Do you prepare and curate data in response to data access requests?
- Do you prepare and curate data in anticipation of future research projects?
- How do researchers assess whether your data will support the research they intend to carry out?
- How do you determine the detail and granularity of data that is shared?
- How do you record the decision-making process?

Who are the main people, groups, departments or external organisations involved in the process? What is your role in the data sharing process?

- Do other external organisations advise you on any aspects of the data sharing process?

7.2 Part Two: Risk factors and controls

When making decisions of sharing data or giving access to data, what are risk factors under consideration?

- Which assurances do you consider on the ethics of research projects?
- Which assurances do you consider on the ability of the people involved to handle sensitive data?
- Do you consider the environment in which the data will be provisioned - for example sharing data directly compared to a TRE?
- What are the barriers to sharing data more widely? What is sensitive in a dataset?
- Do you consider the granularity and detail of what to share on a project-by-project basis? What are the factors that determine this?
- Do you consider the data outputs of a research project and how these will be disseminated?
- Are any other risk factors considered?

What controls or mitigations are available to you for mitigating/reducing the risk?

- How do you manage output controls?

Can you think of example data sharing scenarios where higher risk on one aspect e.g., data/project was compensated with stronger controls on other aspects such as settings/people?

- Do you classify the level of risk for data sharing for different projects? How do you use this classification?

In each of the areas we discussed, what do you consider to be the highest level of risk? Why?

- How to communicate this level of risk and who are the relevant people who need this information?

7.3 Part Three: Decisions and outcomes from risk assessment

What are the decisions and the outcomes linked with the risk assessment?

- What are the most common outcomes?
- Are there exceptions? E.g., inform decisions such as whether a Data Protection Impact Assessment (DPIA) is required?

What frameworks or guidance do you follow to make these decisions?

- What are the challenges for implementing this in practice?
- How are you translating these guidance into practical principles?

7.4 Part Four: Miscellanous

What goes into data sharing agreements and how are they created?

- Why are they time-consuming?
- How can we create standard templates?
- Can trusted research environments (TREs) help?

How do you match the purpose of data collection with the purpose to share data for research?

- Does this change from project to project?

How are the thresholds for functional anonymity/effectively anonymised data determined?

Appendix E: Acknowledgements

We would like to thank all members of our Advisory Board for sharing their valuable insights and helping us shape the framework.

Special thanks the National Data Guardian panel for our presentation and helpful feedback.

We would also like to thank Guy Cohen, Marcus Grazette, Hector Page, Alexandra Mitchell, Nicholas O'Donnell-Hoare, Chris Cannon, Semi Lee, Paul McCormack, Stefan Tzanev for their inputs. Finally, we express our heartfelt gratitude and thanks to Andrew McCallum for coordinating and managing this deliverable.

Any errors in the report are solely of the authors and no information in it should be individually attributed to the advisors or the organisations they are affiliated with.