

# DARE UK

## DARE UK Privacy Risk Assessment Methodology Project (PRiAM) Project: D1 Report v2.0

### Privacy Risk Assessment Requirements for Safe Collaborative Research

Exploring Emerging Data Patterns and Needs of Advanced Analytics  
in Cross Council Research Networks through Use Case Analysis



UK Research  
and Innovation

**HDRUK**  
Health Data Research UK



**ADRUK**  
Data-driven change

# Document Details

<b>Date</b>	16/09/22
<b>Deliverable lead</b>	University of Southampton
<b>Version</b>	2.0
<b>Authors</b>	Boniface, M., Carmichael, L., Hall, W., McMahon, J., Pickering, B., SurrIDGE, M., Taylor, S. (University of Southampton)  Atmaca, U-I., Epiphaniou, G., Maple, C. (University of Warwick)  Murakonda, S., Weller, S. (Privitar Ltd)
<b>Contact</b>	m.j.boniface@soton.ac.uk
<b>Dissemination level</b>	Public

## Licence

This work is licensed under Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)



To view this licence, visit (<https://creativecommons.org/licenses/by-nc-sa/4.0/>). For reuse or distribution, please include this copyright notice.

© Copyright University of Southampton and other members of the DARE UK PRiAM Consortium 2022

## 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.

## Further dissemination

An overview of this work — entitled 'Towards a Socio-Technical Approach for Privacy Requirement Analysis for Next-Generation Trusted Research Environments' — was presented at the CADE 2022 Conference (Competitive Advantage in the Digital Economy) on 13 June 2022.

## Executive Summary

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 of resources and services. These changes are bringing about new and evolving risks. What remains vital is that people are protected from harms associated with data disclosure and re-use — and that public confidence and engagement in health and social care research are maintained. As such, the DARE UK PRIAM project aims to explore methods and tools that can support decision-makers, patients and the public to assess and manage privacy risk when considering emerging data access and re-usage scenarios, such as federation.

This report describes privacy requirements and use cases for cross-domain access and re-use of sensitive data for research purposes, taking into consideration emerging data usage patterns and needs. This report is the first 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. The report specifically focuses on three main areas:

- **Three driver use cases** are outlined as exemplars of cross-domain linkage and analysis related to public health research and integrated care.
- **Emerging data usage patterns and data sharing needs** in operational health data networks are explored, concentrating on trusted research environments (TREs) as facilitators of federated sharing and processing of data.
- **Some different approaches to identifying, organising and using risk factors** for privacy risk assessment are examined through a literature review.

We now summarise some of the key points highlighted in this report:

### Outlining three driver use cases related to public health research and integrated care

UK Research and Innovation (UKRI) cross council research utilising advanced analytics methods — artificial intelligence/machine learning (AI/ML) — for health and social care transformation often require data from multiple sources, including electronic health records, digital health applications and wearable technologies. As part of this project, we focus therefore on research taking place between the Medical Research Council (MRC) — in relation to health, Economic and Social Research Council (ESRC) — concerning social science and social care, and Engineering and Physical Sciences Research Council (EPSRC) — with regard to computer science.

**Three real-world uses cases are outlined as exemplars of access and re-usage of cross-domain sensitive data:**

	Use Case A: Complex Hospital Discharge	Use Case B: Multi-morbidity Prevention	Use Case C: Sub-national federated trusted research environment ecosystem
<b>Project title:</b>	“PROactive, Collaborative and Efficient complex Discharge” PROCED Project	“Multidisciplinary Ecosystem to study Lifecourse Determinants and Prevention of Early-onset Burdensome Multimorbidity” MELD-B Project	“NHSx Wessex Federated TREs”
<b>Project aim:</b>	Focuses on a workflow supporting the proactive	Centres on using Artificial Intelligence (AI) applied to health	Purpose is to pilot a Wessex-wide federated TRE ecosystem

	assessment of patients' onward care needs and discharge risks necessary to minimise discharge delays and improve recovery; and optimising the planning and allocation of community resources against predicted hospital demand.	data, to identify primary health conditions that may progress to burdensome multiple long term conditions if appropriate interventions are not made. The project seeks to identify causal conditions from lifecourse data that may be used to build an intervention strategy.	that brings together the population of Dorset, Hampshire and Isle of Wight Integrated Care System (across two integrated care systems [ICSs]) and the reach of the region's main tertiary referral centre, University Hospital Southampton, and affiliated NHS organisations.
<b>Data linkage:</b>	Involves individual linking of complex multi-stakeholder datasets (e.g., acute care, community care, local authority, etc.) regarding medical stability, patient and family capacity, ongoing care environment, and association with system capacity/demand.	Requires both individual linking and federated learning between longitudinal birth cohorts and routine data sets over the lifecourse.	Necessitates data linkage across clinical care, social care, mental health, and other public administration services.

As a brief summary, the three driver use cases:

- Provide examples of interdisciplinary research.
- Demonstrate how research projects related to public health and integrated care, involving advanced analytics, require a considerable number of connected, multi-stakeholder data sources; all necessitating a degree of federation in terms of the resources and services that are required to achieve their shared goals.
- Emerge and are shaped as a part of wider data ecosystems within health systems.

Note that while our work is driven by use cases taking place in cross-council research networks (UKRI), the principles and learnings from our project research activities concerning cross-domain access and re-use of sensitive data for research purposes have wider relevance beyond the DARE UK programme.

## Examining data usage patterns in operational health data networks

Increased attention is being given to how advanced analytics methods (AI/ML) are being used to discover value in big datasets. These methods are driving new data processing patterns and forms of research collaborations underpinned by the federated sharing and processing of data — e.g., facilitated via next-generation TREs. The relationship between one or more TREs and the health system is important as it influences applicable governance, data flows, tools and benefits expected by stakeholders who have an interest in the system under analysis — all of which have implications for privacy concerns, expectations and associated risks.

**Through our use case analysis, we highlight some key points about data usage patterns in health systems:**

- Health systems are complex and evolving networks of people and service providers whose purpose is to improve and support the health and wellbeing of society.
- Data value flows within such complex and evolving networks of people and service providers are driven by the demands of operational, clinical and research needs.



- Researchers studying health and social care systems will have a wide range of research questions depending on the phenomena they are seeking to understand, and the data value chains of which they are a part.
- A service provider (for healthcare, social care) typically only has partial information about individuals based on the systems that they operate or have access to (e.g., Electronic Health Records tend not to consider other data such as the wider social determinants of health); and therefore, possesses an incomplete view of the complex data network.
- The idea of partial views into complex data networks is important because it shows: (i) there is no centre to the network — in that, there is not a single point of control, or one place where all data can be viewed; (ii) data linkage is established by data controllers who are responsible for views into the network; (iii) views emerge within the network based on service and data value (e.g., a hospital, a curated disease specific dataset); and (iv) a TRE is a specific way of accessing a view on a network and therefore a view onto the whole dataset.
- The nature of data value changes within the complex data network, suggesting that value for research is distributed throughout the health system and cannot be easily integrated into a single place or TRE.

### We further outline some emerging data sharing needs:

- There is a need for greater availability and interoperability of quality data from service providers for research purposes.
- There is a need for TREs to be able to: manage ever-increasing variety, volumes, and velocity of data; offer greater support for a wider range of data analysis tools — e.g., for AI/ML; and be more connected with other TREs — e.g., to support advanced federated analysis and distributed machine learning.
- There is a need to consider how patients, service users, and members of public can have greater involvement with the co-design, testing, and evaluation of research concepts inception through to generated insights and tools (e.g., through interaction capabilities provided by TREs, such as interactive computational notebooks).

### Towards a shared understanding of the Five Safes dimensions for cross-council data sharing

The Five Safes approach is a well-recognised and valuable tool for risk communication, discussion and decision-making about access and re-use of sensitive data. We also highlight the special case of ‘Safe Return’ added by Hubbard et al. (2020). However, given the Five Safes is open to different interpretations, there is a need for both transparency and a shared understanding by all key stakeholders (e.g., data providers, data users, patients, TRE operators) about how the Five Safes (Plus One) are construed in the context of privacy risk assessment for cross-domain access and re-use of sensitive data for research purposes. Given the increasingly complex flows of multi-source data between people and organisations taking place through multiple platforms as part of wider data ecosystems of shared resources and services, federated data sharing and processing requires additional safeguards and controls, such as federated identity management, and specific types of privacy-enhancing technologies (PETs) — e.g., multi-party computation, differential privacy, homomorphic encryption. Therefore, to reflect the elements to be included in privacy risk assessment, we consider three requirements for federation that should be emphasised through the existing Five Safes dimensions:

- **Different types of research collaborations** should be highlighted as part of the **Safe Projects dimension**.
- **A wider-range of stakeholders** should be drawn attention to through the **Safe People dimension**.
- **The complexity of data flows** should be emphasised via the **Safe Settings dimension**.

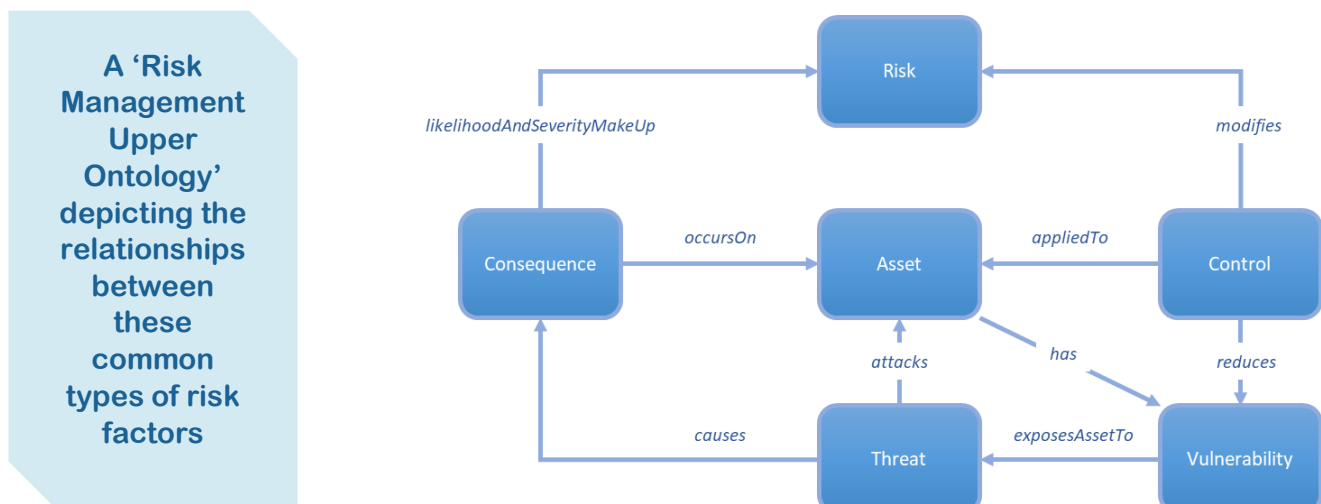
Focus on the Five Safes approach is resumed in the D2 report where specific privacy risk factors are identified and categorised utilising the Five Safes through our engagement with legal, ethics, regulatory and information governance experts and practitioners as part of the DARE UK PRiAM Advisory Board.

## Privacy risk modelling and simulation: mapping common types of risk factors

As part of a conceptual mapping exercise, we **identify common types of risk factors** used by the ISO/IEC 27005 methodology for information security risk management and other selected privacy risk assessment methodologies: **Commission nationale de l'informatique et des libertés (CNIL): Privacy Impact Assessment (PIA) Methodology and Knowledge Bases** (CNIL, 2018a; 2018b); **Inria — Research Centre Grenoble: Privacy Risk Analysis Methodology** (De & Le Métayer, 2016); the **UK Anonymisation Network (UKAN): Anonymisation Decision-Making Framework (ADF)** (Elliot et al., 2020); and, the **U.S. National Institute for Standards and Technology (NIST): NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management (NIST PRAM)** (NIST, 2020a). As part of this exercise, we identify several common types of risk factors used for modelling and simulating security and privacy risks, which include:

- **Assets** which can be ICT components, software, data, and socio-technical aspects such as, places and stakeholders specifically including data subjects;
- **Threats** that can affect Assets;
- **Consequences** of a Threat on an Asset, which is expressed as a **Risk** (i.e., the severity of the Consequence combined with its likelihood);
- Likelihood of Consequences is determined in part by **Vulnerabilities** of Assets, which may be reduced by **Controls** — such as (i) controls on data — i.e., those that transform the data itself, such as de-identification techniques; and (ii) environmental controls — i.e., those that change the environment in which the data is processed.

This conceptual mapping is utilised in the D3 report to model and simulate privacy risk for a given scenario using the System Security Modeller (SSM) (Phillips et al., 2022) — an existing tool used by the project which “automates much of an ISO 27005 risk assessment of socio-technical systems” modelling “both cyber-security and compliance” threats (Phillips et al., 2022).



We therefore further reflect on the relationship between information security and information privacy risk assessment methodologies and outline some key considerations:

- While concerns associated with information security and information privacy are conceptually related, each area offers a **distinct focus**.
- Information security risk assessment primarily focuses on risks arising from unauthorised activities — relating to loss of **confidentiality, availability and integrity**. Whereas privacy risk assessment focuses on risks in relation to **both unauthorised and authorised data-related activities**.
- **Privacy protection goals** from the field of privacy engineering (e.g., the Standard Data Protection Model) aim to address the ethical, legal, organisational and technical aspects of information privacy and data protection in relation to (un)authorised data processing activities in socio-technical systems (e.g., safe research collaborations). These goals therefore do not only concentrate on ensuring confidentiality, availability and integrity ('data security'), but also extend their focus to the goals of **data minimisation, unlinkability, transparency and intervenability** ('information privacy and data protection').
- While some **common methods** can be used to both increase security and protection of privacy (e.g., 'encryption'), some methods can also **cause tensions** in certain situations (e.g., 'identity verification').
- Typically, information security risk assessment centres on impacts to the operator or system stakeholders. In contrast, **information privacy risk assessment takes a much broader view, focusing on the impacts for individuals, groups of people and wider society** from potentially harmful activities.

## Table of Contents

1.	Introduction.....	13
1.1.	Purpose.....	13
1.2.	About the DARE UK PRiAM project .....	13
1.2.1.	Motivation .....	13
1.2.2.	Project objectives .....	15
1.2.3.	Project structure.....	15
1.2.4.	Engagement with the public and other stakeholders .....	16
1.3.	Scope of the D1 report .....	16
2.	Outlining three driver use cases related to public health research and integrated care.....	18
2.1.	Use Case A: Complex hospital discharge — “PROactive, Collaborative and Efficient complex Discharge” (PROCED) project.....	19
2.1.1.	NIHR ARC Wessex .....	19
2.1.2.	Motivation for the PROCED project.....	19
2.1.3.	Aim of the PROCED project.....	19
2.1.4.	Overview of data linkage and re-usage .....	19
2.1.5.	Project partners and research team.....	20
2.1.6.	Patient and public Involvement.....	20
2.2.	Use Case B: Multi-morbidity prevention — “Multidisciplinary Ecosystem to study Lifecourse Determinants and Prevention of Early-onset Burdensome Multimorbidity” MELD-B Project .....	20
2.2.1.	The NIHR AIM research programme.....	20
2.2.2.	Motivation for the MELD-B project .....	20
2.2.3.	Aim of the MELD-B project .....	21
2.2.4.	Overview of data linkage and re-usage .....	21
2.2.5.	Project partners and research team.....	21
2.2.6.	Patient and public Involvement.....	21
2.3.	Use Case C: NHSx Wessex Federated TREs.....	22
2.3.1.	Motivation for sub-national TREs.....	22
2.3.2.	Aim of the Wessex Federated TREs project.....	22
2.3.3.	Overview of data linkage and (re)usage .....	22
2.3.4.	Patient and Public Involvement.....	22
2.4.	Use cases: brief summary .....	22
2.4.1.	These use cases provide examples of multi-disciplinary research collaborations .....	22
2.4.2.	These use cases demonstrate how research projects related to public health research and integrated care require a considerable number of connected, multi-stakeholder data sources .....	23



2.4.3.	These use cases emerge and are shaped as part of wider data ecosystems .....	23
3.	Exploring emerging data usage patterns: Trusted Research Environments as facilitators of federated sharing and processing of data .....	25
3.1.	Operational context of health data networks .....	25
3.1.1.	Partial views into complex data networks.....	26
3.1.2.	Service integration and data aggregation .....	26
3.2.	Trusted Research Environments in healthcare systems.....	28
3.2.1.	A representative example of a research network .....	30
3.3.	Emerging data sharing needs .....	31
3.3.1.	‘An enhanced research experience’: supporting ‘advanced federated analysis’ and ‘distributed machine learning’ .....	31
3.3.1	‘Stakeholder involvement’: co-design and interaction with data .....	32
4.	Examining different approaches to identifying, organising and using risk factors for privacy risk assessment	33
4.1.	The Five Safes: “Safe Projects”, “Safe People”, “Safe Settings”, “Safe Data” and “Safe Outputs” .....	33
4.1.2.	The importance of risk communication.....	35
4.1.3.	Consideration of federation in relation to the Five Safes dimensions .....	36
4.2.	Comparing the ISO/IEC 27005 methodology for information security risk management with selected privacy risk assessment methodologies .....	38
4.2.1.	Sources.....	38
4.2.2.	Identifying common types of risk factors .....	41
4.2.3.	Determining the scope of the risk assessment.....	44
4.3.	Privacy protection goals .....	45
4.3.1.	Tensions between these Privacy Protection Goals.....	46
4.3.2.	Relationship between Privacy Protection Goals and Risk Management.....	47
4.4.	On the relationship between information security and information privacy risk assessment methodologies.....	47
4.4.1.	Consideration of impacts from (un)authorised processing.....	47
4.4.2.	Tensions between approaches .....	48
4.4.3.	Focus on impacts for a wider range of stakeholders.....	48
5.	Conclusion .....	48
5.1.	Outlining the context for privacy risk assessment: data usage patterns in operational health data networks.....	49
5.2.	Towards a shared understanding of the Five Safes dimensions for cross-council data sharing.....	50
5.3.	Privacy risk modelling and simulation: mapping common types of risk factors .....	50
6.	Acknowledgements .....	52
7.	References .....	53

8. Glossary .....	60
9. Appendix: summary of changes to v1.2 report .....	62

## List of Figures

Figure 1: An Overview of the DARE UK PRiAM Project: Deliverables, Stakeholder Engagement and Work Packages .....	13
Figure 2: A Simplified Health Network including Service Integration and Data Aggregation .....	27
Figure 3: Distribution of Data Value and Impact of Data Aggregation .....	28
Figure 4: Typical TRE Deployment Scenarios.....	29
Figure 5: A Representative Example of a Research Network .....	30
Figure 6: ISO 27005 Risk Assessment Identification Activities .....	40
Figure 7: Risk Management Upper Ontology .....	41
Figure 8: ISO 27005 Context Establishment .....	45

## List of Tables

Table 1: Overview of three real-world uses cases selected as exemplars of access and re-usage of cross-domain sensitive data.....	18
Table 2: Three Typical TRE Deployment Scenarios.....	29
Table 3: Original Five Safe Questions (Desai et al., 2016) .....	34
Table 4: Some example interpretations of the Five Safe principles.....	34
Table 5: Mapping Risk Management Concepts to Privacy Risk Assessment.....	41

## Abbreviations

<b>ACONF</b>	Aberdeen Children of the 1950s
<b>AEPD</b>	Agencia Española de Protección de Datos (Spain)
<b>AHSN</b>	Academic Health Science Network
<b>AI/ML</b>	Artificial Intelligence/Machine Learning
<b>AIHW</b>	Australian Institute of Health and Welfare
<b>AIM</b>	Artificial Intelligence for Multiple Long-Term Conditions
<b>BCS70</b>	1970 British Cohort Study
<b>CIA</b>	Confidentiality, Integrity and Availability
<b>CHIE</b>	Care and Health Information Exchange
<b>CIDPSAFL</b>	Conference of the Independent Data Protection Supervisory Authorities of the Federation and the Länder (Germany)
<b>CNIL</b>	Commission nationale de l'informatique et des libertés (France)
<b>COTADS</b>	COdesigning Trustworthy Autonomous Diabetes Systems
<b>CPRD</b>	Clinical Practice Research Datalink
<b>D.</b>	Deliverable
<b>DARE UK</b>	Data and Analytics Research Environments UK
<b>DARE UK PRIAM</b>	DARE UK Privacy Risk Assessment Methodology
<b>DiIS</b>	Dorset Intelligence & Insight Service
<b>DIKW</b>	Data, Information, Knowledge and Wisdom
<b>EHR</b>	Electronic Health Record
<b>EPSRC</b>	Engineering and Physical Sciences Research Council (UK)
<b>ESRC</b>	Economic and Social Research Council (UK)
<b>GDPR</b>	General Data Protection Regulation
<b>HDR UK</b>	Health Data Research UK
<b>HES</b>	Hospital Episode Statistics
<b>ICS</b>	Integrated Care System
<b>ICT</b>	Information Communication Technologies
<b>IEEE</b>	Institute of Electrical and Electronics Engineers
<b>IPC</b>	Information & Privacy Commissioner of Ontario, Canada

<b>ISO</b>	International Organization for Standardization
<b>MELD</b>	Multidisciplinary Ecosystem to study Lifecourse Determinants and Prevention of Early-onset Burdensome Multimorbidity
<b>MLTC-M</b>	Multiple Long-Term Condition Multimorbidity
<b>MRC</b>	Medical Research Council (UK)
<b>NHS</b>	National Health Service
<b>NICE</b>	National Institute for Health and Social Care Excellence
<b>NIHR</b>	National Institute for Health and Care Research (UK)
<b>NIST</b>	National Institute for Standards and Technology (USA)
<b>NIST PRAM</b>	NIST Privacy Risk Assessment Methodology
<b>OECD</b>	Organisation for Economic Co-operation and Development
<b>PETs</b>	Privacy-Enhancing Technologies
<b>PIA</b>	Privacy Impact Assessment
<b>PROCED</b>	PROactive, Collaborative and Efficient complex Discharge
<b>RFC</b>	Request for Comments
<b>SAIL Databank</b>	Secure Anonymised Information Linkage Databank
<b>SCC</b>	Southampton City Council
<b>SCR</b>	Shared Care Record
<b>SDF</b>	Social Data Foundation
<b>SDM</b>	Standard Data Protection Model
<b>SSM</b>	System Security Modeller
<b>TRE</b>	Trusted Research Environment
<b>UHS</b>	University Hospital Southampton
<b>UK</b>	United Kingdom
<b>UKRI</b>	UK Research and Innovation
<b>USA</b>	United States of America
<b>WCR</b>	Wessex Care Records
<b>WP</b>	Work Package
<b>WSI</b>	Web Science Institute



## 1. Introduction

### 1.1. Purpose

This report is Deliverable 1 (D1) “Privacy Risk Assessment Requirements for Safe Collaborative Research: Exploring Emerging Data Patterns and Needs of Advanced Analytics in Cross Council Research Networks through Use Case Analysis” 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.

### 1.2. 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.

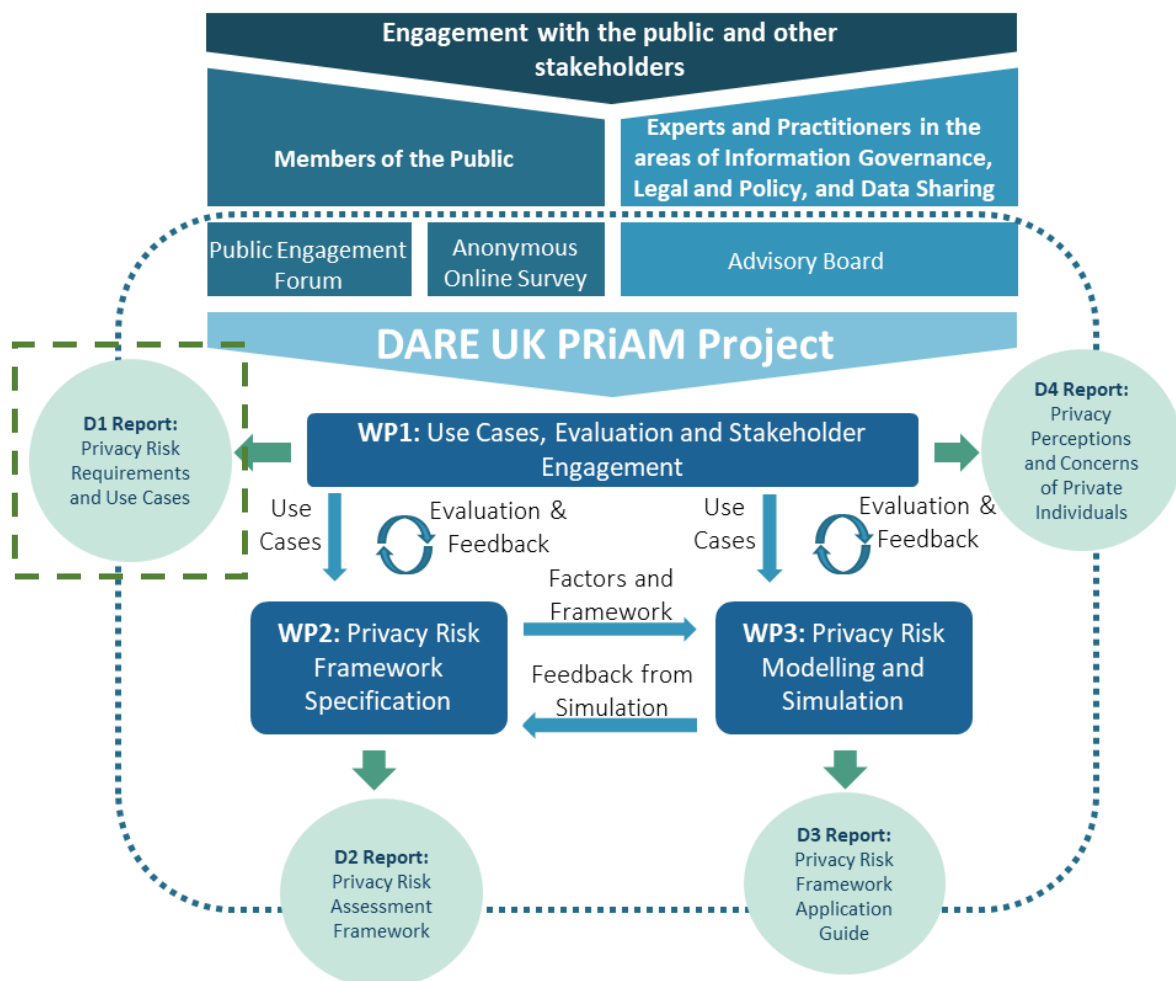


Figure 1: An Overview of the DARE UK PRiAM Project: Deliverables, Stakeholder Engagement and Work Packages

#### 1.2.1. 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<sup>1</sup> 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.<sup>2</sup>

## Some examples of potentially harmful activities and undue harm related to privacy

There are various types of potentially harmful activities that can give rise to different sorts of privacy harms. Such potentially harmful activities not only relate to “re-identification”, but also to other “problematic data actions”<sup>3</sup> such as, “appropriation”, “distortion”, “induced disclosure”, “lapses in data security”, “stigmatization”, “surveillance”, “unanticipated revelation” and “unwarranted restriction” (National Institute for Standards and Technology [NIST], 2019). Problematic data actions are also referred to as “feared events” — a term commonly used in cyber security risk assessment. For instance, the Commission nationale de l’informatique et des libertés (CNIL, 2018b) provides a “Typology of the outcomes of feared events” including: ““Illegitimate access to personal data” — (i) with no tangible outcome, (ii) via “storage”; (iii) through “redistribution” or (iv) by “use”; “Unwanted modification of personal data” — (i) via “malfunction” or (ii) “through use”; and “Disappearance of personal data” — (i) due to “malfunction” or (ii) which leads to “blockage”.

However, “there is no general agreement on how to categorise or rate privacy harms, i.e., on the outcomes one is trying to avoid” (OECD, 2019).<sup>4</sup> As examples of privacy harms, in their “Catalog of Problematic Data Actions and Problems”, the U.S National Standards Institute for Technology (NIST, 2019) set out five key problems for individuals: “dignity loss”; “discrimination”; “economic loss”; “loss of self-determination”, including “loss of autonomy”, “loss of liberty” and “physical harm”; and “loss of trust”. Further, Citron & Solove (2021) provide a typology of privacy harms: “physical harms”, “economic harms”, “reputational harms”, “psychological harms”, “autonomy harms”, “discrimination harms” and “relationship harms”. Also see Solove’s (2006) taxonomy of privacy that includes “four basic groups of harmful activities: (1) information collection, (2) information processing, (3) information dissemination, and (4) invasion.” The Information Commissioner’s Office (ICO, n.d.) also asks those carrying out a data protection impact assessment to consider to what extent processing may contribute to: “inability to exercise rights (including but not limited to privacy rights)”; “inability to access services or opportunities”; “loss of control over the use of personal data”; “discrimination”; “identity theft or fraud”; “financial loss”; “reputational damage”; “physical harm”; “loss of confidentiality”; “re-identification of pseudonymised data”; or “any other significant economic or social disadvantage”.<sup>5</sup>

It is also important to note that the non-use of data for research purposes due to risk aversion can also lead to potential harms (e.g., Laurie et al., 2014).

<sup>1</sup> 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).

<sup>2</sup> 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.

<sup>3</sup> A “problematic data action” is defined by NIST (n.d.) as “a data action that could cause an adverse effect for individuals”.

<sup>4</sup> Note that Calo (2011) defines privacy harm in two senses: the subjective “perception of unwanted observation” and the objective “unanticipated or coerced use of information concerning a person against that person”.

<sup>5</sup> Related to Recital 75 of the General Data Protection Regulation (GDPR).

Price & Cohen (2019) further distinguish between two key types of health privacy concerns — that is, “consequentialist concerns” and “deontological concerns”:

- “Consequentialist concerns” manifest when a “privacy violation” has “negative consequences” for an individual (or groups of people) such as: “discrimination based upon health data” — e.g., “employment discrimination”; “stigma” — e.g., “from others knowing about a sexually transmitted infection”; “embarrassment, paranoia, or mental pain” — e.g., “potential for increased anxiety” over a perceived increased susceptibility to “identity theft”; and, “dignitary harms” — e.g., “it is important that there be a part of an individual’s life that is his or hers alone, that remains unknown to others unless shared” (Price & Cohen, 2019).
- “Deontological concerns” manifest when a “privacy violation” wrongs an individual (or groups of individuals) without their knowledge or “even if no one uses a person’s information against this person” (Price & Cohen, 2019). Such deontological concerns, therefore, are not contingent on an individual (or groups of individuals) “experiencing negative consequences” (Price & Cohen, 2019).

What remains vital is that people are protected from harms associated with data disclosure and re-use — and that public confidence and engagement in health and social care research are maintained. As such, the DARE UK PRiAM project aims to explore methods and tools that can support decision-makers, patients and the public to assess and manage privacy risk when considering emerging data access and re-usage scenarios, such as federation.<sup>6</sup>

### 1.2.2. Project objectives

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 methodology 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).

### 1.2.3. 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.

<sup>6</sup> It is worthwhile to note that the importance of privacy preservation and privacy engineering has been recognised by the recently published “Goldacre Review” on ‘using health data for research and analysis’ commissioned by Secretary of State for Health and Social Care (Goldacre & Morley, 2022). For example, the following two recommendations were made by the review related to this point: “UKRI/NIHR should resource applied methods research into privacy preservation”; and “TRE 9. Evaluate new developments in privacy engineering; adapt accordingly” (Goldacre & Morley, 2022).

## 1.2.4. 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”.

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”.

## 1.3. Scope of the D1 report

**Work Package 1 (WP1): Use Cases, Evaluation & Stakeholder Engagement.** This Deliverable 1 (D.1) report focuses on key research activities undertaken as part of WP1 — which are: (i) privacy risk requirements analysis in relation to research collaborations (including, federated research networks),<sup>7</sup> which describes stakeholders, usage contexts and research purposes; and (ii) specification of the three use cases: ‘complex hospital discharge’, ‘public health multi-morbidity prevention’ and ‘integrated care system’. This D.1 report specifically concentrates on the following project objective:

“  
**Project objective 1 of 5:** Analyse driver use cases (public health prevention, integrated care) and data usage patterns from health and social care research typical of future MRC and ESRC data sharing (WP1, Outcome: understanding of future unmet data sharing needs)  
 ”

**Analyse driver use cases.** Our work is driven by cross UKRI research council use cases (for further information see Section 2) focused on around health (Medical Research Council [MRC]), social science and social care (Economic and Social Research Council [ESRC]), and computer science (Engineering and Physical Sciences Research Council [EPSRC]). The use case approach is championed by the DARE UK programme and ensures that solutions are aligned with funding bodies and research priorities of UKRI, and the needs of researchers themselves. Yet, while this report focuses on particular data sharing and (re)usage situations in the context of the DARE UK programme, the key points raised are more widely applicable to cross-domain access and re-use of sensitive data for research purposes.

**Approaches to identify key factors contributing to privacy risks.** This report further provides an initial step towards addressing the second project objective — a key focus of the D2 report: “Privacy Risk Assessment Framework” — by exploring some different approaches to identifying, organising and using risk factors for privacy risk assessment through a literature review, including the Five Safes:

<sup>7</sup> For definitions of the terms ‘research collaboration’ and ‘federated research network’ see Section 2.4 and the glossary in Section 8.



**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).



This D1 report is divided into the following sections:

- **Section 2. Outlining three driver use cases related to public health research and integrated care.** Three real-world uses cases are provided as exemplars of access to and re-usage of cross-domain sensitive data for research purposes.
- **Section 3. Exploring emerging data usage patterns: Trusted Research Environments as facilitators of federated sharing and processing of data.** The role of TREs within the operational context of health data networks is examined, including a representative example of a federated research network as part of a federated data usage scenario. Some emerging data sharing needs are also considered to better understand how privacy concerns, expectations and associated risks may develop and change as research collaborations become more federated.
- **Section 4. Examining different approaches to identifying, organising and using risk factors for privacy risk assessment.** We specifically focus on the Five Safes (Desai et al., 2016) as a well-known methodology used to provide best practice principles for safe research. We then compare the ISO/IEC 27005 methodology for information security risk management with selected privacy risk assessment methodologies. Finally, we provide an overview of privacy protection goals from the field of privacy engineering.
- **Section 5. Conclusion.** The report then concludes by summarising key points from the previous sections, and highlights further work in subsequent project reports.

See Section 8 of this report for a Glossary of key terms.



## 2. Outlining three driver use cases related to public health research and integrated care

Our work on privacy risk assessment, as part of the DARE UK PRIAM project, is driven by three cross-council use cases revolving around health (MRC), social science and social care (ESRC), and computer science (EPSRC). We specifically selected three real-world advanced analytics use cases related to public health research and integrated care to be used as exemplars of data linkage of cross-domain sensitive data to help drive the identification of factors and situations causing and affecting privacy risks.

A brief overview of these three use cases is provided in the following table — a more detailed outline is given in sections 2.1-2.3:

Table 1: Overview of three real-world uses cases selected as exemplars of access and re-usage of cross-domain sensitive data

	Use Case A: Complex Hospital Discharge	Use Case B: Multi-morbidity Prevention	Use Case C: Sub-national federated trusted research environment ecosystem
<b>Project title:</b>	“PROactive, Collaborative and Efficient complex Discharge” PROCED Project	“Multidisciplinary Ecosystem to study Lifecourse Determinants and Prevention of Early-onset Burdensome Multimorbidity” MELD-B Project	“NHSx Wessex Federated TREs” Project
<b>Project aim:</b>	Focuses on a workflow supporting the proactive assessment of patients’ onward care needs and discharge risks necessary to minimise discharge delays and improve recovery; and optimising the planning and allocation of community resources against predicted hospital demand.	Centres on using Artificial Intelligence (AI) applied to health data, to identify primary health conditions that may progress to burdensome multiple long term conditions if appropriate interventions are not made. The project seeks to identify causal conditions from lifecourse data that may be used to build an intervention strategy.	Purpose is to pilot a Wessex-wide federated TRE ecosystem that brings together the population of Dorset, Hampshire and Isle of Wight Integrated Care System (across two integrated care systems [ICSs]) and the reach of the region’s main tertiary referral centre, University Hospital Southampton, and affiliated NHS organisations.
<b>Data linkage:</b>	Involves individual linking of complex multi-stakeholder datasets (e.g., acute care, community care, local authority, etc.) regarding medical stability, patient and family capacity, ongoing care environment, and association with system capacity/demand.	Requires both individual linking and federated learning between longitudinal birth cohorts and routine data sets over the lifecourse.	Necessitates data linkage across clinical care, social care, mental health, and other public administration services.

‘Use Cases A and B’ are based on real-example scientific problems from the Social Data Foundation (SDF) (Boniface et al., 2020; Boniface et al., 2022). The SDF is a partnership between Southampton City Council (SCC), University Hospital Southampton (UHS) and the University of Southampton (UoS) to transform health and social care through accelerated and trustworthy access to federated linked datasets. The research questions posed by

each use case highlights important patterns of data linkage and usage for federated TREs for advanced analytics (artificial intelligence/machine learning [AI/ML]). Further, the addition of 'Use Case C' (which also involves the SDF supporting development of federation models) not only allows us to examine these data usage patterns at research project-level, but also for one or more programmes of research at sub-national level.

## **2.1. Use Case A: Complex hospital discharge — “PROactive, Collaborative and Efficient complex Discharge” (PROCED) project**

### **2.1.1. NIHR ARC Wessex**

The National Institute for Health and Care Research (NIHR) Applied Research Collaboration (ARC) Wessex (“the NIHR ARC Wessex”) is “one of fifteen NIHR ARCs” across England that aim to “support applied health and care research that responds to, and meets, the needs of local populations and local health and care systems” (NIHR ARC Wessex, n.d.). NIHR ARCs bring together “local providers of NHS services, local providers of care services, NHS commissioners, local authorities, universities, private companies and charities” (NIHR ARC Wessex, n.d.). The NIHR ARC Wessex focuses on four core research areas: “ageing and dementia”, “healthy communities”, “long term conditions” and “workforce & health systems” (NIHR ARC Wessex, n.d.). The PROCED project is one of several projects under the “workforce & health systems” research area — PROCED stands for “PROactive, Collaborative and Efficient complex Discharge”.

### **2.1.2. Motivation for the PROCED project**

In some cases, the results of “discharge assessment” will show that a patient requires “little or no care” after leaving hospital — described as “a minimal discharge (NHS, 2019). In other cases, the results of a discharge assessment will find that a patient needs “more specialised care after leaving hospital — referred to as “a complex discharge” (NHS, 2019). In 2016 and 2017, hospital discharges in the UK were delayed by 2.3 million days (Gardner, 2022). Three-quarters of these delays occurred due to arrangements for community care plans that require many non-acute care services, residential homes, nursing homes, care packages, community equipment, and public funding.

The University Hospital Southampton (UHS) NHS Foundation Trust offers a sophisticated discharge system that enables diverse teams responsible for planning onward care to collaborate in discharge planning decisions. This system allows patients, families, caregivers and services providers to share information about care needs and available resources during discharge choices, allowing operational teams to track, prioritise, plan and provision services to address care needs. However, due to poor information linkage between health and social care, it is difficult to identify patients at risk of discharge delay or readmission, and consequent future demand on community care services. Community service planning and allocations therefore tend to be reactive rather than proactive, requiring human intensive activities to manage cohorts of patients who are medically fit for discharge but who do not have a safe community destination for further assessment, rehabilitation and recovery.

### **2.1.3. Aim of the PROCED project**

By using machine learning, the project aims to develop advanced algorithms that can proactively model patient discharge risks (e.g., extended length of stay, readmission) and expected departure points. Thus, patient scheduling can be optimised by using predictions made by the algorithm to find the best use of resources from available options.

### **2.1.4. Overview of data linkage and re-usage**

The project will include individual linking of complex multi-stakeholder data using the following datasets:

- **UHS Electronic Clinical and Management:** Hospital episodes, Discharge events (messages, pathway state), Discharge report.
- **Social care records from Southampton City Council (SCC):** Social care record (personal, daily, therapies, social/community engagement), Residence (Domiciliary, supported accommodation and nursing home).
- **Community Nursing (Southern Health Care Record):** Visits, Duration, Location Tasks, dependencies, Workforce skills and teams.

Data linkage is coordinated by data providers responsible for establishing a multi-stakeholder pseudonymised dataset in accordance with the requirements of a research protocol specification. The goal is to curate a dataset that can be used to study complex discharge within PROCED, but also reused in subsequent projects in a programme of ongoing work over the coming decade. Even within this localised setting there are multiple TRES operated by each data provider organisation. Due to the close working relationship between stakeholders involved in the project the dataflow and responsibilities (sponsorship, linking, safe provisioning, etc.) are all negotiated.

### 2.1.5. Project partners and research team

The three PROCED project partners are University of Southampton, Southern Health NHS Foundation Trust and University Hospital Southampton NHS Foundation Trust. The research team are from multiple disciplines, including the areas of digital health research and operational research working with healthcare professionals.

### 2.1.6. Patient and public involvement

The research concept for the project has been developed with patients and the public — who will also be involved with ‘co-design, testing and evaluation’. The public will further participate as part of a ‘Steering Committee’ and a ‘Public Patient Involvement (PPI) Committee’ will run eight workshops.

For further information about the PROCED project see: <https://www.arc-wx.nihr.ac.uk/research-areas/workforce-and-health-systems/proced-proactive-collaborative-and-efficient-complex-discharge/>.

## 2.2. Use Case B: Multi-morbidity prevention — “Multidisciplinary Ecosystem to study Lifecourse Determinants and Prevention of Early-onset Burdensome Multimorbidity” MELD-B Project

### 2.2.1. The NIHR AIM research programme

The NIHR Artificial Intelligence for Multiple Long-Term Conditions (“NIHR AIM”) research programme “aims bring together multi-disciplinary Research Collaborations to build on understanding of disease clusters in people with multiple long term conditions (MLTCs) using ground-breaking AI techniques; and to grow capability for multi-disciplinary working in this crucial research area” (NIHR, 2022). The programme has funded seven large scale collaborative projects tackling challenges of understanding disease clusters/trajectory in relation to clinical, social, genetics, polypharmacy and burdensomeness factors. A further Research Support Facility project has been funded to support collaboration across the programme that includes five themes: “Reproducible, secure and interoperable infrastructure”, “Accessible, research ready data”, “Community building and training”, “Patient and public involvement and engagement”, “Sustainability and legacy” (The Alan Turing Institute, 2021).

### 2.2.2. Motivation for the MELD-B project

The number of people living with two or more long-term conditions (multiple long-term condition MLTCs) is increasing. The MELD-B project “will help understand when MLTCs becomes ‘burdensome’ and the best opportunities for intervention”, such as by exploring social determinants of health data (Medicine — University of

Southampton, 2022). The MELD-B project is funded by the National Institute for Health and Care Research (NIHR). It commenced in April 2022 and will complete in September 2024.<sup>8</sup>

### 2.2.3. Aim of the MELD-B project

The aim of the MELD-B project is to “use an Artificial Intelligence (AI) enhanced analysis of birth cohort data and electronic health records to identify lifecourse time points and targets for the prevention of early-onset, burdensome MLTCs”. One area of focus for the project includes developing “safe data environments and readiness for AI analyses across large, representative routine healthcare datasets and birth cohorts” (Medicine — University of Southampton, 2022). These data will then be used by advanced algorithms for clustering populations, modelling trajectories over the lifecourse and identifying optimal timepoints for preventative interventions.

### 2.2.4. Overview of data linkage and re-usage

The project will require both individual linking and federated learning between longitudinal birth cohort and routine data sets over the lifecourse — focusing on datasets from the following sources:

- Secure Anonymised Information Linkage (SAIL)
- Clinical Practice Research Datalink (CPRD)
- Hospital Episode Statistics (HES)
- 1970 British Cohort Study (BCS70)
- Aberdeen Children of the 1950s (ACONF)<sup>9</sup>

MELD-B is positioned within the NIHR AIM programme and there is some overlap between the datasets used in MELD-B and other projects (e.g., CPRD and SAIL), although no project is using identical datasets. The projects in the programme are working together with data providers to ensure that results can be shared. In addition, datasets for MELD-B are accessed through four different TREs each providing data linkage in accordance with the data access application. This adds significant complexity to access and restricts analysis between TRE datasets to analysis of safe outputs only.

### 2.2.5. Project partners and research team

The MELD-B project involves King’s College London, Southampton City Council, Swansea University, University of Aberdeen, University of Glasgow, University Hospitals Southampton NHS Foundation Trust and University of Southampton. The research team are from multiple disciplines, including public health research, digital health science, primary care and mathematics (Medicine — University of Southampton, 2022).

### 2.2.6. Patient and public involvement

The project has a Patient and Public Involvement Officer (Medicine—University of Southampton, 2022).

**For further information about the MELD-B project see:**

[https://www.southampton.ac.uk/medicine/academic\\_units/projects/meld-b.page](https://www.southampton.ac.uk/medicine/academic_units/projects/meld-b.page).

<sup>8</sup> MELD-B follows on from the first phase of MELD — for further background information see: Boniface et al. (2022) where the initial phase of MELD (before MELD-B) is used as a validation test case for the Social Data Foundation (SDF) model.

<sup>9</sup> For more information about these sources see: SAIL (n.d.); CPRD (n.d.); NHS Digital (2022) — with regard to HES; Centre for Longitudinal Studies, UCL (n.d.) — with regard to BCS70; and University of Aberdeen (n.d.) — with regard to ACONF.

## 2.3. Use Case C: NHSx Wessex Federated TREs

### 2.3.1. Motivation for sub-national TREs

The NHS is focused on enriching “data-driven research and innovation” — one key area of interest being the planning and testing of “NHS-owned, ‘Subnational Trusted Research Environments’” which would allow “researchers to conduct de-identified data analysis at a significant ‘regional’ scale, whilst being able to work closely with local clinical teams who provide critical expertise and context” (Jhutti & Bloomfield, 2022).

The Wessex Federated TREs project is “one of four geographies in the country” selected “to pilot federating data at scale across more than one Integrated Care System” (Wessex Academic Health Science Network [AHSN], 2022). The University Hospital Southampton NHS Foundation Trust, on behalf of Wessex Health Partners, led the successful bid (Wessex AHSN, 2022). Note: the Dorset Integrated Care System (ICS) and the Hampshire and Isle of Wight ICS are “founding partners for Wessex Health Partners” (Wessex AHSN, 2022). In the words of the Wessex Academic Health Science Network (Wessex AHSN, 2022): “Beyond the immediate project deliverables, a successful TRE could support more rapid adoption of innovation, particularly disruptive innovations where benefits may sit outside the conventional process of care, by making it much easier to undertake real-world evaluations of system benefits.”

### 2.3.2. Aim of the Wessex Federated TREs project

The aim therefore is to pilot a sub-national federated TRE ecosystem for Wessex that brings together the population of Dorset, Hampshire and Isle of Wight (ICSs) and the reach of the region's main tertiary referral centre, University Hospital Southampton, and affiliated NHS organisations. For instance, one of the use cases for the pilot, provided by the Wessex Academic Health Science Network (AHSN), focuses on “federating data to provide a real world evaluation of the system impact of adopting FeNO, a diagnostic tool that can be used in primary care for the more accurate diagnosis and treatment of asthma” (Wessex AHSN, 2022). The project further aims to test model architecture of federated data sharing and governance based on the Social Data Foundation (SDF) (Boniface et al., 2020; 2022).

### 2.3.3. Overview of data linkage and (re)usage

This use case necessitates data linkage across clinical care, social care, mental health, and related services, including Wessex Care Records (WCR) and a dataset from Dorset Intelligence and Insight Service (DiiS, 2022). Real-world data from two ICSs (Dorset and Hampshire IoW) and specialist hospitals (UHS) are utilised to accomplish this.

### 2.3.4. Patient and Public Involvement

The Wessex Academic Health Science Network is “supporting the design of patient and citizen engagement in the Wessex TRE” (Wessex AHSN, 2022).

## 2.4. Use cases: brief summary

### 2.4.1. These use cases provide examples of multi-disciplinary research collaborations

Research collaborations can be described as communities of people and organisations, often across different sectors and disciplines, working together for one or more shared goals, who contribute to research activities by undertaking or otherwise informing them. They may be *ad hoc*, short-lived collaborations — such as, for specific



research projects, or long-term formal resources — such as, those provided by professional bodies.<sup>10</sup> Research collaborations can take many forms — all varying in nature, size, and scale (e.g., sub-national, national, international). One example being “patient-powered research networks”; e.g., where “patients, their families and caregivers” can “generate and contribute data about themselves” and “collaborate with researchers in prioritizing and answering clinical research questions about the effectiveness of treatments” (Fleurence et al., 2014).

#### 2.4.2. These use cases demonstrate how research projects related to public health research and integrated care require a considerable number of connected, multi-stakeholder data sources

The research collaborations, represented by the use cases, all involve a degree of federation in terms of the resources and services that are required to achieve their shared goals.<sup>11</sup> For instance, all three use cases depend on the availability and linkage of quality data (‘shared resources and services’) for research purposes from a number of connected, multi-stakeholder data sources (e.g., healthcare providers, social care providers) at network-level. Further, the NHSx Wessex Federated TREs project (Use Case C) aims to pilot a federated research network at sub-national level through connected TREs (‘shared resources and services’) to provide enhanced support for research projects and innovation activities strongly associated with local clinical teams. Note that Harris et al. (2021) define a ‘federated research network’<sup>12</sup> as

*“collaborations among partners who, through coordination at an overarching network level, bring together, share, and optimize resources and services in order to enable research that exploits this new data-intensive and connected scientific environment” (Harris et al., 2021)<sup>13</sup>*

#### 2.4.3. These use cases emerge and are shaped as part of wider data ecosystems

An important aspect of privacy risk assessment is understanding how privacy risks occur in relation to the ‘interaction between people, resources and services’ as part of research networks — and “the (soft and hard) structures that shape that interaction (such as national policies on data sharing and access, the legal framework,

<sup>10</sup> For further illustration, in some cases, a key purpose of a research collaboration can be to establish and participate in a ‘data sharing initiative’ i.e., where two or more organisations come together to share data for “an agreed purpose” (Ada Lovelace Institute & AI Council, 2021). These shared goals “will subsequently determine the benefits and drive the **nature of the relationship** between the actors involved in a data-sharing initiative” (Ada Lovelace Institute & AI Council, 2021 [bold emphasis as part of original text]). Data sharing initiatives come in many different forms — e.g., “data commons”, “data exchanges and markets”, “data trusts”, “open data platforms and open APIs”, “data collaboratives”, “data co-ops” and “research partnerships and data hackathons” (Smart Dubai & Nesta, 2020).

<sup>11</sup> Note this may involve different forms of governance arrangements, technical infrastructure, and legal structure (e.g., “multi-party contracts or corporate structures” [Stalla-Bourdillon et al., 2019b]). As an example, multiple organisations agree to share certain resources and services for the purposes and duration of a specific project under a multi-party contractual arrangement. Another example is where multiple organisations decide to collectively govern shared resources and services, such as those provided by a group of federated TREs, through a data institution to support various programmes of projects — in some cases, this data institution could be an independent legal entity.

<sup>12</sup> Also note similar definitions of federation: e.g., Nokkala & Dahlberg (2019) provide a definition of a ‘federative approach’, in the context of health and social care transformation, as follows: “[b]y federative approach, we mean governance, methods and practices that make data interoperable through the shared attributes (=data elements) of information systems (IS) and/or data storages. Interoperability implies that data are linked and made available from their original data sources by using shared attributes.” Note the term federation is defined by NIST (Singhal et al., 2007) based on Bajaj et al. (2003) as “A collection of realms (domains) that have established trust among themselves. The level of trust may vary but typically includes authentication and may include authorization”. In some cases, such an approach can be both federated and distributed, such as the World Economic Forum (WEF, 2020): ‘Federated Data Consortium Model’ and the Etic Lab and Open Data Institute: ‘Data Federations Model’ (Woodall, 2021) — also see: Eder & Shekhovtsov (2021) for discussion concerning “federated medical data lakes”. For instance, the Information Commissioner’s Office (ICO, 2021) describes a “distributed and federated approach” as where “data is not held centrally but is distributed amongst various controllers in the system. This allows interoperability, and only essential and minimised information sharing between de-centrally organised controllers, providing increased control to individuals and increased security to their data.” Further, Chaterji et al. (2019) explain the concept of “federation” in the context of “distributed cyberinfrastructures” within genomics, as allowing “the end users to transparently access a set of resources and services, distributed among several independent service providers”. Also, see: Peeters (2013) for discussion about how the label ‘federated’ has been applied to network architecture.

<sup>13</sup> This definition aligns with the rationale for ‘federated TRE ecosystems’ given by UK HDRA & NHSx (2021) — that is “[t]o maximise the potential of using TREs, common agreed specifications and systems are needed to simplify processes for researchers, lowering barriers to access multiple TREs and supporting federated analysis.” Further, it is important to highlight that HDR UK (2021b) are exploring how such as federation may operate in practice e.g., by setting out an “open, federated and interoperable technology stack for trusted research environments” and “capability maturity model” for a “federated data analytics infrastructure”.

IT systems, governance practices, cultural attitudes to data sharing and privacy, etc.)” (Elliot et al., 2020).<sup>14</sup> It is therefore important to consider how these research collaborations (e.g., projects, programmes of work, long-term formal resources provided by professional bodies) emerge and are shaped (including the extent of their federation) as part of wider ‘data ecosystems’<sup>15</sup> within health systems; a point further discussed in Section 3 of this report. These data ecosystems are where data can be collected/generated, made accessible and linked, such as, for primary usage — e.g., by health and social care providers for direct care (such as, for medical diagnoses, treatments); and for secondary usage — e.g., by researchers for indirect care (such as, for assessing the effectiveness of health and social care provision and policies).<sup>16</sup>

Individual TREs and federated ecosystems of TREs must have robust socio-technical data governance models in place to facilitate trustworthy data sharing and re-usage, including clearly-defined data governance objectives, roles, rules, processes and workflows combined with robust risk assessment and mitigation procedures (e.g., for data ingress, data egress, algorithm portability, monitoring re-usage activities etc.).<sup>17</sup> The aim of the project therefore is to inform the development of such socio-technical data governance models by focusing on the risk factors associated with cross-domain access and re-usage of sensitive data for research utilising advanced analytics methods (AI/ML).

---

<sup>14</sup> Note this is referred to as the “Comprehensiveness Principle: *You cannot decide whether or not data are safe to share/release by looking at the data alone, but you still need to look at the data*” (Elliot et al., 2020) by the UK Anonymisation Network (UKAN): Anonymisation Decision-Making Framework (ADF).

<sup>15</sup> The term ‘data ecosystem’ is described by Oliveira et al. (2019) as “socio-technical complex networks in which actors interact and collaborate with each other to find, archive, publish, consume, or reuse data as well as to foster innovation, create value, and support new businesses.”

<sup>16</sup> For further discussion about data sharing in relation to ‘direct’ and ‘indirect care’ see: Information Governance Review (2013).

<sup>17</sup> For further discussion of key components of data governance models for trustworthy data sharing, see Stalla-Bourdillon et al. (2019b).

## 3. Exploring emerging data usage patterns: Trusted Research Environments as facilitators of federated sharing and processing of data

Increased attention is being given to how cross-domain access and re-usage of sensitive data for research utilising advanced analytics methods (AI/ML) can be most appropriately supported, governed and provisioned. One proposed solution is the establishment of federated ecosystems of TREs,<sup>18</sup> fostering “next-generation TRE capabilities” (Kavianpour et al., forthcoming/ in press) required for advanced analytics methods, to act as facilitators of trustworthy, collaborative sharing and processing of data.<sup>19</sup> As part of the DARE UK PRiAM project, we therefore seek to explore emerging data usage patterns in relation to TRE ecosystems as facilitators of federated sharing and processing of data as part of cross-council research networks — and the privacy risk factors arising in this context.

From the analysis of the three driver use cases, it is clear that data usage patterns for research collaborations related to a TRE, or otherwise federation of TREs, should be considered in the context of the system they are established to study.<sup>20</sup> For instance, as shown by ‘Use Case C’, a key driving factor for sub-national TREs is to enable researchers “to work closely with local clinical teams” (Jhutti & Bloomfield, 2022). The relationship between one or more TREs and the health system is important as it influences applicable governance, data flows, tools and benefits expected by stakeholders who have an interest in the system under analysis — all of which have implications for privacy concerns, expectations and associated risks.<sup>21</sup>

In Section 3, we therefore explore the operational context of health data networks, including the role of TREs. We further provide a representative example of a federated research network as part of a federated data usage scenario. We then focus on emerging data sharing needs to better understand how privacy concerns, expectations and associated risks may develop and change as research collaborations become more federated.

### 3.1. Operational context of health data networks

Health systems are complex and evolving networks of people and service providers whose purpose is to improve and support the health and wellbeing of society. The World Health Organization (WHO, 2007) defines a ‘health system’ as follows

*“A health system consists of all organizations, people and actions whose primary intent is to promote, restore or maintain health. This includes efforts to influence determinants of health as well as more direct health-improving activities. A health system is therefore more than the pyramid of publicly owned facilities that deliver personal health services.” (WHO, 2007)*

<sup>18</sup> E.g., at regional (e.g., Use Case C: NHSx Wessex Federated TREs) and national levels (Hubbard et al., 2020).

<sup>19</sup> Examples of these next-generation capabilities being: “Support for big, non-structured data [...] Ability to parallelise computational jobs to either a High-Performance Computing (HPC) cluster or GPU farm [...] Support for software development within the TRE [...] Freedom to install software packages of researcher choice [...] Ability to export software and artificial intelligence (AI) algorithms from the environment [...] Ability to connect to certain internet locations, e.g. code repositories (GitHub)” (Kavianpour et al., forthcoming/in press)

<sup>20</sup> For instance, Bourne et al. (2015) state: “Current practices typically use many disparate sources of data to conduct a study. These data are located in a variety of repositories, often with different modes of access. This lack of centralization and commonality may hinder their ease of use and reduce productivity. We need a better understanding of usage patterns across multiple data resources to use as a basis for redesigning such resources to preserve valuable expertise and curation, and for improving how the data are found, accessed, integrated and reused.”

<sup>21</sup> For instance, privacy risks at network-level can be viewed as being “primarily operational” in nature (Information & Privacy Commissioner of Ontario, Canada [IPC], 2010). The IPC (2010) identifies four key types of operational risks that have “a chance of causing direct or indirect loss” for individuals, groups of people and wider society — these are: (i) “inadequate or failed internal processes and systems”; (ii) “issues related to staff”; (iii) “external events”; and (iv) “outsourced service providers”.

Data value flows within such complex and evolving networks of people and service providers are driven by the demands of operational, clinical and research needs. Researchers studying health and social care systems will have a wide range of research questions depending on the phenomena they are seeking to understand, and the ‘data value chains’<sup>22</sup> of which they are a part. For instance, what is needed in terms of data, tools and interdisciplinary expertise would be distinctly different e.g., for studying public health prevention, real-time decision support tools for emergency care pathways, and biomarkers for cancer detection.

### 3.1.1. Partial views into complex data networks

The people in health systems include individuals and communities along with the service provider workforce. These service providers encompass those provided by the NHS and a range of commercial companies offering point of care and wellbeing services all supported by industries across life sciences, MedTech, and information and communication technologies (ICT).

Over the life course, people’s health and wellbeing will vary. At times, when a person is sufficiently ill, they will seek support from healthcare services, at which point that person becomes patient. This highlights that being a patient is only part of a person’s life and that systems developed to support patients such as Electronic Healthcare Records (EHRs) tend not to consider other data such as the wider social determinants of health — i.e., data about individuals outside of the care setting. As such, a service provider typically only has partial information about individuals based on the systems that they operate or have access to.

**The idea of partial views into complex data networks is important because it shows:**

- There is no centre to the network — in that, there is not a single point of control, or one place where all data can be viewed.
- Data linkage is established by data controllers<sup>23</sup> who are responsible for views into the network.
- Views emerge within the network based on service and data value (e.g., a hospital, a curated disease specific dataset).
- A TRE is a specific way of accessing a view on a network and therefore a view onto the whole dataset.

### 3.1.2. Service integration and data aggregation

Data sharing within health systems is supported by two well-known mechanisms that implement the principles of distributed system architecture:

- **Service Integration:** connectivity between services to create business processes and care pathways (e.g., referral and discharge).
- **Data Aggregation:** sharing of data between service providers that is then used as a resource to deliver services (e.g., shared care records, public health management).

<sup>22</sup> Note that, in the context of big data, Curry (2016) describes ‘data value chains’ as “the information flow within a big data system as a series of steps needed to generate value and useful insights from data”. These data value chains include “key high-level activities” such as “data acquisition”, “data analysis”, “data curation”, “data storage” and “data usage” (Curry, 2016).

<sup>23</sup> The term ‘controller’ is defined by Article 4(7) of the GDPR as “the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data [...]”.

Figure 2 (below) shows a simplified health network including service integration (dotted lines) and data aggregation (black lines). Services are offered through Service Provider Components where data is created through service user and workforce interactions. Data can flow through service integration using service requests and service response documents (dotted lines) or via care records at the level of Service Provider, Regional or National (black lines). Service Provider Components may have access to higher level Care Records at regional and national levels (e.g., Hampshire and Isle of Wight Care and Health Information Exchange [CHIE] or Dorset ICS Diis), but this is not the case for all data aggregation (e.g., access to operational data provided to NHS Business Services Authority).

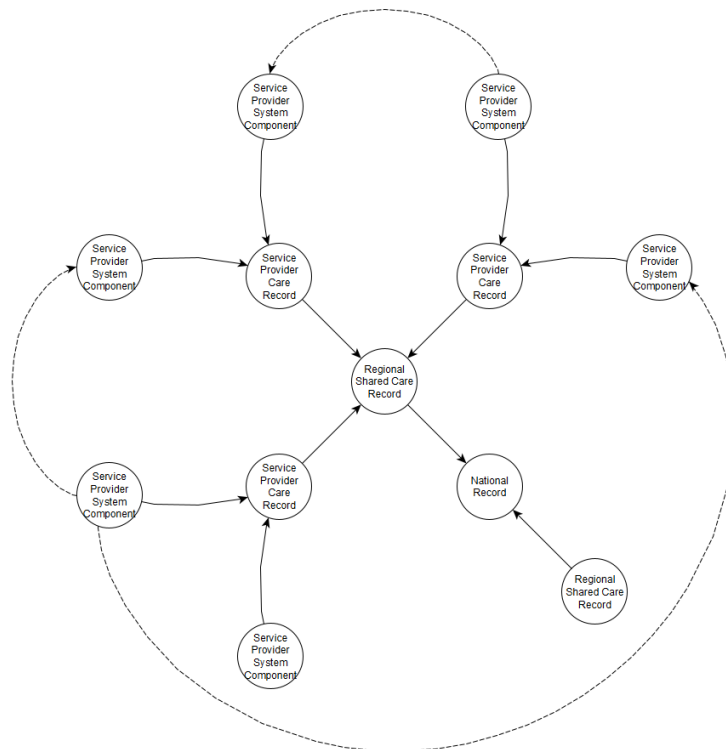


Figure 2: A Simplified Health Network including Service Integration and Data Aggregation

It is worth noting that health networks have evolved from the data structure of an Electronic Healthcare Record (EHR). An EHR is a data aggregation approach that allows different system components to access, update and share a patient record. Shared Care Records (SCRs) use the same data aggregation approach but at a higher level, combining care records from multiple providers (e.g., NHS Foundation Trusts, Local Authorities, etc.). What emerges is a hierarchy of data aggregation from acquisition at the network edge to population level at the network centre.

The process of data aggregation through EHRs and SCRs is often therefore ‘lossy’ — i.e., data and information are lost due to factors such as generalisation, size of datasets, privacy preservation and time. Data value therefore changes through data aggregation as resolution, context (metadata), and specialism is replaced by increasing numbers of data subjects available. Data aggregation also takes time — so if timely availability of information is necessary, it may not always be possible to share data through data aggregation processes, and in which case point-to-point exchange would be needed.

The changing nature of data value within the network is important because it suggests that value for research is distributed throughout the health system and cannot be easily integrated into a single place or TRE.

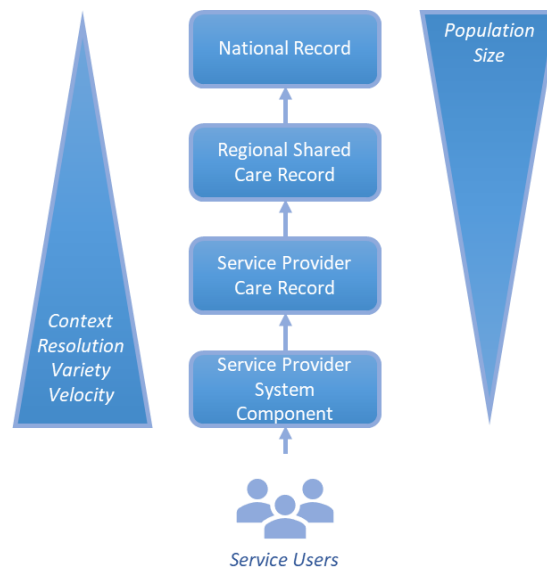


Figure 3: Distribution of Data Value and Impact of Data Aggregation

### 3.2. Trusted Research Environments in healthcare systems

Trusted Research Environments (TREs) have emerged to provide safe access to sensitive data for analysis. Real-world data is acquired from operational services, then integrated and de-identified/anonymised, before it is ingested into a TRE. The data flows for operational services and TREs are therefore different. Note the term ‘data flow’ is defined by the UKAN Anonymisation Decision-Making Framework (UKAN ADF) as

*“The movement or transfer of data through a system, describing who has responsibility for and access to them, and the contexts in which it is held.” (Elliot et al., 2020)*

The ability to understand the multiple ways in which data flows to different stakeholders and environments across its lifecycle — that is, from collection, aggregation, analyses through to insight dissemination and (where applicable) deletion — is essential for helping to assess privacy risks relating to data (re)usage (e.g., De & Le Métayer, 2016; Elliot et al., 2020).<sup>24</sup>

Many organisations across academia, industry, and health and social care service provision have designed, developed, and deployed their own versions of TREs (Madden et al., 2021). It is worthwhile to note that, in relation to re-usage of data specifically provided by the National Health Service (NHS) for research purposes, there have been some calls for a “future data access ecosystem made up of a relatively small number of TREs” (Madden et al., 2021) at national and regional levels (also see the Goldacre Review [Goldacre & Morley, 2022]). Furthermore, Our Future Health (2022) (“in partnership with the NHS”) has recently opened a consultation on a “trusted research environment accreditation process”.

<sup>24</sup> It is further worthwhile to note that in a consultation on TREs undertaken by the UK HDRA (Hubbard et al., 2020), data flow mapping is identified as a key tool for helping to “demonstrate” not just “explain” safe settings for research to key stakeholders (e.g., “data researchers”, “custodians”, “the public”).

As part of the DARE UK PRIAM project, we have identified three typical TRE deployment scenarios that have emerged to provide safe access to personal and sensitive data for analysis — i.e., ‘Data User TRE’, ‘Service Provider TRE’, and ‘Broker TRE’, see Figure 4 and Table 2 as follows:

Table 2: Three Typical TRE Deployment Scenarios

	Data User TRE	Service Provider TRE	Broker TRE
<b>Deployed at:</b>	A research institution/company	A service provider	A legal entity operating a data marketplace
<b>Data are ingested from:</b>	Primary research datasets and third-party data providers for the purpose of specific projects	Operational services and third-party data providers — there is tight coupling of operational services and the TRE	Third-party data providers
<b>Access to TRE:</b>	Limited to the institution or company employees; although some delegated access maybe possible depending on TRE capability	Service provider employees or remote access to research data users	The data marketplace brings together third party data providers and research data users from multiple organisations
<b>Needed for situations e.g., where:</b>	Data providers do not have: TREs; TREs that offer the necessary analysis tools; or third-party data agreements in place	Data providers want to retain the highest level of control over data usage	Data discovery is challenging; data aggregation and curation can significantly increase data value

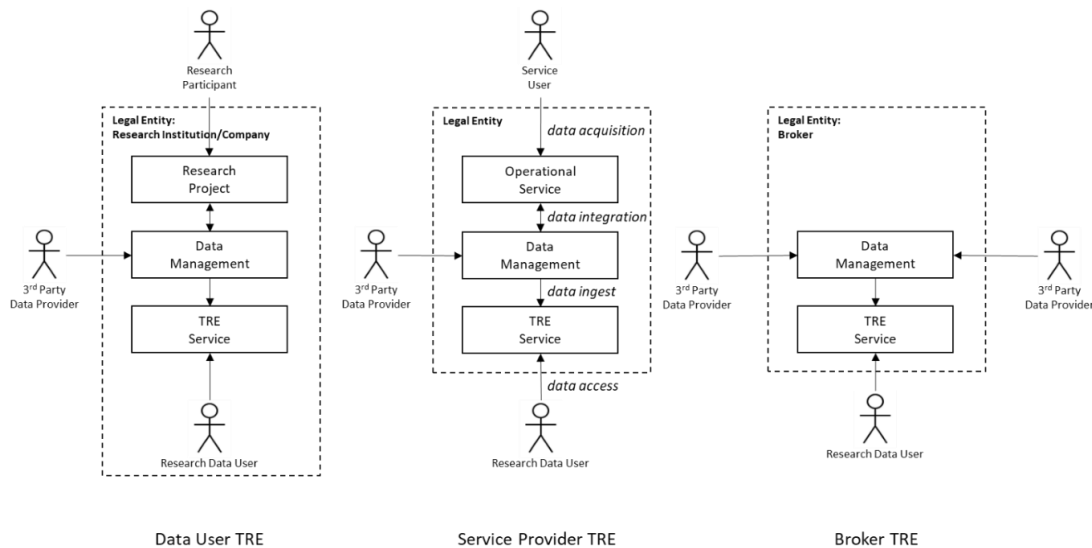


Figure 4: Typical TRE Deployment Scenarios

In all cases, the TRE operator (Legal entity) either own the data (research or operational) or have the necessary agreements/licensing in place with third-parties to process data, including consent for primary and secondary use from data subjects (where applicable). Figure 4 provides an illustration of these three typical TRE deployment scenarios. Each of these scenarios has its own advantages and disadvantages depending on the context and



purpose for data (re)usage. For example, a Service Provider TRE works well for situations where the data has been aggregated already — e.g., through service integration or data aggregation in operational systems. However, it may work less well for situations where a Research Data User wants to study new links between datasets — e.g., to explore a new service pathway between previously disconnected service providers; or to consider interplay between data in heterogeneous systems within different contexts of operation such as, mobility data and health data.

### 3.2.1. A representative example of a research network

Our use case analysis has identified that: (i) TREs exist within complex data networks; and (ii) TREs offer partial views of an overall data network. To illustrate this point further, Figure 5 (below) provides a representative example of the situation for an integrated care system (ICS), which is derived from ‘Use Case C: NHSx Wessex Federated TREs’. As discussed, the health data network is organised into layers of service integration, data aggregation and federated research:

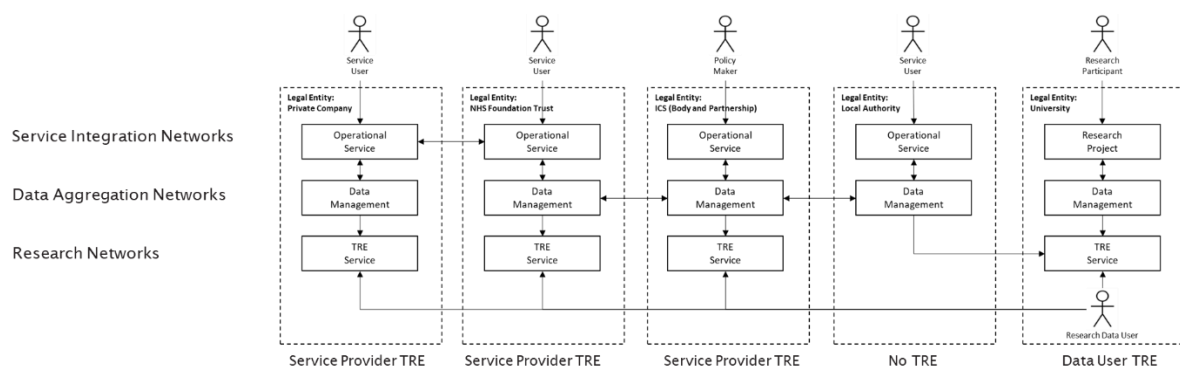


Figure 5: A Representative Example of a Research Network

We now explore these layers from this representative example of a research network (presented in Figure 5 above) in more detail:

#### Service Integration Layer:

##### Data Acquisition—‘Operational Services’ and ‘Research Projects’

- **Brief description.** Data are acquired by legal entities in operational services or individuals participating in research projects. Some operational services are connected by business to business (B2B) processes whilst others may be isolated.
- **Example.** In this case, an NHS Foundation Trust may interact with a Private Company providing patient support for a long-term condition, such as diabetes, with data shared as single or bidirectional flows depending on the care arrangements.

#### Data Aggregation Layer:

##### ‘Data Management’

- **Brief description.** Each legal entity typically operates a data management infrastructure (e.g., data warehouse, data lake) that integrates operational services and offers interfaces for information exchange with others.
- **Example.** In this case, both the NHS Foundation Trust and Local Authority exchange data via an information exchange operated by the ICS, whilst the ICS uses this data for public health management and clinical commissioning. For instance, the Private Company may curate diabetes related datasets —

e.g., continuous glucose monitoring (CGM), vital signs monitoring, activity diaries — and some may have been extracted from NHS systems (under the necessary data sharing agreements/ licensing).

## Federated Research Layer

### ‘TRE Services’

- **Brief description.** TREs are then deployed by legal entities to provide access to data for approved research collaborations (e.g., via Integrated Research Application System (IRAS), institutional ethics review). While all TREs must comply with legal, ethical and cyber-security requirements, the applicable governance arrangements and processes for each TRE is varied. The data available within each TRE are those linked by the legal entity through data management in the context of an operational service. This arrangement demonstrates the variety of access points within a federated research network and that today, TREs are largely isolated silos with data only flowing within service integration and data aggregation networks, and then ingested into TREs.
- **Example.** In this case, the Private Company, NHS Foundation Trust, ICS and University all operate TREs, but the Local Authority does not.

### 3.3. Emerging data sharing needs

In the UK Health Data Research Alliance (UK HDRA) Green Paper on TREs, Hubbard et al. (2020) outline two key themes for emerging data sharing needs in relation to research projects involving advanced analytics (AI/ML):

- ‘An enhanced research experience’ including support for ‘advanced federated analysis’ and ‘distributed machine learning’; and
- Effective ‘communications, engagement and involvement’ with stakeholders about these changing research needs.

We now explore these two themes in further detail:

#### 3.3.1. ‘An enhanced research experience’: supporting ‘advanced federated analysis’ and ‘distributed machine learning’

Researchers and data analysts require TREs that ‘enhance the research experience’ — in that, they are easy to use and efficient, provide adequate training and support, and fulfil functionality requirements, such as for analytical tools — however, risks to privacy must also be minimised (Hubbard et al., 2020). The ability to perform advanced “federated analysis” and “distributed machine learning” of data accessible via a group of national and/or cross-border TREs is identified as a “a key concern” for researchers (Hubbard et al., 2020). For example, as part of a project, a researcher needs to analyse multiple datasets that are accessible via several TREs (Hubbard et al., 2020) and as identified in all three driver use cases (Section 2). However, some TREs may restrict the export of individual-level data to another TRE for analysis (Hubbard et al., 2020; e.g., MELD-B project [Medicine — University of Southampton, 2022]) or require data providers to negotiate new data flows (e.g., PROCED project [(NIHR ARC Wessex, n.d.)]). There are further issues where a researcher needs to analyse multiple datasets accessible via TREs located in different countries, given potential restrictions on cross-border data flows (Hubbard et al., 2020).

TREs are not only expected to deal with ever increasing volumes and velocity of data and offer greater support for a wider range of data analysis tools<sup>25</sup> — such as for AI/ML — but also to be more connected with other TREs.

<sup>25</sup> For further discussion on the ‘next generation capabilities’ of TREs see: Kavianpour et al. (forthcoming/in press).

Further, there is a need for greater availability and interoperability<sup>26</sup> of quality data from service providers (e.g., for health care providers, social care providers) for research purposes.<sup>27</sup>

### 3.3.1 ‘Stakeholder involvement’: co-design and interaction with data

Hubbard et al. (2020) also highlight the crucial need for effective “communications, engagement and involvement” with key stakeholders — e.g., “public and patients”, “data custodians”, “researchers and innovators”, “TRE service providers”, “funders” — in relation to these emerging data sharing needs (Hubbard et al., 2020). As highlighted by ‘Use Case A: Complex Hospital Discharge — PROCED Project’, we should further consider how patients, service users and members of wider publics can have greater involvement with the co-design, testing and evaluation of research concept inception through to generated insights and tools.

#### 3.3.1.1. Building greater capabilities for interaction with data

One area of interest for TREs therefore is how to build greater capabilities for interaction with data (e.g., Stalla-Bourdillon et al., 2019a) for data subjects, co-designers and (where possible) other interested members of the public whilst minimising risks to privacy. For instance, TREs can act as an interface for people other than researchers and data analysts to interact with data. One example being an interactive computational notebook (Duckworth & Boniface, 2022; Duckworth et al., 2022a) — produced as part of the COdesigning Trustworthy Autonomous Diabetes Systems’ (“COTADS”) project (e.g., Duckworth et al., 2022b) — which allows users to explore “text, diagrams and interactive widgets and facilitates “codesign sessions for the application of machine learning in type-1 diabetes” (Duckworth et al., 2022a).

Further, such capabilities for interaction can also support greater ‘intervenability’ — described by Agencia Española de Protección de Datos (AEPD, 2019) in their guide to privacy by design as “the data subject’s capacity for intervention and control in the processing”. In other words, we need to consider how TREs can be built with greater capabilities for interaction with data, which empower data subjects to exercise their data-related rights<sup>28</sup> under the GDPR (subject to exemptions and restrictions) (Stalla-Bourdillon et al., 2019a). It is therefore apparent that we need to examine emerging data sharing needs from the perspectives of multiple stakeholders to ensure that value from these activities is distributed amongst them effectively.

<sup>26</sup> It is worthwhile to note that the European Interoperability Framework (European Commission, Directorate-General for Informatics, 2017) sets out a “four-layer interoperability governance model” for “integrated public service governance” as part of its “interoperability-by-design paradigm” — these layers are “legal interoperability”, “organisational interoperability”, “semantic interoperability” and “technical interoperability”.

<sup>27</sup> For instance, in response to the recent publication of the Goldacre review, HDR UK (2022) states: “Although the pandemic has shown the transformative impact of data research and innovation the potential of health data is far from being realised in full. Only a fraction of NHS, biomedical and health-relevant data is accessible to inform research. Data is of variable quality. Many datasets are still held, unconnected, in individual Institutions and/or on data platforms that lack the computing infrastructure required for advanced analysis. There are major research and technological skills shortages. Public trust and confidence in the use of health data for research remains vulnerable.”

<sup>28</sup> These data-related rights include: “Notification” (Article 16, GDPR); “Erasure” (Article 17, GDPR); “Restriction of processing” (Article 18, GDPR); “Data portability” (Article 20, GDPR); “Object” (Article 21, GDPR); and “Not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her” (Article 22, GDPR).

## 4. Examining different approaches to identifying, organising and using risk factors for privacy risk assessment

We now explore some different approaches to identifying, organising and using risk factors for privacy risk assessment through a literature review. We specifically focus on:

- The **Five Safes** (Desai et al., 2016) as well-known principles and dimensions to structure discussions and decision-making about access to sensitive data, exploring existing interpretations and adaptations. Given the ways in which data and analytics systems are changing and the new means for sharing and accessing data that are emerging (such as, the potential for federation), we also give thought to the considerations that need additional emphasis or should be featured as part privacy risk discussions and decision-making when structured using the Five Safes approach.

- Note: focus on the Five Safes approach is resumed in the **D2 report** where specific privacy risk factors are identified and categorised utilising the Five Safes (Desai et al., 2016) through our engagement with legal, ethics, regulatory and information governance experts and practitioners as part of the DARE UK PRIAM Advisory Board.

- Four selected privacy risk assessment methodologies: **Commission nationale de l'informatique et des libertés (CNIL): Privacy Impact Assessment (PIA) Methodology and Knowledge Bases** (CNIL, 2018a; 2018b); **Inria — Research Centre Grenoble: Privacy Risk Analysis Methodology** (De & Le Métayer, 2016); the **UK Anonymisation Network (UKAN): Anonymisation Decision-Making Framework (ADF)** (Elliot et al., 2020); and, the **U.S. National Institute for Standards and Technology (NIST): NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management (NIST PRAM)** (NIST, 2020a).

- Note: the System Security Modeller (SSM) (Phillips et al., 2022) is the tool used by the project to model and simulate privacy risk for a given scenario in the **D3 report**. The purpose of the SSM is to automate “cyber-security risk assessment” for “socio-technical systems”, including threats of non-compliance with the GDPR (Phillips et al., 2022). Given the SSM “follows the process of ISO 27005 and thereby supports ISO 27001 compliance” (Phillips et al., 2022), we compare these four privacy risk assessment methodologies with the **ISO/IEC 27005 methodology for information security risk management** to identify common risk factor types. We further consider the relationship between information security and information privacy risk assessment methodologies.

- **Privacy protection goals** from the field of privacy engineering.

### 4.1. The Five Safes: “Safe Projects”, “Safe People”, “Safe Settings”, “Safe Data” and “Safe Outputs”

The Five Safes is a well-established and popular approach used by many types of organisations both nationally and internationally, principally, to help structure discussion and decision-making around access to sensitive data between multiple organisations and individuals with different interests and expertise, such as law, ethics, statistical disclosure and technology (Desai et al., 2016; Ritchie, 2017). For instance, the Five Safes has been adopted by HDR UK (2021a) to help “researchers and custodians to meet the principles of transparency, safety and privacy throughout the data use cycle”. In the words of Ritchie (2017), the main purpose of the Five Safes is to “deal with these competing issues in a structured way that allows all factors to be discussed but without requiring that everything must be settled at the same time or in particular order”.

The original Five Safes devised by the Office for National Statistics (ONS) in 2003 (Desai et al., 2016) are as follows:

Table 3: Original Five Safe Questions (Desai et al., 2016)

Five Safe Dimensions	Original Five Safe Questions Devised by ONS in 2003 (Desai et al., 2016)
“Safe Projects”	“Is this use of the data appropriate?”
“Safe People”	“Can the researchers be trusted to use it in an appropriate manner?”
“Safe Data”	“Is there a disclosure risk in the data itself?”
“Safe Settings”	“Does the access facility limit unauthorised use?”
“Safe Outputs”	“Are the statistical results non-disclosive?”

#### 4.1.1.1. Interpretations of the Five Safes

Some example interpretations of the Five Safes principles are outlined in the following table:

Table 4: Some example interpretations of the Five Safe principles

	“Safe Projects”	“Safe People”	“Safe Data”	“Safe Settings”	“Safe Outputs”
<b>Desai et al. (2016):</b>	“[...] refers to the legal, moral and ethical considerations surrounding use of the data. [...] ‘Grey’ areas might exist when ‘exploitation of data’ may be acceptable if an overall ‘public good’ is realised. [...]”	“[...] reviews the knowledge, skills and incentives of the users to store and use the data appropriately. It considers the confidence of the data owner that those who will access to the data can be trusted to use it appropriately. In this context, ‘appropriately’ means ‘in accordance with the required standards of behaviour’ [...]”	“[...] refers primarily to the potential for identification in the data. It could also refer to the sensitivity of the data itself, but for argument’s sake we focus on the former case; without identification of an individual or group there is no breach. [...]” <sup>29</sup>	“[...] relates to the practical controls on the way the data is accessed. At one extreme researchers are restricted to using the data in a supervised physical location [...]; at the other, there are no restrictions on data downloaded from the internet. [...] Safe settings encompasses both the physical environment [...] but also procedural arrangements such as the supervision and auditing regimes.”	“[...] covers the residual risk in publications from sensitive data.”
<b>Health Data Research UK (HDR UK, n.d.):</b>	“Data is only used for ethical, approved research with the potential for clear public benefit”	“Only trained and specifically accredited researchers can access the data”	“Researchers only use data that have been de-identified to protect privacy”	“Access to data is only possible using secure technology systems – the data never leaves the TRE”	“All research outputs are checked to ensure they cannot be used to identify subjects”
<b>Australian Institute of Health and Welfare (AIHW, 2021):</b>	“Use of the data is legal, ethical and the project is expected to deliver public benefit”	“Researchers have the knowledge, skills and incentives to act in accordance with required standards of behaviour”	“Data has been treated appropriately to minimise the potential for identification of individuals or organisations”	“There are practical controls on the way the data is accessed – both from a technology perspective and considering the physical environment”	“A final check can be required to minimise risk when releasing the findings of the project”
<b>UK Data Service, SecureLab (2022):</b>	“Research projects are approved by data owners for the public good”	“Researchers are trained and authorised to use data safely”	“Data is treated to protect any confidentiality concerns”	“a SecureLab environment prevents unauthorised use”	“Screened and approved outputs that are non-disclosive”

<sup>29</sup> Note: that the ACS (2018) follow the interpretation of the Five Safes by Desai et al. (2016), but also highlight further elements to be considered as part of ‘safe data’ — relating to the ‘quality’, ‘completeness’, ‘richness’ and ‘sensitivity’ of the data.

<b>Arbuckle &amp; Ritchie (2019)—in the context of risk-based anonymisation:</b>	“Will personal data be anonymized? What are the legal/ethical boundaries?”	“Evaluate recipient trust, and manage their motives”	“To determine the data transformations necessary to deal with residual risk, we need to understand the risk from the data” <sup>30</sup>	“Assess security and privacy controls of the recipient”	“Evaluate context and data risk, and transform data to achieve a very low risk while maintaining ethical oversight”
--	--	--	--	---	---

It is worthwhile to note that, in some cases, the Five Safes have been adapted and extended — such as, to include “Safe Transit” as a sixth safe — i.e., “are the data sufficiently encrypted during transit” an extension by Groos & van Veen (2020); and to incorporate “Safe Algorithm” on application to AI by ACS<sup>31</sup> (2018). 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).

### 4.1.2. The importance of risk communication

Privacy is a nebulous concept<sup>32</sup> — holding various meanings for people<sup>33</sup> as well as for different stakeholder groups and disciplines.<sup>34</sup> Privacy concerns, attitudes and expectations held by individuals and, at a more generalised level, by stakeholder groups may vary depending on the circumstances in which they are being considered and can develop and change over time. Risk communication is a fundamental aspect of privacy risk assessment and is described as

*“the process of exchanging or sharing risk-related data, information and knowledge between and among different groups such as scientists, regulators, industry, consumers or the general public”* (International Risk Governance Center, 2017)

As individuals and stakeholders will have different perceptions of privacy risk — consultation with internal and external stakeholders during the course of privacy risk assessment is vital to ensure that stakeholders have the opportunity to “highlight privacy risks and solutions based on their own area of interest or expertise” and “all relevant perspectives are taken into account” (Information Commissioner’s Office, 2014). In other words, there needs to be “co-production of good governance” between stakeholders, which goes “beyond the mere provision of information” (Laurie et al., 2015). Risk communication is a crucial element in the ongoing development of a standard privacy risk assessment framework.

<sup>30</sup> Note: this is not included in original Figure 1 from the article (Arbuckle & Ritchie, 2019), which provides an overview of the Five Safes in the context of risk-based anonymisation but is instead added from the main text of article.

<sup>31</sup> The ACS is the professional association for the technology sector in Australia.

<sup>32</sup> For discussion related to different conceptualisations of privacy — including, “the right to be let alone”, “limited access to the self”, “secrecy”, “control over personal information”, “personhood” and “intimacy” — see Solove (2002). For a further conceptualisation: “privacy as contextual integrity” see: Nissenbaum (2004).

<sup>33</sup> For instance, in response to the question “what does privacy mean?” the International Association of Privacy Professionals (IAPP, n.d.) states: “Well, it depends on who you ask. Broadly speaking, privacy is the right to be let alone, or freedom from interference or intrusion. Information privacy is the right to have some control over how your personal information is collected and used.”

<sup>34</sup> The notion of privacy is viewed from many different perspectives — from fields of study (e.g., jurisprudence, privacy engineering) and disciplines across the humanities, social sciences and sciences (e.g., computer science, history, law, philosophy, psychology). For instance, The Office of the Australian Information Commissioner (n.d.) states: “Information privacy is about promoting the protection of information that says who we are, what we do and what we believe.” For further discussion on a “contextual approach to information privacy” see Wu et al. (2020). The potential limitations of focusing on ‘information privacy’ should also be noted. For instance, Wright and Raab (2014) conceive information privacy to be closely affiliated with data protection and therefore the rights of individuals — where sole focus on data protection can be “to the detriment of other types of privacy and privacy rights, which may be affected by policies and practices”. It is also worthwhile to note, in their “taxonomy of information privacy for collaborative environments”, Skinner et al. (2006) highlight key privacy dimensions, including a “structural view” of information privacy — that is, at “individual level”, “group level” and “organisation level”.



It is vital that the privacy requirements identified by the DARE UK PRiAM project are validated against the expectations of key stakeholders. For instance, for members of the public to engage (as mentioned above), they must perceive direct relevance to them and what they believe to be important. **Given the importance of risk communication, our work both involves an Advisory Board for engagement with legal, ethics, regulatory and information experts and practitioners, and a Privacy Risk Assessment Forum for engagement with members of the public. For more information on project engagement, see the D2 and D4 reports.**

#### 4.1.3. Consideration of federation in relation to the Five Safes dimensions

Our use case analysis draws attention to how advanced analytics methods are being used and supported as part of interdisciplinary collaborative research to discover value in big datasets. These methods are driving new data processing patterns (such as, to support advanced federated analysis, distributed machine learning [Hubbard et al., 2020]) and forms of research collaborations underpinned by the federated sharing and processing of data (including, types of federated research networks). All of which point towards increasingly complex flows of multi-source data between people and organisations taking place through multiple platforms as part of wider data ecosystems of shared resources and services. Given the complexity of these data flows, federation requires additional safeguards and controls — e.g., federated identity management, common data models<sup>35</sup> [UK Health Data Research Alliance, & NHSX, 2021]), the application of specific types of privacy-enhancing technologies (PETs) such as multi-party computation, differential privacy and homomorphic encryption (The Royal Society, 2019). Such controls and safeguards are required to provide the necessary assurances to stakeholders (e.g., data providers, data users, members of the public) that re-usage is trustworthy, safe and secure.

The analysis of privacy risk factors for such interdisciplinary collaborative research is therefore particularly challenging since multiple organisations are responsible for facilitating (e.g., data providers, funding bodies, TRE providers) and carrying out interdisciplinary collaborative research (e.g., universities, healthcare providers), often with different privacy concerns, expectations and attitudes. The Five Safes therefore remains an important, familiar tool for bringing together these organisations around common principles and dimensions, in this case to assess, communicate and make decisions related to privacy risk and data access for specified research activities (e.g., a project, programme of work). We highlight some key reasons for focusing on the Five Safes (Plus One), as part of privacy risk assessment:

- **Familiarity** — in that, many stakeholders are likely to have awareness of or previous experience of using the Five Safes.
- **Comprehension** — in that, seemingly more complex notions around privacy (e.g., those related to privacy engineering, conceptualisations of privacy) can be made more accessible to all stakeholders, for discussion and decision-making, if surfaced through the Five Safes (Plus One) where the high-level categories of ‘people, projects, settings, data, outputs and return’ are likely to be more clearly understood.
- **Cohesion** — in that, as a common approach for structuring decision-making and discussion as well as categorising privacy risk factors, the Five Safes (Plus One) can bring together those issues related to privacy risk assessment with issues raised in other pertinent areas (e.g., intellectual property rights clearance and management). Such synthesis of diverse requirements is of vital importance to inform socio-technical data governance models for individual and federated ecosystems of TREs.

It is therefore of paramount importance that the ways in which data and analytics systems are changing and the new means for sharing and accessing data that are emerging (such as, the potential for federation) are made to

---

<sup>35</sup> For instance, UK Health Data Research Alliance, & NHSX (2021) outline ‘best practice principles’ for a group of federated TREs (in this case, an Alliance TRE ecosystem) in terms of the Federated Five Safes.



clearly stand out throughout the Five Safes (Plus One) dimensions. It is worthwhile to note the references to the Federated Five Safes made by UK Health Data Research Alliance, & NHSX (2021) and as part of “Federated Data Analytics Infrastructure - Capability Maturity Model” outlined by HDR UK (2021b). In particular, the following aspects must be highlighted when the Five Safes (Plus One) are utilised for this type of assessment:

#### 4.1.3.1. *Safe Projects: Highlighting different types of research collaborations*

As part of the ‘Safe Projects’ dimension it should be recognised that organisations are coming together not only to share resources and services for collaborative research at project-level, but also at scale e.g., to support a programme of work, a federated research network. The Five Safes (Plus One) therefore have a wider remit than individual projects, as organisations may also utilise the Five Safes (Plus One) to structure discussion and decision-making concerning privacy risks related to data access and re-usage at scale e.g., involving multiple connected projects and platforms.

#### 4.1.3.2. *Safe People: Drawing attention to a wider-range of stakeholders*

As part of the ‘Safe People’ dimension, it should be highlighted that alongside researchers and data analysts (often the main focus of this dimension) — a much wider range of people and organisations have responsibility for, access to and influence over resources and services utilised as part of research collaborations e.g., co-designers, data providers, data subjects, data stewards, TRE operators.<sup>36</sup> Further, in terms of privacy risk assessment, there is a need to consider the relationships between all stakeholders (De & Le Métayer, 2016) — not just researchers and data analysts for a particular project. ‘Stakeholders’ therefore is a key component as part of the Privacy Risk Analysis Methodology set out by De & Le Métayer (2016) who describe stakeholders as individuals and organisations who data “relates to”, “processes” data, or has (un)lawful “access” to data during “any stage of its lifecycle”. Further, according to De & Le Métayer (2016), the stakeholders under consideration should be described with regard to the following attributes: (1) “the data flow view” — that is, “depicting how data flows across stakeholders” which is distinct from the ‘data flows’ attribute under the ‘system component’; and (2) “stakeholder relationships” (as appropriate) — including “trust, hierarchical dependency, economic dependency”. Therefore, in this context, the ‘Safe People’ dimension (and ‘Safe Return’ dimension) should further motivate organisations to deliberate on the relationships between organisations and people involved in a research collaboration — and the possible implications for privacy risk.

#### 4.1.3.3. *Safe Settings: Emphasising complex data flows*

The ‘Safe Settings’ dimension (typically focused on a single data access facility or platform, see Table 4) should call attention to how emerging data usage patterns and needs for collaborative research involves increased flows of data to and from multiple platforms as part of a wider data ecosystem of shared resources and services (including federated ecosystems of TREs). For instance, in this context, the ‘Safe Settings’ dimension should encourage organisations to consider e.g., the governance and technical mechanisms to ensure safe and secure data flows between specified platforms, the degree of interoperability between different platforms, provenance of data flows — and the possible implications for privacy risk.<sup>37</sup>

---

<sup>36</sup> For instance, in some cases, there may be obvious and known links between projects (e.g., those with the same principal investigator, researchers from the same organisation working on different projects). In other cases, these links may be less obvious. We need to consider the links between projects — ‘how does data flow between projects within a programme?’ — to understand privacy risks related to: connecting analyses; combined learning and projects; expectations for data access as part of a programme of work; and the temporal nature of some programmes of projects and work. For further background information on data institutions see: Dodds et al. (2020). For discussion of “Determining your responsibilities across a data flow” see Elliot et al., (2020).

<sup>37</sup> In terms of privacy risk assessment, there is a need to consider e.g., how third-party data sources (those outside the boundaries of a specified research collaboration) may introduce risk; the potential for ‘mosaic effects’ as datasets can be used multiple times as part of the same programme of work or by specified researchers and data analysts; and risks of data flows between multiple systems and stakeholders.

## 4.2. Comparing the ISO/IEC 27005 methodology for information security risk management with selected privacy risk assessment methodologies

While there are several risk assessment methodologies in existence, some of which address information security and others that concentrate on information privacy specifically, there is a need for a standard privacy risk assessment framework that can fully deal with privacy risks arising from emerging data patterns and needs of advanced analytics in cross council research networks (as explored in Section 3). We now explore a selection of these methodologies in more detail.

### 4.2.1. Sources

Common types of risk factors are identified by taking a baseline of cybersecurity risk assessment, a closely related (but distinct) field to privacy risk assessment (see Sub-Section 4.4 for more detail), plus existing privacy risk assessment methodologies themselves; and considering the common concepts of risk management used. Note that a risk factor is described by NIST (2012) as “[a] characteristic used in a risk model as an input to determining the level of risk in a risk assessment”. In other words, such factors are elements that cause and affect risks<sup>38</sup> e.g., assets, consequences, controls, threats, vulnerabilities — as outlined by ISO/IEC 27005. The consideration of key risk factors is necessary to help specify the types of information required for assessing privacy risks arising in relation to collaborative research.

Across these cybersecurity and privacy risk assessment approaches, there are different terms used, but there are considerable crossovers in the underlying concepts the terms represent. The main objectives of this exercise are: (i) to determine the core concepts; (ii) how they relate to each other; and (iii) to propose a common naming convention for the types of factors that affect privacy risk in federated situations. These selected sources are:

- Commission nationale de l'informatique et des libertés (CNIL): Privacy Impact Assessment (PIA) Methodology and Knowledge Bases (CNIL, 2018a; 2018b).
- UK Anonymisation Network (UKAN): Anonymisation Decision-Making Framework (ADF) (Elliot et al., 2020).
- Inria — Research Centre Grenoble: Privacy Risk Analysis Methodology (De & Le Métayer, 2016).
- U.S. National Institute for Standards and Technology (NIST): NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management (NIST PRAM) (NIST, 2020a).
- ISO/IEC 27005:2018. Information technology—Security techniques—Information security risk management.
- Request for Comments: 4949 (RFC 4949). Internet Security Glossary, Version 2. Network Working Group, August 2007 (Shirey, 2007).
- ISO/IEC 27000:2018. Information technology—Security techniques—Information security management systems—Overview and vocabulary.

**We now provide a brief overview of each of these approaches:**

#### 4.2.1.1. CNIL PIA (CNIL, 2018a; 2018b)

The **CNIL PIA** (2018a) provides guidance for people (e.g., “project owners”, “data protection officers”, “decision-making authorities”) and organisations on how to carry out a data protection impact assessment (DPIA) pursuant

<sup>38</sup> For instance, in some cases, one or more key factors will contribute to privacy risks by ‘amplifying’ the likelihood and severity of a privacy risk (International Risk Governance Council, 2010). Some examples of such increasing factors include “the volume of the breached data (for the same individual)”; “special characteristics of the data controller”; and “special characteristics of the individuals” (European Union Agency for Cybersecurity (ENISA), 2013). In other cases, one or more key factors will contribute to privacy risks by decreasing the likelihood and severity of a privacy risk (International Risk Governance Council, 2010). Some examples of such decreasing factors include “public availability”; and “nature of data” (ENISA, 2013).

to Article 35 of the GDPR. This methodology is compatible with “international standards on risk management (such as [ISO 31000])” and best practice guidance from the Art 29 Data Protection Working Party (2017) on DPIAs (CNIL, 2018a). The CNIL PIA (2018a) refers to a construction of risks where threats lead to feared events. Threats are comprised of risk sources and operate on assets; and feared events comprise potential impacts on personal data. Controls are referred to as a means of modifying the risk level. The risk assessment focuses on the determination of the likelihood and severity of a feared event, where severity is the impact on data subjects’ privacy and likelihood is how likely the feared event is given the threats, the vulnerabilities of the assets concerned and the controls already in place.

The **CNIL PIA Knowledge Bases** (2018b) resource provides “a catalogue of controls aimed at complying with the legal requirements and treating the risks”, including: various typologies, such as for “risk sources” and “outcomes of feared events”; “scales and rules” for “estimating severity” and “likelihood”; and good practices e.g., for measures used to reduce risks, for empowering data subjects to exercise their rights through intervenability, for data security.

#### 4.2.1.2. UK ADF (Elliot et al., 2020)

The **UKAN ADF** is a practical anonymisation guide offering “operational advice” to people and organisations who need to be able to “anonymise” data “with confidence” (Elliot et al., 2020). It aims to offer “a way of thinking about anonymisation and the reuse of personal data that breaks out of the constraints of overly technical or overly legal framings of the problem” (Elliot et al., 2020). The ADF covers “two frames of action”: (1) “technical element” — to enable the user of the guide “to think about both the quantification of disclosure risk and how to manage it”; and (2) “contextual element” — to enable the user of the guide “to think about and address the factors that affect that risk” (Elliot et al., 2020).

The UKAN ADF approach is a situational analysis of data usage with a view to assessing disclosure risks and functional anonymisation. It assesses data, data flows, data analysis situations, data processing environments and stakeholders. The UKAN ADF considers the notion of multiple data environments, where environments are described in terms of agents, other data, governance and infrastructure. Given these situations can concern data flows between legal entities, this approach has significance for safe research collaborations. The UKAN ADF assesses the likelihood of an “adverse event”, which — given the focus of the UKAN ADF on “disclosure risk assessment and control” — is “most often the re-identification of a data unit” and the “data situation sensitivity, will be used to describe those elements which affect the impact of an adverse event” (all quotes Elliot et al., 2020).

#### 4.2.1.3. Inria Privacy Risk Analysis Methodology (De & Le Métayer, 2016)

The **Inria Privacy Risk Analysis Methodology** takes the perspective of a systemic analysis of data flows and processing:

*“The framework relies on the definition of appropriate categories and attributes of seven components (system, stakeholders, data, risk sources, feared events, harms and privacy weaknesses). The methodology is made of two main phases: information gathering and risk assessment leading to a well-defined risk assessment process based on harm trees” [...] “which are used to establish a relationship among privacy weaknesses, feared events and harms” (De & Le Métayer, 2016)*

#### 4.2.1.4. NIST PRAM (NIST, 2020a)

The **National Institute of Standards and Technology (NIST)** published the **Privacy Risk Assessment Methodology (PRAM)** to help better analyse, manage, and develop mitigation methods for privacy risks in a system. It requires

a comprehensive and systematic review of a system, which includes an examination of the data flow of each asset in the system to undertake a risk assessment of data disclosure and harms for each asset (NIST, 2020a). It analyses data processing for “problematic data actions”<sup>39</sup> (NIST, n.d.; 2019) using the privacy risk model and privacy engineering objectives defined in NISTIR 8062 (Brooks et al., 2017). For instance, data processing includes “collection”, “retention”, “logging”, “generation”, “transformation”, “disclosure”, “transfer”, and “disposal” (NIST 2020a: ‘Worksheet 2 – Supporting Data Map’; 2020b). ‘An occurrence or prospective occurrence of problematic data actions’ is considered as a “privacy event” (NIST 2020b). A privacy event can generate harm, including loss of self-determination, discrimination, loss of trust, and economic loss (NIST 2020b). As a result of risk assessment, it aims to prioritise privacy risks based on two dimensions, “likelihood” and “impact” (NIST 2020b). Impact is analysed based on factors such as, “non-compliance costs, direct business costs, reputational costs, and internal culture costs” (NIST 2020b).

#### 4.2.1.5. ISO/IEC 27005

**ISO/IEC 27005** is an asset-based cybersecurity risk modelling methodology. This approach aims to defend systems against cybersecurity attacks by assessing the key assets of a system, the threats that could affect them, their vulnerabilities that expose them to the threats and the consequences if the threats successfully affect the assets. Its key factor identification activities are shown in Figure 6. The ISO/IEC 27005 methodology helps organisations to identify systemic assets together with vulnerabilities and threats that can exploit them. An "incident" links a vulnerability on an asset with a threat that may exploit it, and the associated consequences that may result. It further helps organisations to identify controls that can be enabled to reduce weaknesses and therefore lower the likelihood of successful attacks.

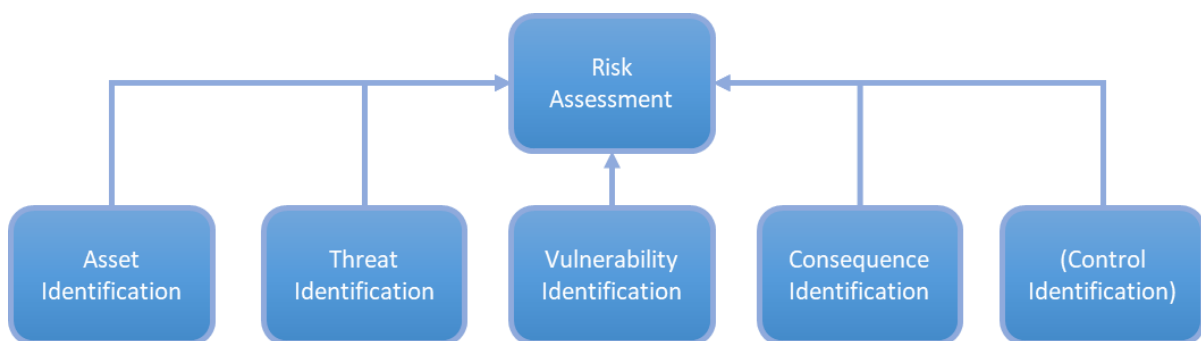


Figure 6: ISO 27005 Risk Assessment Identification Activities

#### 4.2.1.6. RFC 4949 and ISO 27000

**RFC 4949** and **ISO 27000** provide vocabularies of terms that are relevant to risk assessment. ISO 27000 determines the underlying nomenclature used by ISO2005. Both ISO 27000 and RFC 4949 are well accepted *de facto* resources for term definition in cybersecurity risk analysis, hence their inclusion.

<sup>39</sup> See section 1.1 of this report for further information.

## 4.2.2. Identifying common types of risk factors

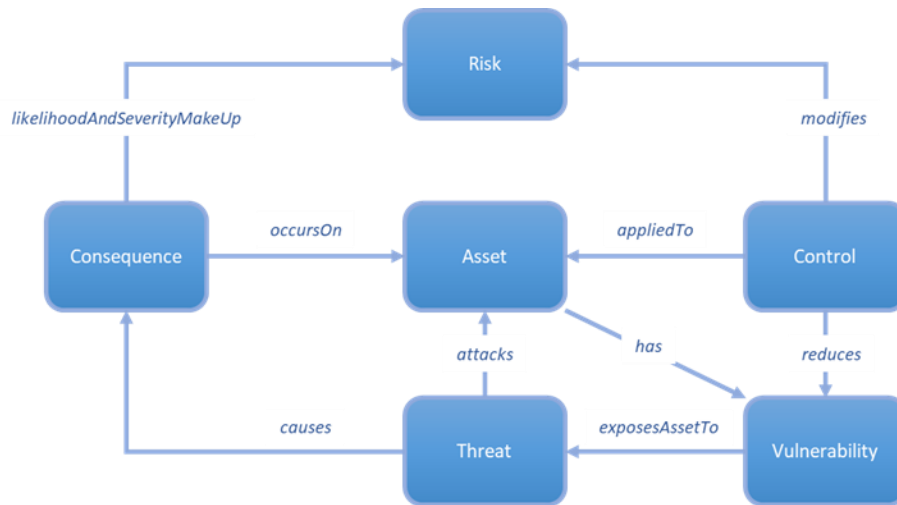


Figure 7: Risk Management Upper Ontology

In order to understand the types of risk factors needed for privacy risk management, common concepts of risk management need to be defined. The above sources have been used to determine an upper ontology for privacy risk management, which reflects common concepts shared between these methodologies. The first draft of this ontology is shown in Figure 7 (above), followed by definitions and their mapping to the sources in Table 5, which are further elaborated in the DARE UK PRIAM D3 report.

The following table provides a mapping of these risk management concepts between the areas of cybersecurity risk assessment and privacy risk assessment:

Table 5: Mapping Risk Management Concepts to Privacy Risk Assessment

Risk Management Concepts	Definition(s) from Cybersecurity Risk Assessment	Mapping these Risk Management Concepts to Privacy Risk Assessment
<b>Asset</b>	<p>“A system resource that is (a) required to be protected by an information system's security policy, (b) intended to be protected by a countermeasure, or (c) required for a system's mission” — as defined by RFC 4949 (Shirey, 2007).</p> <p>“An asset is anything that has value to the organization and which, therefore, requires protection. For the identification of assets, it should be borne in mind that an information system consists of more than hardware and software” — as defined by ISO 27005.</p>	<p>CNIL PIA (2018a) defines Supporting Asset as “Asset on which personal data rely. [/] Note: this may be hardware, software, networks, people, paper or paper transmission channels.”</p> <p>Inria Privacy Risk Analysis Methodology also considers Supporting Asset as “such as hardware, applications, data stores, software environment, etc.” (De &amp; Le Métayer, 2016).</p> <p><b>Also, Data Actions, Data and Relevant Contextual Factors:</b></p> <p>NIST PRAM focuses on identifying and classifying “Data actions being performed by the system”; “Data being processed by the data actions” and “Relevant contextual factors” — as outlined by “Worksheet 2: Assessing System Design; Supporting Data Map (version February 2019)” (NIST, 2020a).</p>

		The main focus of NIST PRAM therefore is on data actions rather than assets.
<b>Threat</b>	<p>“Potential cause of an unwanted incident, which may result in harm to a system or organisation” — as defined by ISO 27000.</p> <p>“A potential for violation of security, which exists when there is an entity, circumstance, capability, action, or event that could cause harm. A threat consists of a 'threat action' and 'threat consequences’” — as defined by RFC 4949 (Shirey, 2007).</p>	<p>CNIL PIA (2018a) defines Threat as “Procedure comprising one or more individual actions on data supporting assets”.</p> <p><b>Problematic Data Action is used by NIST PRAM rather than Threat and Vulnerabilities:</b> “A data action that causes an adverse effect, or problem, for individuals” (Brooks et al., 2017).</p>
<b>Consequence</b>	<p>“Outcome of an event affecting objectives” — as defined by ISO 27000.</p> <p><b>Also, Threat Consequence:</b> “A security violation that results from a threat action. The basic types are 'unauthorized disclosure', 'deception', 'disruption' and 'usurpation’” — as defined by RFC 4949 (Shirey, 2007).</p> <p>ISO 27000 notes that events can have a range of consequences, that can be certain or uncertain but usually negative, expressed qualitatively or quantitatively. Also, initial consequences (from an event) can escalate through knock-on effects. Consequence is the conjunction of the impact and the likelihood of the events that cause the consequence.</p>	<p>For privacy risk assessment, Consequence can be viewed in relation to the occurrence of “feared events” that generate “impacts on the privacy of data subjects” (CNIL PIA) — i.e., Privacy Harms. These two concepts are defined as follows:</p> <p><b>Feared Event:</b></p> <p>CNIL PIA (2018a) defines Feared Event as “Potential data breach likely to have impacts on data subjects’ privacy”.</p> <p>Inria Privacy Risk Analysis Methodology defines Feared Event as “an event of the system that occurs as a result of the exploitation of one or more privacy weaknesses and may lead to privacy harms” (De &amp; Le Métayer, 2016).</p> <p><b>Privacy Harm:</b></p> <p>Inria Privacy Risk Analysis Methodology defines Privacy Harm as “the negative impact on a data subject, or a group of data subjects, or the society as a whole, from the standpoint of physical, mental, or financial well-being or reputation, dignity, freedom, acceptance in society, self-actualization, domestic life, freedom of expression, or any fundamental right, resulting from one or more feared events” (De &amp; Le Métayer, 2016).</p> <p>NIST defines Privacy Harms as “any adverse effects that would be experienced by an individual whose [personal identifiable information] PII was the subject of a loss of confidentiality, as well as any adverse effects experienced by the organization that maintains the PII” — as defined by NIST 800-12 (McCallister et al., 2010).</p>



		<p>Privacy Harms can be considered a specific type of Consequence.</p> <p><b>Also, Problems:</b></p> <p>In their Catalog of Problematic Data Actions and Problems”, NIST (2019) set out five key problems for individuals: “dignity loss”; “discrimination”; “economic loss”; “loss of self-determination”, including “loss of autonomy”, “loss of liberty” and “physical harm”; and “loss of trust” (as also highlighted in Section 1.1 of this report).</p>
<b>Vulnerability</b>	<p>“Weakness of an asset or control that can be exploited by one or more threats” — as defined by ISO 27000.</p> <p>“(I) A flaw or weakness in a system's design, implementation, or operation and management that could be exploited to violate the system's security policy” — as defined by RFC 4949 (Shirey, 2007).</p> <p>The term 'vulnerability' is sometimes used to mean 'software vulnerabilities' (a specific type of vulnerability), and sometimes to mean 'threats to a system for which there are no controls' (a restriction based on vulnerability status). ISO 27000 does not include either of these restrictions and our interpretation of vulnerability can apply to any systemic asset including ICT hardware, computer software, networking, places, people and governance to reflect weaknesses that may increase the likelihood of their being affected by threats.</p>	<p>CNIL PIA (2018a) refers to the “the level of vulnerabilities of personal data supporting assets”.</p> <p>As a “more general term than vulnerabilities”, Inria Privacy Risk Analysis Methodology utilises the term <b>Privacy Weakness</b>: “a weakness in the data protection mechanisms (whether technical, organizational or legal) of a system or lack thereof that can ultimately result in privacy harms” (De &amp; Le Métayer, 2016).</p> <p>Again, note <b>Problematic Data Action</b> in NISTIR 8062 is used rather than Threat and Vulnerabilities (Brooks et al., 2017).</p>
<b>Risk</b>	<p>“Effect of uncertainty on objectives” — as defined by ISO 27000.</p>	<p>Definitions of risk typically refer to the combined likelihood and severity on assets of consequences arising from threats: “A measure of the extent to which an entity [Asset] is threatened by a potential circumstance or event [Threat], and typically a function of: (i) the adverse impacts [Consequences] that would arise if the circumstance or event occurs; and (ii) the likelihood of occurrence.” (NIST, 2020b).</p>
<b>Control</b>	<p>“Measure that is modifying risk. May include any process, policy, device, practice or other</p>	<p><b>Privacy Control:</b><sup>40</sup> “The administrative, technical, and physical safeguards employed</p>

<sup>40</sup> Note that, in general terms, privacy controls can be divided into two groups: (i) controls on data — i.e., those that transform the data itself, such as de-identification techniques; and (ii) environmental controls—i.e., those that change the environment in which the data is processed. There are therefore various types of action that can be taken to mitigate privacy risk, including privacy enhancing technologies (PETs) (e.g., The Royal Society, 2019) — for further examples of different types of privacy controls e.g., see: AEPD (2019), CNIL PIA Knowledge Base (CNIL, 2018b), Conference of the Independent Data Protection Supervisory Authorities of the Federation and the Länder (CIDPSAFL, 2020). Further, note that Stalla-Bourdillon et al. (2019a) classify controls as “corrective controls”, “detective controls”, “directive controls” and “preventative controls”.



	<p>action. Controls may not always exert the intended or assumed modifying effect” — as defined by ISO 27000.</p> <p>Also, Security Control: “The management, operational, and technical controls (safeguards or countermeasures) prescribed for an information system which, taken together, satisfy the specified security requirements and adequately protect the confidentiality, integrity, and availability of the system and its information” — as defined by RFC 4949 (Shirey, 2007).</p>	<p>within an agency to ensure compliance with applicable privacy requirements and manage privacy risks” — as defined by NISTIR 8062 (Brooks et al., 2017).</p> <p>CNIL PIA defines Control as “Action to be taken. [/] Note: this may be technical or organisational and may entail putting fundamental principles into practice or avoiding, reducing, transferring or assuming all or part of the risks”.</p> <p>Inria Privacy Risk Analysis Methodology describes controls consisting of “legal measures” (e.g., “contracts”, “privacy statements”); “organizational measures” (e.g., “training”, “incident management”) and “technical measures” (e.g., “encryption schemes”, “access controls”) (De &amp; Le Métayer, 2016). Further, Inria Privacy Risk Analysis Methodology highlights that an assessment of the controls already implemented can “provide information about the strength of the data protection mechanisms already in place” and “is therefore a major determinant of the privacy weaknesses of the system” (De &amp; Le Métayer, 2016).</p>
--	---	--

### 4.2.3. Determining the scope of the risk assessment

The boundaries and scope of the risk assessment need to be identified (ISO 27005). These will determine the types of risk factors that are of concern, specifically including those under the control of key stakeholders, and other factors that influence the risk assessment but are beyond their control. The risk assessment scope is analogous to the concept of **System** from the Inria Privacy Risk Analysis Methodology, which

*“defines the logical boundary of the PRA [Privacy Risk Assessment]. It should encompass the entire life-cycle of the personal data for the application (or set of applications) considered. It consists of various hardware and software components.” (De & Le Métayer, 2016)*

The risk assessment scope also corresponds with the concept of the **Data Situation** from the UKAN ADF, which is described as

*“the data and their environment as a total system (which we call the data situation)” [...] “Formally, a data situation is the aggregate set of relationships between some data and the set of their environments.” (Elliot et al., 2020)*

ISO 27005 has a Context Establishment setup step in its methodology (Figure 8 below) where the

*“external and internal context for information security risk management should be established, which involves setting the basic criteria necessary for information security risk*

*management [...], defining the scope and boundaries [...], and establishing an appropriate organization operating the information security risk management” (ISO 27005)*

Interpreted in the context of research collaborations, this setup step can help with identification of

- **‘Scope and Boundaries’** — e.g., the physical, organisational and technical boundaries of the socio-technical system under examination and its governance processes in relation to a research collaboration (e.g., a specified research project, programme of projects, provision of long-term resources as part of federated research networks).
- **‘Purpose of Risk Management’** — e.g., the risks that need to be avoided or minimised to ensure safe and useful research whilst taking into consideration the expectations of key stakeholders for a research collaboration.
- **‘Criteria for Risk Management’** — e.g., the impact and criticality of key systemic elements, and the associated acceptance criteria such as, the levels of risk that are tolerable on them.
- **‘Key Stakeholders’** that have interest in the socio-technical system and its risks, e.g., the actors with related individual (such as, a researcher) and institutional roles (such as, healthcare provider).

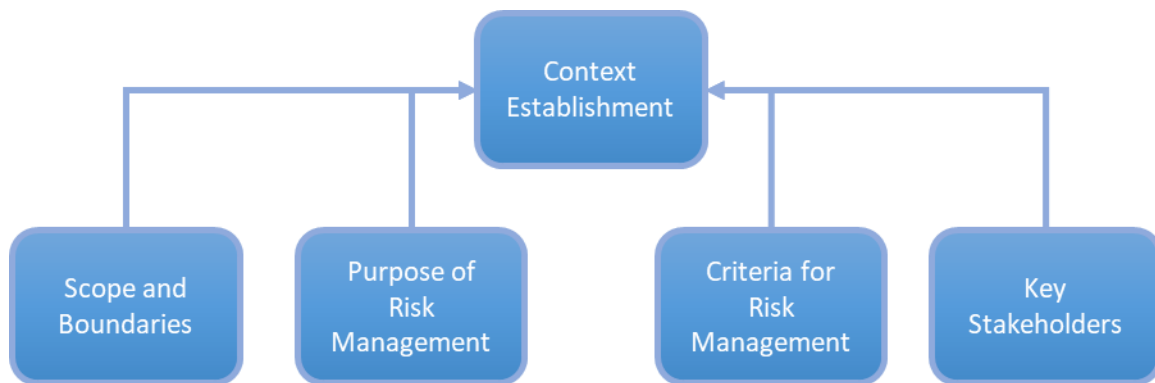


Figure 8: ISO 27005 Context Establishment

While traditional risk assessments, e.g., ISO/IEC 27005, often focus on a fixed scope (such as, the operations of a company), a key distinguishing aspect of emerging data usage patterns for research collaborations is in their varying degrees of fluidity across different levels. For example, where combinations of resources and services are utilised by multiple types of users for specific (connected) projects as well as within and across different programmes of work that encapsulate specified projects.

### 4.3. Privacy protection goals

For privacy risk assessment, best practice from the field of privacy engineering exists in the form of ‘Privacy Protection Goals’, which represent design, implementation and operational principles (these are discussed in the following sub-section).<sup>41</sup> Privacy protection goals are described as providing “a common scheme for addressing the legal, technical, economic, and societal dimensions of privacy and data protection in complex IT systems” (Hansen et al., 2015).<sup>42</sup> The Standard Data Protection Model (Conference of the Independent Data Protection

<sup>41</sup> Note that NIST refers to ‘privacy goals’ as ‘privacy engineering objectives’ — and alongside the CIA triad proposes the following three privacy engineering objectives: “predictability”, “manageability” and “disassociability” aligned with the ‘Fair Information Practice Principles’ (FIPPs) and Circular A-130 FIPPs (Brooks et. al., 2017). For the purposes of the project, we are focusing on the data protection goals outlined by the SDM methodology (Conference of the Independent Data Protection Supervisory Authorities of the Federation and the Länder, 2020) — as these directly relate to the GDPR.

<sup>42</sup> As a further description, a key purpose for these goals is to “establish a global framework of protection in personal data processing and determine, by means of a risk assessment, other non-functional attributes or requirements that the system must satisfy and which become entry points to privacy design processes” (AEPD, 2019).

Supervisory Authorities of the Federation and the Länder [CIDPSAFL, 2020) provides a “systematic approach” to GDPR compliance and risk assessment by transforming “the regulatory requirements of the GDPR to qualified technical and organisational measures” through seven protection goals:<sup>43</sup>

- **“Data minimisation”** — that is, “the fundamental requirement under data protection law to limit the processing of personal data to what is appropriate, substantial and necessary for the purpose” (CIDPSAFL, 2020)
- **“Availability”** — that is, “the requirement that access to personal data and their processing is possible without delay and that the data can be used properly in the intended process” (CIDPSAFL, 2020)
- **“Integrity”** — that is, “(i) “the requirement that information technology processes and systems continuously comply with the specifications that were defined for them to perform their intended functions” and (ii) “the data to be processed remain intact” (CIDPSAFL, 2020)
- **“Confidentiality”** — that is, “the requirement that no unauthorised person can access or use personal data” (CIDPSAFL, 2020)
- **“Unlinkability”** — that is, “the requirement that personal data shall not be merged, i.e., linked” (CIDPSAFL, 2020)
- **“Transparency”<sup>44</sup>** — that is, “the requirement that [...] data subjects [...] and system operators [...] and competent supervisory bodies [...] shall be able to identify to varying degrees which data are collected and processed when and for what purpose [...], which systems and processes are used [...], and who has legal responsibility for the data and systems in the various phases of data process” (CIDPSAFL, 2020)
- **“Intervenability”** — that is, “the requirement that the data subjects’ rights [...] are granted without undue delay and effectively if the legal requirements exist” (CIDPSAFL, 2020)

The Standard Data Protection Model also identifies various “generic organisational and technical measures” as a means to guarantee each of these seven protection goals (CIDPSAFL, 2020).<sup>45</sup> These measures have been “tried and tested in the data protection audit practices of several data protection supervisory authorities for many years” (CIDPSAFL, 2020).

#### 4.3.1. Tensions between these Privacy Protection Goals

It is important to note these protection goals are ‘complementary’ and may “overlap” (AEPD, 2019) — further “there is no possibility to ensure 100% of each of the goals simultaneously” (Hansen et al., 2015). For instance, tensions have been highlighted by Hansen et al. (2015) between:

- **Confidentiality and availability** — e.g., where restrictions on data accessibility may conflict with access to data without delay (Hansen et al., 2015).
- **Integrity and intervenability** — e.g., where the need to modify data may conflict with a requirement to keep data intact (Hansen et al., 2015).

<sup>43</sup> These regulatory requirements include the seven core data protection principles outlined by Article 5 of the GDPR — that is, “lawfulness, fairness and transparency”, “purpose limitation”, “data minimisation”, “accuracy”, “storage limitation”, “integrity and confidentiality” and “accountability” — as well as individuals rights and other requirements. For full information about how these seven protection goals are mapped to GDPR rights and obligations, see the Standard Data Protection Model (CIDPSAFL, 2020). Also, note that in a slightly different approach, AEPD (2019) and Hansen et al. (2015) both focus six protection goals, all those outlined by the Standard Data Protection Model apart from data minimisation, which is presented instead as a sub-category of unlinkability.

<sup>44</sup> Note that one of the key recommendations from a recent “public-dialogue” on “building trustworthy national infrastructure” carried out by DARE UK and Kohlrabi Consulting focuses on transparency — i.e., “Proactive transparency should be practiced by those handling and using sensitive data for research” (Harkness et al., 2022).

<sup>45</sup> Stalla-Bourdillon (2019c) states: “The SDM correctly conceives Data Protection Principles as goals, because the GDPR does not offer an exhaustive list of controls for each principle, and ultimately, the choice of the applicable controls should depend upon a trade-off between privacy and utility set in context.”

- **Unlinkability and transparency** — e.g., where audit monitoring regimes to ensure transparency of data processing activities may conflict with unlinkability aims to reduce information about data processing. (Hansen et al., 2015).

In addition to these conflicting examples, tensions also often arise between data minimisation and data utility. Therefore, an essential part of privacy risk assessment is striking an appropriate balance between these conflicting demands that is acceptable to key stakeholders — by acknowledging the implications of this balancing exercise on the scope and nature of privacy protection measures selected and implemented and their effect on the overall data utility and generated insights.<sup>46</sup> This balancing exercise therefore requires employment of robust risk communication mechanisms, including meaningful stakeholder involvement, and readiness of appropriate and effective technical and organisational measures to minimise existing privacy risks.

### 4.3.2. Relationship between Privacy Protection Goals and Risk Management

As a pointer to further work in WP3, a proposed approach for incorporating these Privacy Protection Goals into Risk Management is to consider the risk of failure to achieve a privacy goal. This can be modelled as a Consequence that leads to a privacy harm (examples of which are in Section 1.1 of this report) — i.e., failure to meet a privacy goal is caused by a threat and associated with a risk likelihood and severity. Given the modelling tool for WP3 — that is, the University of Southampton “System Security Modeller” (SSM) (Phillips et al., 2022) — already supports the Consequences: “Loss of Confidentiality”, “Loss of Availability” and “Loss of Integrity”, a precedent already exists for this concept. These other goals therefore can be modelled in a similar fashion — i.e., Consequences representing the compromise of data minimisation, unlinkability, transparency, and intervenability, represented e.g., as “Loss of Data Minimisation”, “Loss of Unlinkability”, etc.

In the SSM tool, the modeller sets the severity of a Consequence (i.e., how bad it would be if the Consequence happened), and the tool automatically determines the likelihood of the Consequence based on the vulnerabilities of the assets affected and the combined likelihood of the threats that cause the Consequence. By this mechanism, the combined severity and likelihood of compromise of a goal reflects that some goals may be more important than others in some situations (i.e., the severity of failure to meet the goal); and may be harder or easier to achieve in others (i.e., the likelihood of failure). Conflicts or tensions between goals can be therefore explored via modelling of the risk via different combinations in terms of the relative severity and likelihood of failure for the Consequences representing the privacy protection goals.

## 4.4. On the relationship between information security and information privacy risk assessment methodologies

While concerns associated with information security and information privacy are conceptually related, each area offers a distinct focus (Bambauer, 2013; Brookes et al., 2017; Kuner et al., 2017; IAPP, n.d.). We highlight some key considerations:

### 4.4.1. Consideration of impacts from (un)authorised processing

Information security risk assessment primarily focuses on risks arising from unauthorised activities that result in unwanted consequences, such as losses of confidentiality, availability, and integrity of important elements in an ICT system (Brooks et al., 2017). In contrast, information privacy risk assessment concentrates on risks in relation to both unauthorised and authorised data-related activities (Brooks et al., 2017; Dayarathna, 2011) that can cause

<sup>46</sup> For instance, according to Wottrich et al., willingness to share personal information is frequently determined by a privacy calculus in which conflicting factors are weighed to maximise benefits while minimising risks (Wottrich et al., 2018).

potential harm to individuals, groups of people, and wider society.<sup>47</sup> Breaches of information privacy can therefore result from malicious actions — such as, to disclose or modify personal data intentionally without authorisation; and from inadvertent actions — such as, to accidentally disclose or modify personal data, e.g., through “loss of equipment” or “sending mail to the wrong recipients” (Dayarathna, 2011).

Privacy protection goals from the field of privacy engineering aim to address the ethical, legal, organisational, and technical aspects of information privacy and data protection in relation to both authorised and unauthorised data processing activities in socio-technical systems. These goals therefore do not only concentrate on ensuring confidentiality, availability, and integrity (‘data security’), but also extend their focus to the goals of data minimisation, unlinkability, transparency, and intervenability (information privacy and data protection), e.g., to help achieve the overall objective of safe and secure data sharing and (re)usage.

#### 4.4.2. Tensions between approaches

Although “common tools” such as “encryption” and “data minimization” can “advance” both information privacy and information security, some approaches can cause “tensions” between these two areas (Kuner et al., 2017). For instance, “proposals to enhance cybersecurity by requiring identity verification, reducing online anonymity, and sharing potentially personal information about cyberattacks all pose risks for personal privacy” (Kuner et al., 2017).

#### 4.4.3. Focus on impacts for a wider range of stakeholders

Typically, information security risk assessment centres on impacts on the operator or system stakeholders. In contrast, information privacy risk assessment takes a much broader view, focusing on the impacts on individuals, groups of people, and wider society of potentially harmful activities.

## 5. Conclusion

---

<sup>47</sup> For further illustration, Brooks et al. (2017) provide the following example: “smart meters are the part of the system collecting the information and thereby creating the problems for individuals (e.g., loss of trust; chilling effect on ordinary behavior). An information security risk model would be unlikely to perceive this behavior of the smart meter as a “threat” since the activity is an authorized part of the functioning of the system itself.”

This report is the first in a series of four DARE UK PRIAM project reports, which together focus on working towards standardisation of privacy risk assessment for cross-domain access and re-usage of sensitive data for research utilising advanced analytics methods (AI/ML). In this report, we specifically concentrated on three key areas:

- First, we outlined **three driver use cases as exemplars of cross-domain linkage and analysis** related to public health research and integrated care. Two of these use cases centre on interdisciplinary, collaborative research projects re-using data from multiple sources in the areas of complex hospital discharge and multi-morbidity prevention. The third use case focuses on supporting cross-domain access and re-use of sensitive data for research purposes through a sub-national federated TRE ecosystem pilot.
- Second, we explored **emerging data usage patterns and data sharing needs in operational health data networks**, concentrating on TREs as facilitators of federated sharing and processing of data.
- Third, we examined **some different approaches to identifying, organising and using risk factors for privacy risk assessment** through a literature review, specifically focusing on the CNIL PIA, the Five Safes, Inria Privacy Risk Assessment Methodology, ISO/IEC 27005, NIST PRAM, Privacy Protection Goals and UKAN ADF.

Note that while our work is driven by use cases taking place in cross-council research networks (UKRI), the principles and learnings from our project research activities concerning cross-domain access and re-use of sensitive data for research purposes have wider relevance beyond the DARE UK programme. We now summarise some of the key points highlighted in this report:

## 5.1. Outlining the context for privacy risk assessment: data usage patterns in operational health data networks

Increased attention is being given to how advanced analytics methods (AI/ML) are being used to discover value in big datasets. These methods are driving new data processing patterns and forms of research collaborations underpinned by the federated sharing and processing of data — e.g., facilitated via next-generation TREs. The relationship between one or more TREs and the health system is important as it influences applicable governance, data flows, tools and benefits expected by stakeholders who have an interest in the system under analysis — all of which have implications for privacy concerns, expectations and associated risks.

**Through our use case analysis, we highlight some key points about data usage patterns in health systems:**

- Health systems are complex and evolving networks of people and service providers whose purpose is to improve and support the health and wellbeing of society.
- Data value flows within such complex and evolving networks of people and service providers are driven by the demands of operational, clinical and research needs.
- Researchers studying health and social care systems will have a wide range of research questions depending on the phenomena they are seeking to understand, and the data value chains of which they are a part.
- A service provider (for healthcare, social care) typically only has partial information about individuals based on the systems that they operate or have access to (e.g., Electronic Health Records tend not to consider other data such as the wider social determinants of health); and therefore, possesses an incomplete view of the complex data network.
- The idea of partial views into complex data networks is important because it shows: (i) there is no centre to the network — in that, there is not a single point of control, or one place where all data can be viewed; (ii) data linkage is established by data controllers who are responsible for views into the network; (iii)



views emerge within the network based on service and data value (e.g., a hospital, a curated disease specific dataset); and (iv) a TRE is a specific way of accessing a view on a network and therefore a view onto the whole dataset.

- The nature of data value changes within the complex data network, suggesting that value for research is distributed throughout the health system and cannot be easily integrated into a single place or TRE.

## We further outline some emerging data sharing needs:

- There is a need for greater availability and interoperability of quality data from service providers for research purposes.
- There is a need for TREs to be able to: manage ever-increasing variety, volumes, and velocity of data; offer greater support for a wider range of data analysis tools — e.g., for AI/ML; and be more connected with other TREs — e.g., to support advanced federated analysis and distributed machine learning.
- There is a need to consider how patients, service users, and members of public can have greater involvement with the co-design, testing, and evaluation of research concepts inception through to generated insights and tools (e.g., through interaction capabilities provided by TREs, such as interactive computational notebooks).

## 5.2. Towards a shared understanding of the Five Safes dimensions for cross-council data sharing

The Five Safes approach is a well-recognised and valuable tool for risk communication, discussion and decision-making about access and re-use of sensitive data. We also highlighted the special case of ‘Safe Return’ added by Hubbard et al. (2020). However, given the Five Safes is open to different interpretations, there is a need for both transparency and a shared understanding by all key stakeholders (e.g., data providers, data users, patients, TRE operators) about how the Five Safes (Plus One) are construed in the context of privacy risk assessment for cross-domain access and re-use of sensitive data for research purposes. Given the increasingly complex flows of multi-source data between people and organisations taking place through multiple platforms as part of wider data ecosystems of shared resources and services, federated data sharing and processing requires additional safeguards and controls, such as federated identity management and specific types of PETs (e.g., multi-party computation, differential privacy, homomorphic encryption). Therefore, to reflect the assets to be included in privacy risk assessment, we considered three requirements for federation that should be emphasised through the existing Five Safes dimensions:

- **Different types of research collaborations** should be highlighted as part of the **Safe Projects dimension**.
- **A wider-range of stakeholders** should be drawn attention to through the **Safe People dimension**.
- **The complexity of data flows** should be emphasised via the **Safe Settings dimension**.

Focus on the Five Safes approach is resumed in the D2 report where specific privacy risk factors are identified and categorised utilising the Five Safes through our engagement with legal, ethics, regulatory and information governance experts and practitioners as part of the DARE UK PRiAM Advisory Board.

## 5.3. Privacy risk modelling and simulation: mapping common types of risk factors

In this report, we undertook a conceptual mapping exercise by identifying common types of risk factors utilised by the ISO/IEC 27005 methodology for information security risk management — and comparing these concepts with selected privacy risk assessment methodologies. As part of this exercise, we identified several common types of risk factors used for modelling and simulating security and privacy risks, which include:



- **Assets** which can be ICT components, software, data, and socio-technical aspects such as, places and stakeholders specifically including data subjects;
- **Threats** that can affect Assets;
- **Consequences** of a Threat on an Asset, which is expressed as a **Risk** (i.e., the severity of the Consequence combined with its likelihood);
- Likelihood of Consequences is determined in part by **Vulnerabilities** of Assets, which may be reduced by **Controls** — such as (i) controls on data — i.e., those that transform the data itself, such as de-identification techniques; and (ii) environmental controls — i.e., those that change the environment in which the data is processed.

This conceptual mapping is utilised in the D3 report to model and simulate privacy risk for a given scenario using the System Security Modeller (SSM) — a tool used by the project which “automates much of an ISO 27005 risk assessment of socio-technical systems” modelling “both cyber-security and compliance” threats (Phillips et al., 2022). We therefore further reflected on the relationship between information security and information privacy risk assessment methodologies and outlined some key considerations:

- While concerns associated with information security and information privacy are conceptually related, each area offers a **distinct focus**.
- Information security risk assessment primarily focuses on risks arising from unauthorised activities — relating to loss of **confidentiality, availability and integrity**. Whereas privacy risk assessment focuses on risks in relation to **both unauthorised and authorised data-related activities**.
- **Privacy protection goals** from the field of privacy engineering (e.g., the Standard Data Protection Model) aim to address the ethical, legal, organisational and technical aspects of information privacy and data protection in relation to (un)authorised data processing activities in socio-technical systems (e.g., safe research collaborations). These goals therefore do not only concentrate on ensuring confidentiality, availability and integrity (‘data security’), but also extend their focus to the goals of **data minimisation, unlinkability, transparency and intervenability** (‘information privacy and data protection’).
- While some **common methods** can be used to both increase security and protection of privacy (e.g., ‘encryption’), some methods can also **cause tensions** in certain situations (e.g., ‘identity verification’).
- Typically, information security risk assessment centres on impacts to the operator or system stakeholders. In contrast, **information privacy risk assessment takes a much broader view, focusing on the impacts for individuals, groups of people and wider society** from potentially harmful activities.

## 6. Acknowledgements

---

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). The DARE UK programme, HDR UK and ADR UK are not responsible for any use that may be made of the information it contains.

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.

We acknowledge the feedback and discussions with medConfidential on the first version of this report released via the DARE UK website (v1.2, 15 July 2022), and an earlier draft of this updated report.

\* \* \*

**Please note that all views and opinions expressed in this report are those of the authors, and do not necessarily represent those named above.**

## 7. References

- ACS. (2018, November).** Privacy in Data Sharing: A Guide for Business and Government. An ACS White Paper. Retrieved from: <https://www.acs.org.au/insightsandpublications/reports-publications/privacy-in-data-sharing.html>.
- Ada Lovelace Institute and AI Council. (2021).** Exploring legal mechanisms for data stewardship. Retrieved from: <https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/>.
- Agencia Española de Protección de Datos (AEPD). (2019, October).** A Guide to Privacy by Design. Retrieved from: [https://www.aepd.es/sites/default/files/2019-12/guia-privacidad-desde-diseno\\_en.pdf](https://www.aepd.es/sites/default/files/2019-12/guia-privacidad-desde-diseno_en.pdf).
- Arbuckle, L., & Ritchie, F. (2019).** The Five Safes of Risk-Based Anonymization. *IEEE Security & Privacy*, 17(5), 84-89, Sept.-Oct. 2019. <https://doi.org/10.1109/MSEC.2019.2929282>.
- Article 29 Data Protection Working Party. (2017).** Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, wp248rev.01. As last Revised and Adopted on 4 October 2017. Available via: <https://ec.europa.eu/newsroom/article29/items/611236>.
- Australian Institute of Health and Welfare (AIHW). (2021).** The Five Safes framework. Australian Government. Last updated: 2021, September 9; v3.0. Retrieved from: <https://www.aihw.gov.au/about-our-data/data-governance/the-five-safes-framework>.
- Bajaj, S., Della-Libera, G., Dixon, B., Dusche, M., Hondo, M., Hur, M., Kaler, C. (Ed.), Lockhart, H., Maruyama, H., Nadalin, A. (Ed.), Nagaratnam, N., Nash, A., Prafullchandra, H., & Shewchuk, J. (2003, July 8).** Web Services Federation Language (WS-Federation). V1.0. Retrieved from: <https://specs.xmlsoap.org/ws/2003/07/secext/WS-Federation.pdf>.
- Bambauer, D.E. (2013).** Privacy versus security. *Journal of Criminal Law and Criminology*, 103(3), 667-684.
- Boniface M., Carmichael, L., Hall, W., Pickering, B., Stalla-Bourdillon, S., & Taylor, S. (2020).** A blueprint for a social data foundation: Accelerating trustworthy and collaborative data sharing for health and social care transformation. Web Science Institute (WSI) White Paper #4. Retrieved from: [www.socialdatafoundation.org/](http://www.socialdatafoundation.org/).
- Boniface M., Carmichael, L., Hall, W., Pickering, B., Stalla-Bourdillon, S., & Taylor, S. (2022).** The Social Data Foundation model: Facilitating health and social care transformation through datatrust services. *Data & Policy*, 4, E6. <https://doi.org/10.1017/dap.2022.1>.
- Bourne, P.E., Lorsch, J.R., & Green, E.D. (2015).** Perspective: Sustaining the big-data ecosystem. *Nature*, 527, S16-S17. <https://doi.org/10.1038/527S16a>.
- Brooks, S., Garcia, M., Lefkowitz, N., Lightman, S., & Nadeau, E. (2017, January).** An Introduction to Privacy Engineering and Risk Management in Federal Systems [NISTIR 8062]. National Institute of Standards and Technology (NIST); Internal Report 8062. <https://doi.org/10.6028/NIST.IR.8062>.
- Burton, P.R., Murtagh, M.J., Boyd, A., Williams, J.B., Dove, E.S., Wallace, S.E., Tass'e, A.-M., Little, J., Chisholm, R.L., Gaye, A., Hveem, K., Brookes, A.J., Goodwin, P., Fistein, J., Bobrow, M., & Knoppers, B.M. (2015, October).** Data Safe Havens in health research and healthcare. *Bioinformatics*, 31(20), 3241-3248. <https://doi.org/10.1093/bioinformatics/btv279>.
- Calo, R. (2011).** The Boundaries of Privacy Harm. *Indiana Law Journal*, 86(3). Available at SSRN: <https://ssrn.com/abstract=1641487>.
- Centre for Longitudinal Studies, UCL. (n.d.).** 1970 British Cohort Study. Retrieved from: <https://cls.ucl.ac.uk/cls-studies/1970-british-cohort-study/>.
- Chaterji, S., Koo, J., Li, N., Meyer, F., Grama, A., & Bagchi, S. (2019, January).** Federation in genomics pipelines: techniques and challenges. *Briefings in Bioinformatics*, 20(1), 235-244. <https://doi.org/10.1093/bib/bbx102>.
- Citron, D.K., & Solove, D.J. (2021, February 9).** Privacy Harms. GWU Legal Studies Research Paper No. 2021-11, George Washington University (GWU) Law School, Public Law Research Paper No. 2021-11, 102 Boston University Law Review 793 (2022), Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3782222>.
- Clinical Practice Research Datalink (CPRD). (n.d.).** Available at: <https://cprd.com/>.
- Commission nationale de l'informatique et des libertés (CNIL). (2018a, February).** Privacy Impact Assessment (PIA): Methodology. February 2018 edition. Retrieved from:

<https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf>.

**Commission nationale de l'informatique et des libertés (CNIL). (2018b, February).** Privacy Impact Assessment (PIA): Knowledge Bases. February 2018 edition. Retrieved from:

<https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-3-en-knowledgebases.pdf>.

**Conference of the Independent Data Protection Supervisory Authorities of the Federation and the Länder (Provider). (2020).** Standard Data Protection Model (SDM): A method for Data Protection advising and controlling on the basis of uniform protection goals.

Version 2.0b, Adopted by the 99. Conference of the Independent Data Protection Supervisory Authorities of the Federation and the Länder on the 17 April 2020. Publisher: AK Technik of the Independent Data Protection Supervisory Authorities of the Federation and the Länder; Editor: UAG „Standard Data Protection Model“ of the AK Technik of the Independent Data Protection Supervisory Authorities of the Federation and the Länder. Retrieved from:

[https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methodology\\_V2.0b.pdf](https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methodology_V2.0b.pdf).

**Curry, E. (2016).** The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In: Cavanillas, J., Curry, E., Wahlster, W. (eds) *New Horizons for a Data-Driven Economy*. Springer, Cham, pp. 29-37.

[https://doi.org/10.1007/978-3-319-21569-3\\_3](https://doi.org/10.1007/978-3-319-21569-3_3).

**Dayarathna, R. (2011).** Taxonomy for Information Privacy Metrics. *Journal of International Commercial Law and Technology*, 6(4), 194–206.

**De, S.J., & Le Métayer, D. (2016).** PRIAM: A Privacy Risk Analysis Methodology. Research Report, RR-8876, Inria - Research Centre Grenoble – Rhône-Alpes, fhal-01302541f. Retrieved from: <https://hal.inria.fr/hal-01302541/document>.

**Desai, T., Ritchie, F., & Welpton, R. (2016).** Five Safes: designing data access for research. Faculty of Business and Law, University of the West of England (UWE), Economics Working Paper Series 1601. Retrieved from: <https://www2.uwe.ac.uk/faculties/bbs/documents/1601.pdf>.

**Dodds, L., Szász, D. Keller, J.R, Snaith, B. and Duarte, S. (2020, April).** Designing sustainable data institutions. Open Data Institute (ODI) report. Contributions from Hardinges, J. and Tennison, J. Retrieved from:

<https://theodi.org/article/designing-sustainable-data-institutions-paper/>.

**Dorset Intelligence and Insight Service (DiiS). (2022).** A Guide to the Dorset Intelligence & Insight Service (DiiS). NHS Dorset Clinical Commissioning Group (CCG). Retrieved from: <https://www.dorsetccg.nhs.uk/wp-content/uploads/2022/01/Introducing-DiiS-v2.0-25012022.pdf>.

**Duckworth, C., & Boniface, M. (2022).** soton-hub/ai-codesign-notebook: First Release of COTADS Codesign notebook (v0.1). COdesigning Trustworthy Autonomous Diabetes Systems' ("COTADS") project. Available at: Zenodo. <https://doi.org/10.5281/zenodo.6376782>.

**Duckworth, C., Ayobi, A., Dylag, J., O'Kane, A., Marshall, P., Kumaran, A., Guy, M., & Boniface, M. (2022a).** ai-codesign-notebook. COdesigning Trustworthy Autonomous Diabetes Systems' ("COTADS") project. Retrieved from: <https://github.com/soton-hub/ai-codesign-notebook>.

**Duckworth, C., Guy, M. J., Kumaran, A., O'Kane, A. A., Ayobi, A., Chapman, A., Marshall, P., & Boniface, M. (2022b).** Explainable Machine Learning for Real-Time Hypoglycemia and Hyperglycemia Prediction and Personalized Control Recommendations. *Journal of Diabetes Science and Technology*. <https://doi.org/10.1177/19322968221103561>.

**Eder, J., & Shekhovtsov, V.A. (2021).** Data quality for federated medical data lakes. *International Journal of Web Information Systems*, 17(5), 407-426. <https://doi.org/10.1108/IJWIS-03-2021-0026>.

**Elliot, M., Mackey, E., & O'Hara, K. (2020).** The Anonymisation Decision-Making Framework: European Practitioners' Guide. 2<sup>nd</sup> Edition; Published in the UK in 2020 by UKAN, University of Manchester; UKAN Publications. Retrieved from: <https://ukanon.net/framework/>.

**European Commission, Directorate-General for Informatics. (2017).** New European interoperability framework: promoting seamless services and data flows for European public administrations. Publications Office. Retrieved from: <https://data.europa.eu/doi/10.2799/78681>.

**European Union Agency for Cybersecurity (ENISA). (2013, December 20).** Recommendations for a methodology of the assessment of severity of personal data breaches. Authors: Data Protection Authorities of Greece and Germany, Clara Galan Manso, ENISA,

Sławomir Górniak, ENISA. Retrieved from:  
<https://www.enisa.europa.eu/publications/dbn-severity>.

**Fleurence, R.L., Beal, A.C., Sheridan, S.E., Johnson, L.B. & Selby, J.V. (2014).** Patient-Powered Research Networks Aim to Improve Patient Care and Health Research. *Health Affairs*, 33(7). <https://doi.org/10.1377/hlthaff.2014.0113>.

**Gardner, T., (2022, January 11).** NHS winter pressures: Going home from hospital. The Health Foundation. Retrieved from: <https://www.health.org.uk/blogs/nhs-winter-pressure-going-home-from-hospital>.

**General Data Protection Regulation (GDPR).** Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). Retrieved from: <http://data.europa.eu/eli/reg/2016/679/oj>.

**Goldacre, B., & Morley, J. (2022, April).** Better, Broader, Safer: Using health data for research and analysis. A review commissioned by the Secretary of State for Health and Social Care. Department of Health and Social Care. Retrieved from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1067053/goldacre-review-using-health-data-for-research-and-analysis.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1067053/goldacre-review-using-health-data-for-research-and-analysis.pdf).

**Gross, D., & van Veen, E.-B. (2020).** Anonymised data and the rule of law. *European Data Protection Law Review (EDPL)*, 6(4), 498-508.

**Hansen, M., Jensen, M., & Rost, M. (2015).** Protection Goals for Privacy Engineering. 2015 IEEE Security and Privacy Workshops, pp. 159-166. <https://doi.org/10.1109/SPW.2015.13>.

**Harkness, F., Blodgett, J., Rijneveld, C., Waind, E., Amugi, M., & McDonald, F. (2022).** Building a trustworthy national data research infrastructure: A UK-wide public dialogue. Carried out by DARE UK & Kohlrabi Consulting; ADR UK, HDR UK, UKRI. Retrieved from: [https://dareuk.org.uk/wp-content/uploads/2022/05/DARE\\_UK\\_Building\\_a\\_Trustworthy\\_National\\_Data\\_Research\\_Infrastructure\\_Public\\_Dialogue\\_May-2022.pdf](https://dareuk.org.uk/wp-content/uploads/2022/05/DARE_UK_Building_a_Trustworthy_National_Data_Research_Infrastructure_Public_Dialogue_May-2022.pdf).

**Harris, M., Ferguson, L., & Luo, A. (2021).** Infrastructuring an organizational node for a federated research and data network: A case study from a sociotechnical perspective.

*Journal of Clinical and Translational Science*, 5(1), E186. <https://doi.org/10.1017/cts.2021.846>.

**Health Data Research UK (HDR UK). (2021a).** Data Insights in a Pandemic: Annual Review 2020/2021. Retrieved from: [https://www.hdruk.ac.uk/wp-content/uploads/2021/08/HDRUK\\_AnnualReview\\_2021-compressed.pdf](https://www.hdruk.ac.uk/wp-content/uploads/2021/08/HDRUK_AnnualReview_2021-compressed.pdf).

**Health Data Research UK (HDR UK). (2021b, September 9).** Innovation Gateway Open Door: Trusted Research Environments. Retrieved from: <https://www.hdruk.ac.uk/wp-content/uploads/2021/09/TRE-Open-Door-September.pdf>.

**Health Data Research UK (HDR UK). (2021c).** "What is a TRE?" Retrieved from: [https://www.hdruk.ac.uk/wp-content/uploads/2021/09/HDRUK\\_TRE-One-Pager.pdf](https://www.hdruk.ac.uk/wp-content/uploads/2021/09/HDRUK_TRE-One-Pager.pdf).

**Health Data Research UK (HDR UK). (2022, April 8).** Health Data Research UK's response to the publication of the Goldacre Review. HDR UK, News. Retrieved from: <https://www.hdruk.ac.uk/news/health-data-research-uks-response-to-the-publication-of-the-goldacre-review/>.

**Health Data Research UK (HDR UK). (n.d.).** What is a TRE? Retrieved from: [https://www.hdruk.ac.uk/wp-content/uploads/2021/09/HDRUK\\_TRE-One-Pager.pdf](https://www.hdruk.ac.uk/wp-content/uploads/2021/09/HDRUK_TRE-One-Pager.pdf).

**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>.

**Information & Privacy Commissioner of Ontario, Canada. (2010).** Privacy Risk Management: Building privacy protection into a Risk Management Framework to ensure that privacy risks are managed, by default. Retrieved from: <https://www.ipc.on.ca/wp-content/uploads/2010/04/privacy-risk-management-building-privacy-protection-into-a-risk-management-framework-to-ensure-that-privacy-risks-are-managed.pdf>.

**Information Commissioner's Office (ICO). (2014).** Conducting Privacy Impact Assessments: Code of Practice. Retrieved from: <https://www.pdpjournals.com/docs/88317.pdf>.

**Information Commissioner's Office (ICO). (2021, April 22).** The Information Commissioner's position paper on the UK Government's proposal for a trusted digital identity system. V2.0. Retrieved from: <https://ico.org.uk/media/about-the->



[ico/documents/2619686/ico-digital-identity-position-paper-20210422.pdf](https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/how-do-we-do-a-dpia/#how9).

**Information Commissioner's Office (ICO). (n.d.).** How do we do a DPIA? Online Guide to the General Data Protection Regulation (GDPR). Retrieved from: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/how-do-we-do-a-dpia/#how9>.

**Integrated Research Application System (IRAS).** Retrieved from: <https://www.myresearchproject.org.uk/>.

**International Association of Privacy Professionals (IAPP). (n.d.).** What is privacy. Retrieved from: <https://iapp.org/about/what-is-privacy/>.

**International Risk Governance Center. (2017).** Introduction to the IRGC Risk Governance Framework. Revised version. Lausanne: EPFL International Risk Governance Center. Retrieved from: <https://irgc.org/wp-content/uploads/2018/09/IRGC.-2017.-An-introduction-to-the-IRGC-Risk-Governance-Framework.-Revised-version..pdf>.

**International Risk Governance Council. (2010).** The Emergence of Risks: Contributing Factors. IRGC Report, Geneva. Retrieved from: [https://irgc.org/wp-content/uploads/2018/09/irgc\\_ER\\_final\\_07jan\\_web.pdf](https://irgc.org/wp-content/uploads/2018/09/irgc_ER_final_07jan_web.pdf).

**ISO 27000 ISO/IEC 27000:2018.** Information technology — Security techniques — Information security management systems — Overview and vocabulary. <https://www.iso.org/standard/73906.html>.

**ISO 27005 ISO/IEC 27005:2018.** Information technology — Security techniques — Information security risk management. <https://www.iso.org/standard/75281.html>

**Jhutti, H., & Bloomfield, C. (2022, March 2).** Collaboration across the system to increase the privacy-protection and speed of life-saving research. NHS Blog. Retrieved from: <https://www.england.nhs.uk/blog/collaboration-across-the-system-to-increase-the-privacy-protection-and-speed-of-life-saving-research/>.

**Kavianpour, S., Sutherland, J., Mansouri-Benssasi, E., Coull, N., & Jefferson, E. (forthcoming/in press).** A Review of Trusted Research Environments to Support Next Generation Capabilities based on Interview Analysis. Journal of Medical Internet Research. 30/05/2022:33720 (forthcoming/in press). Retrieved from: <https://preprints.jmir.org/preprint/33720>.

**Kuner, C., Cate, F.H., Millard, C., Svantesson, D.J.B., & Lynskey, O. (2015, May).** Risk management in data protection. International Data Privacy Law, 5(2), 95–98. <https://doi.org/10.1093/idpl/ipv005>.

**Kuner, C., Svantesson, D.J.B., Cate, F.H., Lynskey, O., & Millard, C. (2017, May).** The rise of cybersecurity and its impact on data protection. International Data Privacy Law, 7(2), 73–75. <https://doi.org/10.1093/idpl/ix009>.

**Laurie, G., Ainsworth, J., Cunningham, J., Dobbs, C., Jones, K.H., Kalra, D., Lea, N., & Sethi, N. (2015).** On moving targets and magic bullets: Can the UK lead the way with responsible data linkage for health research?" International Journal of Medical Informatics, 84, (11): 933–940. <https://doi.org/10.1016/j.ijmedinf.2015.08.011>.

**Laurie, G., Jones, K.H., Stevens, L., & Dobbs, C. (2014, June 30).** A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data. Prepared for the Nuffield Council on Bioethics Working Party on Biological and Health Data and the Expert Advisory Group on Data Access. The Farr Institute Scotland & the Mason Institute for Medicine, Life Sciences & the Law. Retrieved from: <https://www.nuffieldbioethics.org/wp-content/uploads/A-Review-of-Evidence-Relating-to-Harms-Resulting-from-Uses-of-Health-and-Biomedical-Data-FINAL.pdf>.

**Madden, S., Pollard, C., & Donohoe, T. (2021, November 10).** Trusted Research Environments (TREs): Our ambition and plans. NHSX. Retrieved from: <https://www.england.nhs.uk/aac/wp-content/uploads/sites/50/2022/03/Trusted-research-environments-TRE-development-update.pdf>.

**McCallister, E., Grance, T., & Scarfone, K. (2010, April).** Guide to Protecting the Confidentiality of Personally Identifiable Information (PII): Recommendations of the National Institute of Standards and Technology. NIST Special Publication 800-122. <https://doi.org/10.6028/NIST.SP.800-122>.

**Medicine — University of Southampton. (2022).** Research project: Multidisciplinary Ecosystem to study Lifecourse Determinants and Prevention of Early-onset Burdensome Multimorbidity (MELD-B). Retrieved from: [https://www.southampton.ac.uk/medicine/academic\\_units/projects/meld-b.page](https://www.southampton.ac.uk/medicine/academic_units/projects/meld-b.page).

**National Health Service (NHS). (2019, February 18).** Being discharged from hospital. Retrieved from: <https://www.nhs.uk/nhs-services/hospitals/going-into-hospital/being-discharged-from-hospital/>.



**National Institute for Health and Care Research (NIHR). (2022, March).** Artificial Intelligence for Multiple Long-Term Conditions (AIM) - Research Specification. Published: 17 April 2020; Version 1.2 – March 2022. Retrieved from: <https://www.nihr.ac.uk/documents/nihr-artificial-intelligence-for-multiple-long-term-conditions-aim-clusters-call-research-specification-finalised/24646>.

**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>.

**National Standards Institute for Technology (NIST), U.S. Department of Commerce. (2019, February).** NIST Privacy Framework: Catalog of Problematic Data Actions and Problems. Available to download via: <https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/resources>.

**National Standards Institute for Technology (NIST), U.S. Department of Commerce. (2020a, April 8).** Privacy Engineering Program: Resources. Retrieved from: <https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/resources>.

**National Standards Institute for Technology (NIST), U.S. Department of Commerce. (2020b, January 16).** NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management. Version 1.0. Retrieved from: [https://www.nist.gov/system/files/documents/2020/01/16/NIST%20Privacy%20Framework\\_V1.0.pdf](https://www.nist.gov/system/files/documents/2020/01/16/NIST%20Privacy%20Framework_V1.0.pdf).

**NHS Digital. (2022).** Hospital Episode Statistics (HES). Last edited: 2022, June 6. Retrieved from: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics>.

**NIHR ARC Wessex. (n.d.).** About us. Retrieved from: <https://www.arc-wx.nihr.ac.uk/about-us/>.

**Nissenbaum, H. (2004).** Privacy as contextual integrity. *Washington Law Review*, 79(1), 119-158.

**Nokkala, T. A., & Dahlberg, T. (2019).** Empowering citizens through data interoperability - data federation applied to consumer-centric healthcare. *Finnish Journal of EHealth and EWellfare*, 11(4), 246-257. <https://doi.org/10.23996/fjhw.82599>.

**Office of the Australian Information Commissioner. (n.d.)** What is privacy? Australian Government. Retrieved

from: <https://www.oaic.gov.au/privacy/your-privacy-rights/what-is-privacy>.

**Oliveira, M.L.S., Barros Lima, G.d.F., & Farias Lóscio, B. (2019).** Investigations into Data Ecosystems: a systematic mapping study. *Knowledge and Information Systems*, 61, 589-630. <https://doi.org/10.1007/s10115-018-1323-6>.

**Organisation for Economic Co-operation and Development (OECD). (2019).** 4: Risks and Challenges of Data Access and Sharing in: *OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, Paris, <https://doi.org/10.1787/276aaca8-en>.

**Our Future Health. (2022, June 30).** Our Future Health opens consultation on trusted research environment accreditation process. Retrieved from: <https://ourfuturehealth.org.uk/our-future-health-opens-consultation-on-trusted-research-environment-accreditation-process/>.

**Peeters, S. (2013).** Beyond distributed and decentralized: what is a federated network? Unlike Us #3, *Social Media: Design or Decline*; Institute of Network Cultures, Amsterdam University of Applied Sciences. Retrieved from: <https://networkcultures.org/unlikeus/resources/articles/what-is-a-federated-network/>.

**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>.

**Price, W.N., & Cohen, I.G. (2019).** Privacy in the age of medical big data. *Nature Medicine*, 25, 37-43. <https://doi.org/10.1038/s41591-018-0272-7>.

**Ritchie, F. (2017).** The 'Five Safes': a framework for planning, designing and evaluating data access solutions. *Data for Policy 2017: Government by Algorithm? (Data for Policy)*, London. Zenodo. <https://doi.org/10.5281/zenodo.897821>.

**Secure Anonymised Information Linkage (SAIL) Databank. (n.d.).** About us. Retrieved from: <https://saildatabank.com/about-us/>.

**Sharon, T., & Lucivero, F. (2019).** Introduction to the Special Theme: The expansion of the health data ecosystem – Rethinking data ethics and governance. *Big Data & Society*. <https://doi.org/10.1177/2053951719852969>.

**Shaw, S., & Barrett, G. (2006).** Research Governance: Regulating Risk and Reducing Harm? *Journal of the Royal Society of Medicine*, 99(1), 14–19.

<https://doi.org/10.1177/014107680609900109>.

**Shirey, R. (2007, August).** Internet Security Glossary, Version 2. Request for Comments: 4949 (RFC 4949); Network Working Group; the IETF Trust. Retrieved from: <https://datatracker.ietf.org/doc/html/rfc4949>.

**Singhal, A., Winograd, T., & Scarfone, K.A. (2007).** Guide to Secure Web Services. (National Institute of Standards and Technology, Gaithersburg, MD), NIST. Special Publication (SP) 800-95.

<https://doi.org/10.6028/NIST.SP.800-95>.

**Skinner, G., Han, S., & Chang, E. (2006).** An information privacy taxonomy for collaborative environments. *Information Management & Computer Security*, 14 (4), 382-394. <https://doi.org/10.1108/09685220610690835>.

**Smart Dubai and Nesta. (2020, March).** Data Sharing Toolkit: Approaches, guidance and resources to unlock the value of data. Retrieved from: [https://media.nesta.org.uk/documents/Data\\_sharing\\_toolkit.pdf](https://media.nesta.org.uk/documents/Data_sharing_toolkit.pdf).

**Solove, D. J. (2006).** A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3), 477–564. <https://doi.org/10.2307/40041279>.

**Solove, D.J. (2002).** Conceptualizing Privacy. *California Law Review*, 90(4), 1087–1155. <https://doi.org/10.2307/3481326>.

**Stalla-Bourdillon, S., Rossi, A., & Zafir-Fortuna, G. (2019a, December).** Data Protection by Process: How to Operationalize Data Protection by Design for Machine Learning. V1.0. Immuta & Future of Privacy Forum White Paper. Retrieved from: [https://fpf.org/wp-content/uploads/2019/12/WhitePaper\\_DataProtectionByProcess.pdf](https://fpf.org/wp-content/uploads/2019/12/WhitePaper_DataProtectionByProcess.pdf).

**Stalla-Bourdillon, S., Wintour, A. & Carmichael, L. (2019b, December).** Building Trust Through Data Foundations: A Call for a Data Governance Model to Support Trustworthy Data Sharing. Web Science Institute (WSI) White Paper #2. Retrieved from: [https://eprints.soton.ac.uk/443715/1/White\\_Paper\\_2.pdf](https://eprints.soton.ac.uk/443715/1/White_Paper_2.pdf).

**Stalla-Bourdillon, S. (2019c).** Data protection by design and data analytics: can we have both? *Privacy & Data Protection*, 19(5), 8-10. Available at: <https://www.pdpjournals.com/back-issues-privacy-and->

[data-protection](#); also <https://www.immuta.com/downloads/data-protection-and-data-analytics/>.

**The Alan Turing Institute. (2021).** AI for multiple long-term conditions: Research Support Facility. Retrieved from: <https://www.turing.ac.uk/research/research-projects/ai-multiple-long-term-conditions-research-support-facility>.

**The Information Governance Review. (2013, March).** Information: To share or not to share? Retrieved from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/192572/2900774\\_InfoGovernance\\_accv2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf).

**The Royal Society. (2019, March).** Protecting privacy in practice: The current use, development and limits of Privacy Enhancing Technologies in data analysis. ISBN: 978-1-78252-390-1. Retrieved from: <https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/>.

**The Scottish Government. (2015).** The Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics. Retrieved from: <https://www.gov.scot/publications/charter-safe-havensscotland-handling-unconsented-data-national-health-servicepatient-records-support-research-statistics/pages/3/>.

**UK Data Service, SecureLab (2022).** What is the Five Safes framework? Retrieved from: <https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/#:~:text=The%20Five%20Safes%20framework%20is,data%20providers%20in%20the%202010s>.

**UK Health Data Research Alliance, & NHSX. (2021).** Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems (1.0). Zenodo. <https://doi.org/10.5281/zenodo.5767586>.

**University of Aberdeen. (n.d.).** Aberdeen Birth Cohorts: Children of the 1950s—For researchers. Retrieved from: <https://www.abdn.ac.uk/birth-cohorts/1950s/for-researchers/>.

**Wessex Academic Health Science Network (AHSN). (2022).** Business Plan 2022-23. NHS. Retrieved from: <https://wessexahsn.org.uk/img/publications/Wessex%20Academic%20Health%20Science%20Network%20Business%20Plan%2022-23.pdf>.

**Woodall, R. (2021, March 15).** Data Federations: Digital Collaboration Without Data Sharing. Open Data Institute Report (ODI); Etic Lab. Retrieved from:

<https://theodi.org/article/etic-lab-developing-a-data-federation-model/>.

**World Economic Forum (WEF). (2020, July).** Sharing Sensitive Health Data in a Federated Data Consortium Model: An Eight-Step Guide. Insight Report. Retrieved from:

[https://www3.weforum.org/docs/WEF\\_Sharing\\_Sensitive\\_Health\\_Data\\_2020.pdf](https://www3.weforum.org/docs/WEF_Sharing_Sensitive_Health_Data_2020.pdf).

**World Health Organization (WHO). (2007).** Everybody's business: strengthening health systems to improve health outcomes: WHO's framework for action. WHO Document Production Services, Geneva, Switzerland. Retrieved from:

[https://apps.who.int/iris/bitstream/handle/10665/43918/9789241596077\\_eng.pdf](https://apps.who.int/iris/bitstream/handle/10665/43918/9789241596077_eng.pdf).

**Wottrich, V.M., van Reijmersdal, E.A., & Smit, E.G. (2018).** The privacy trade-off for mobile app downloads: The roles of app value, intrusiveness, and privacy concerns. *Decision Support Systems*, 106, 44-52.

<https://doi.org/10.1016/j.dss.2017.12.003>.

**Wright, D., & Raab, C. (2014).** Privacy principles, risks and harms. *International Review of Law, Computers & Technology*, 28(3), 277-298.

<https://doi.org/10.1080/13600869.2014.913874>.

**Wu, P.F., Vitak, J., & Zimmer, M.T. (2020).** A contextual approach to information privacy research. *Journal of the Association for Information Science and Technology*, 71(4), 485-490. <https://doi.org/10.1002/asi.24232>.

## 8. Glossary

For the purposes of the DARE UK PRIAM project, we present the following definitions for key terms:

Key term	Definition
<b>Complex Discharge</b>	A patient who requires “more specialised care after leaving hospital” — as defined by NHS (2019).
<b>Controller</b>	“the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data [...]” — as defined by Article 4(7) of the GDPR.
<b>Data Aggregation</b>	Sharing of data between service providers that is then used as a resource to deliver services (e.g., shared care records, public health management).
<b>Data Ecosystem</b>	“socio-technical complex networks in which actors interact and collaborate with each other to find, archive, publish, consume, or reuse data as well as to foster innovation, create value, and support new businesses” — as defined by Oliveira et al. (2019).
<b>Data Flow</b>	“The movement or transfer of data through a system, describing who has responsibility for and access to them, and the contexts in which it is held” — as defined by the UKAN ADF (Elliot et al., 2020).
<b>Data Subject</b>	An “identified or identifiable natural person” to whom personal data relates — as defined by Article 4(1) of the GDPR.
<b>Data Value Chain</b>	“the information flow within a big data system as a series of steps needed to generate value and useful insights from data” — as defined by Curry (2016).
<b>Federated Research Network</b>	“collaborations among partners who, through coordination at an overarching network level, bring together, share, and optimize resources and services in order to enable research that exploits this new data-intensive and connected scientific environment” — as defined by Harris et al. (2021).
<b>Five Safes</b>	Well-known best practice principles for safe research — focused on five key dimensions: ‘Safe Projects’, ‘Safe People’, ‘Safe Settings’, ‘Safe Data’ and ‘Safe Outputs’ — originally devised for the Office for National Statistics (Desai et al., 2016).
<b>Five Safes Plus One</b>	Refers to the addition of ‘Safe Return’ to the Five Safes by the HDRA UK (Hubbard et al., 2020).
<b>Health System</b>	“consists of all organizations, people and actions whose <i>primary intent</i> is to promote, restore or maintain health. This includes efforts to influence determinants of health as well as more direct health-improving activities” — as defined by the World Health Organization (WHO, 2007).
<b>Intervenability</b>	“the data subject’s capacity for intervention and control in the processing” — as defined by Agencia Española de Protección de Datos (AEPD, 2019).
<b>Multimorbidity</b>	The co-occurrence of two or more long-term health conditions.
<b>Personal Data</b>	“any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person” — as defined by Article 4(1) of the GDPR.
<b>Privacy Protection Goals</b>	“a common scheme for addressing the legal, technical, economic, and societal dimensions of privacy and data protection in complex IT systems” — as defined by Hansen et al. (2015).
<b>Privacy Risk Assessment</b>	“A privacy risk management sub-process for identifying and evaluating specific privacy risks” — as defined by NIST Privacy Framework (NIST, 2020).
<b>Processing</b>	“any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction” — as defined by Article 4(2) of the GDPR.

<b>Research Collaboration</b>	Communities of people and organisations, often across different sectors and disciplines, working together for one or more shared goals, who contribute to research activities by undertaking or otherwise informing them. They may be <i>ad hoc</i> , short-lived collaborations — such as, for specific research projects, or long-term formal resources — such as, those provided by professional bodies through federated research networks.
<b>Risk Communication</b>	“the process of exchanging or sharing risk-related data, information and knowledge between and among different groups such as scientists, regulators, industry, consumers or the general public” — as defined by the International Risk Governance Center (2017).
<b>Risk Factor</b>	“A characteristic used in a risk model as an input to determining the level of risk in a risk assessment” — as defined by NIST (2012).
<b>Service Integration</b>	Connectivity between services to create business processes and care pathways (e.g., referral and discharge).
<b>Trusted Research Environment (TRE)</b>	Safe and secure platform supporting workspaces for approved research that can be remotely accessed by authorised researchers and data analysts (also referred to as ‘data safe havens’).

## 9. Appendix: summary of changes to v1.2 report

---

The first publicly released version of the D1 report (v1.2) was made available via the [DARE UK PRIAM project page](#) on the DARE UK website (15 July 2022). In the first version, we decided to reformulate the Five Safes, considering the requirements for federation, with new definitions that accurately reflect the assets to be included in privacy risk assessment. However, following feedback from medConfidential, to avoid any confusion that could be caused through renaming the Five Safes, the existing Five Safes are now qualified through description instead. We have subsequently revised the original report in response to this feedback as well as to provide further updates as the project completes.

The following key changes have been made to the report:

- **Removal of text and other references related to the renaming of the Five Safes** (i.e., ‘Enhanced Five Safes Plus One’) throughout the report.
- **Section 4.1 modified. New text** provided in section 4.1.3 where the **existing Five Safes dimensions are qualified through description** (rather than renamed) to take into consideration requirements for federation. **Links to D2 and D4 reports highlighted** in section 4.1.2 on risk communication. The original Five Safes questions (Table 3) and some example interpretations of the Five Safes dimensions (Table 4) are now in **separate tables**.
- **Deletion of Matrix** (Mapping Risk Factor Types to Scope of Risk Assessment) and associated text (previously section 5.5 in v1.2 of report).
- Report has been **reformatted using the DARE UK template** (e.g., removal of Parts A and B heading sections).
- **Document details** have been updated.
- **Restructuring and re-wording of Executive Summary** to reflect other modifications to report.
- **Introduction updated** — including more detail provided about the project (section 1.1).
- Modification of **section 2 heading** to emphasise focus on emerging data usage patterns in relation to TREs
- **Additional text added to section 2.4.5** to emphasise focus on privacy risk factors.
- **Additional text added to section 3** introductory paragraphs and section 3.2 on TREs.
- **Previous sections 4 and 5 brought under a new section heading:** “Examining different approaches to identifying, organising and using risk factors for privacy risk assessment” to emphasise a key area focus of the report.
- **Slight re-wording of sections 4.2.3 and 4.4** for improved clarity.
- **Conclusion modified** to reflect changes to report; and references **to next steps in relation to other ongoing project work have been removed** as the project has now completed.
- New **acknowledgements section** added.
- New **references added** related to additional text.