University of Southampton | WARWICK THE UNIVERSITY OF WARWICK | PRIVITAR

PRiAM
PRIVACY RISK ASSESSMENT METHODOLOGY

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

Deliverable 1 Report | DARE UK PRiAM Project

v1.1 | July 2022

DARE UK (Data and Analytics Research Environments UK) Programme

UK Research and Innovation | HDRUK Health Data Research UK | ADRUK

Document Details

Date	12/07/22
Deliverable lead	University of Southampton
Version	1.1
Authors	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)
Contact	m.j.boniface@soton.ac.uk
Dissemination level	Public

Funding statement

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

Disclaimer

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

Publication Acknowledgement

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

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.

Abstract

The aim of the DARE UK PRiAM project is to lay the foundations for a standard privacy risk assessment framework that can describe and automatically assess privacy risk for safe research collaborations. This report describes privacy requirements and use cases for collaborative research, taking into consideration emerging data usage patterns and needs of advanced analytics in cross council research networks (related to Work Package 1 project activities). This report is the first in a series of three reports to be delivered by the project.

This report is divided into two main parts. Part A sets out the context for our privacy risk assessment framework. We first describe our three real-world advanced analytics use cases related to public health research and integrated care, specifically selected to be used as exemplars of data linkage to drive the identification of factors and situations causing and affecting privacy risks in cross council research networks.

Part B focuses on an initial conceptualisation of risk factors. As a key part of our methodology, we first examine the Fives Safes (Plus One)—'Safe Projects', 'Safe People', 'Safe Settings', 'Safe Data' and 'Safe Outputs' (plus 'Safe Return)—as a well-established and popular approach for structuring transdisciplinary discussion and decision-making around access to sensitive data, and its important role in risk communication. We further present an 'Enhanced Five Safes Plus One' suited to the emerging data usage patterns and needs of advanced analytics in cross council research networks. We then we propose an initial privacy risk assessment approach, combining ISO/IEC 27005 methodology for information security risk management and other key sources related to privacy risk assessment, to determine the factor types needed to identify privacy risks for safe research collaborations. This initial privacy risk assessment approach will be further refined and developed in WP2 and WP3.

Executive Summary

This report describes privacy requirements and use cases for safe collaborative research, taking into consideration emerging data usage patterns and needs of advanced analytics in cross council research networks. The key findings highlighted by this report will help to lay the foundations for a standard privacy risk assessment framework, which can describe and automatically assess privacy risk for safe research collaborations, as part of ongoing DARE UK PRiAM project work.

Motivation

UK Research and Innovation (UKRI) cross council research revolving around advanced analytics—artificial intelligence/machine learning—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.

Organisations responsible for carrying out and facilitating such research activities must ensure these remain **trustworthy, safe, secure and useful with reasonable and acceptable levels of privacy protection in place** so that individuals, groups, and wider society are not put at risk of undue harm (e.g., from re-identification, surveillance).

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.**

Overview of the DARE UK PRiAM Project

A DARE UK Sprint Exemplar Project. The 'Privacy Risk Assessment Methodology' project ("DARE UK PRiAM project") is one of nine projects funded by 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 will complete in August 2022. The DARE UK PRiAM project involves three partner organisations—University of Southampton, University of Warwick and Privitar Ltd—and brings 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.

The project aim is to lay the foundations for a standard privacy risk assessment framework that can describe and automatically assess privacy risk for safe research collaborations in cross council research networks.

The project objectives are to:

a) Define use cases and data patterns for advanced analytics;
b) Identify privacy risk factors;
c) Define a risk tier classification framework;
d) Assess privacy risks for use cases (related to public health research and integrated care) using cyber security risk modelling and simulation; and
e) Develop, evaluate and disseminate the framework and lessons learnt through engagement with experts and the public.

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

- WP1 "Use Cases, Evaluation & Stakeholder Engagement" will analyse use cases, requirements, conduct evaluation and capture/disseminate lessons learnt to maximise impact.
- WP2 "Privacy Risk Framework Specification" will identify privacy risk factors and develop the risk tier classification framework.
- WP3 "Privacy Risk Modelling and Simulation" will model risk factors and assess use cases using the ISO/IEC 27005 information security risk management methodology.

Outlining the context for privacy risk assessment

The first half of this report (Part A) sets out the context for our privacy risk assessment framework. We first describe our three real-world advanced analytics use cases related to public health research and integrated care (see below), specifically selected to be used as exemplars of data linkage to drive the identification of factors and situations causing and affecting privacy risks in cross council research networks; which will serve as validation cases for ongoing project work. We then explore emerging data usage patterns and needs of advanced analytics in cross council research networks in relation to these use cases.

Our three driver use cases

Use Case A

Complex Hospital Discharge:

PROCED Project

Project aim. Focuses on how to: (i) proactively model patient discharge risks (e.g., elongated length of stay, readmission) and expected departure points; and (ii) schedule and optimise provision of ongoing community care services.

Data linkage. 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.

Use Case B

Multi-morbidity Prevention:

MELD-B Project

Project aim. Centres on how to: safely deliver an artificial intelligence (AI) enhanced epidemiological analytic system in which optimal lifecourse time points and targets for prevention of early-onset, burdensome multimorbidity are identified through analysis of birth cohorts and electronic health and care records.

Data linkage. Requires both individual linking and federated learning between longitudinal birth cohort and routine data sets over the lifecourse.

Use Case C

Wessex Federated Trusted Research Environment (TREs)

Project aim. 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.

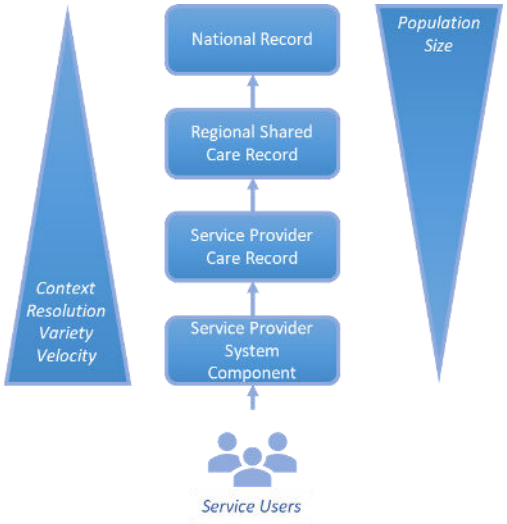
Data linkage. Necessitates data linkage across clinical care, social care, mental health, and other public administration services.

As a **brief summary**, these three driver use cases: (i) provide examples of multi-disciplinary research collaborations; (ii) demonstrate how research projects related to public health and integrated care, involving advanced analytics, require a considerable number of connected, multi-stakeholder data sources—and are increasingly federated in nature; and (iii) emerge and are shaped as part of wider data ecosystems within health systems.

Definition of 'research collaboration'. For the purposes of the DARE UK PRiAM project, we define a 'research collaboration' 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 through federated research networks."

Data usage patterns in operational health data networks

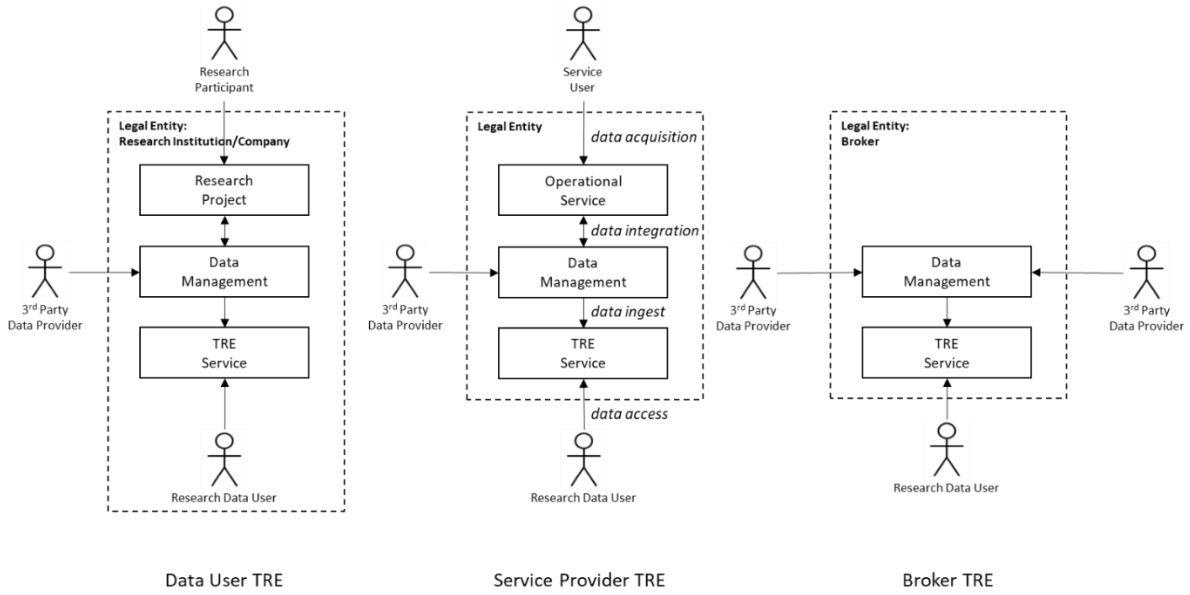
From the analysis of the three driver use cases, it is clear that data usage patterns for research collaborations related to a TRE, or otherwise federation of TREs, should be considered in the context of the system they are established to study. The relationship between one or more TREs and the health 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.**



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 a partial 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**; (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**.
- **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**.
- **Three typical TRE deployment scenarios 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 and Table below). 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**.

TREs in health systems



THREE TYPICAL TRE DEPLOYMENT SCENARIOS			
	Data User TRE	**Service Provider TRE**	**Broker TRE**
Deployed at:	A research institution/company	A service provider	A legal entity operating a data marketplace
Data are ingested from:	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
Access to TRE:	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
Needed for situations e.g., where:	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

Some emerging data sharing needs:

- A need for **greater availability and interoperability of quality data from service providers** for research purposes.
- A need for **TREs to be able to manage ever increasing types, volumes and velocity of data** and offer greater support for a **wider range of data analysis tools** (e.g., for AI/ML), and to be **more connected with other TREs** (e.g., to support advanced federated analysis, distributed machine learning).
- A need to 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 (e.g., through interaction capabilities provided by TREs such as interactive computational notebooks).

Laying foundations for a standard privacy risk assessment

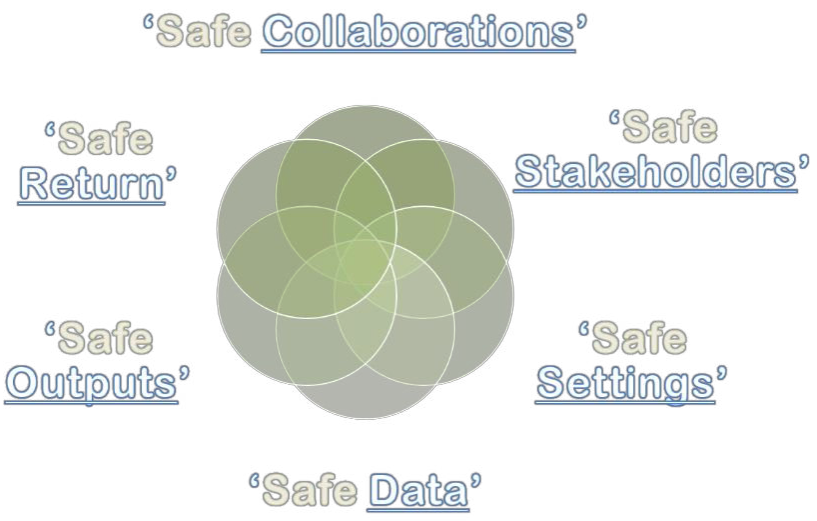
The second half of the report (Part B) helps towards laying the foundations for a standard privacy risk assessment through an initial conceptualisation of risk factors related to the emerging patterns and needs of research collaborations. Privacy is a nebulous concept—holding various meanings for people as well as for different stakeholder groups and disciplines. 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 it is being considered and can develop and change over time. Therefore, as a key part of our methodology, we first examine the Fives Safes—'Safe Projects', 'Safe People', 'Safe Settings', 'Safe Data' and 'Safe Outputs'—as a well-established and popular approach for **structuring transdisciplinary discussion and decision-making around access to sensitive data, and its important role in risk communication.**

We then we propose an **initial privacy risk assessment approach**, combining ISO/IEC 27005 methodology for information security risk management and other key sources related to privacy risk assessment, to determine the factor types needed to identify privacy risks for safe research collaborations. This initial privacy risk assessment approach will be further refined and developed in WP2 and WP3.

Effective risk communication: our approach to enhancing the 'Five Safes Plus One' to better align with emerging data patterns and needs for safe collaborative research

The Five Safes were first devised in 2003 for the Office for National Statistics—and are typically used a **best practice principles for safe research** in relation to a single data access facility or platform. The UK Health Data Research Alliance recently added 'Safe Return' (i.e., the 'Plus One'). From our use case analysis, a key data usage pattern emerging is that of the federated research network—a type of research collaboration involving multiple flows of data between an ecosystem of TREs providing shared resources and services for approved research purposes. As such, **the increasingly federated nature of research collaborations is disrupting the particular assumptions and context on which best practice principles for safe research are based**, such as the original Five Safes—and changes the risk factors associated with each of these dimensions.

Our 'Enhanced Five Safes Plus One'



We present our 'Enhanced Five Safes Plus One' principles[1] as follows:

'Safe Collaborations'. The use of resources and services as part of a research collaboration (involving two or more organisations) is lawful, ethical and aligned with stakeholder expectations. The research collaboration is for public benefit.

'Safe Stakeholders'. All people and organisations with responsibility for, accesses and utilises resources and services as part of a research collaboration have 'the skills, knowledge and incentives to act in accordance with required standards of behaviour'.

'Safe Settings'. All data processed as part of a research collaboration takes place within one or more specified environments; all environments and data flows in-between have effective and appropriate privacy and security controls.

'Safe Data'. All data processed as part of a research collaboration are reviewed to ensure adequate standards of data quality, data completeness and data richness. Data are treated appropriately and effectively to minimise privacy risks to individuals, organisations, groups of people and wider society.

'Safe Outputs'. Insights generated from a research collaboration will undergo appropriate checks to ensure any residual risks remain very low to individuals, organisations, groups of people and wider society.

'Safe Return'. The re-combination of outputs from a research collaboration with other data at the 'clinical setting that originated the data' can only take place where permitted and consented by the data subjects concerned.

We have therefore enhanced the Five Safes Plus One by:

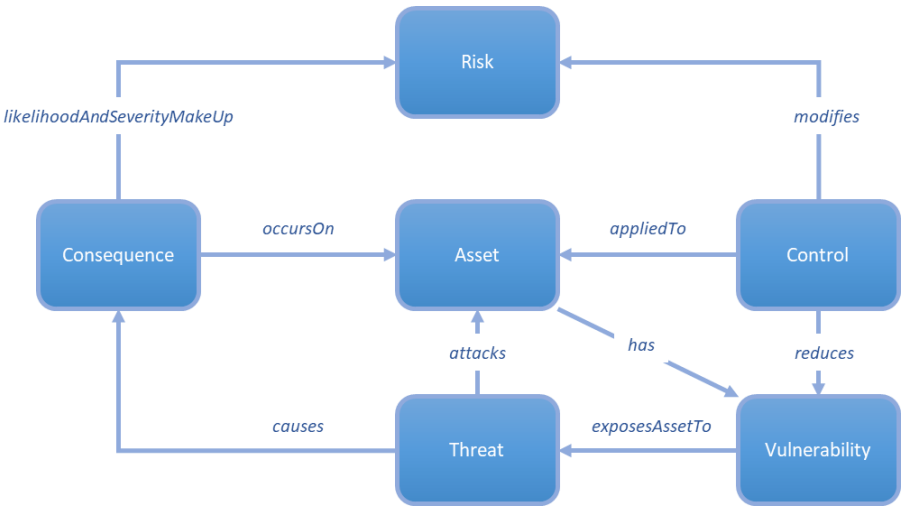
- **Expanding the 'Safe Projects' dimension to 'Safe Collaborations'** to better-reflect the diverse ways in which organisations are coming together to share resources and services for collaborative research, as this is not just happening at project-level but also at scale in cross council research networks—e.g., programme, institutional levels (such as part of federated TRE ecosystems).
- **Extending the 'Safe People' dimension to 'Safe Stakeholders'** to better-highlight the wide range of people and organisations that have responsibility for, access to and influence over resources and services (such as, co-designers, data providers, data subjects, data stewards, TRE operators) as part of collaborative research in cross council research networks. (Given the 'Safe People' dimension is typically concerned with researchers and data analysts.)
- **Emphasising data flows in the 'Safe Stakeholders' and 'Safe Data'** to draw greater 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). (Given the 'Safe Settings' dimension typically concentrates on a single data access facility or platform.)

Risk factors

As part of a conceptual mapping exercise, we **identify common risk factors** used by the ISO/IEC 27005 methodology for information security risk management and other selected privacy risk assessment methodologies. These risk factors will help to determine the types of information needed in order to assess the privacy risk for collaborative research relating to the emerging data patterns and needs of advanced analytics in cross council research networks.

[1] Note these principles are based on the original Five Safes (Plus One) and example interpretations—for further detail see Section 4 in Part B of the main report.

A 'Risk Management Upper Ontology' depicting the relationships between these common risk factors



Relationship between information privacy and information security—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.

Boundaries and scope

The boundaries and scope of the risk assessment need to be identified. These will determine the risk factors above that are of concern, specifically including those in the control of key stakeholders, and the outside factors that influence the risk assessment but cannot be controlled by them. While traditional risk assessments often focus on a fixed scope (e.g., the operations of a company), **a key distinguishing aspect of emerging data usage patterns for research collaborations is in their fluidity**. E.g., 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.

Mapping Scope of Risk Assessment to Risk Factors via an 'Enhanced Five Safes Plus One'

We also propose an initial mapping between the 'Enhanced Five Safe Plus One' and the Risk Factors identified, as a matrix to provide a classification scheme for the distinct types of information needed to determine the privacy risks arising from emerging data patterns and needs of advanced analytics in cross council research networks. An example is shown below—note that Figure is not yet populated as this will be evaluated and developed as necessary in further work in WP2 and WP3.

			Risk Factor Types				
			Asset Identification	Threat Identification	Vulnerability Identification	Consequence Identification	(Control Identification)
Scope of Risk Assessment	Safe Collaborations	Safe Stakeholders					
		Safe Data Flows: Safe Data					
		Safe Data Flows: Safe Settings					
		Safe Data Flows: Safe Outputs					
		Safe Data Flows: Safe Return					

Next steps for the project

We continue our project research activities as part of WP2 "Privacy Risk Framework Specification" and WP3 "Privacy Risk Modelling and Simulation". Key areas for further work include:

- Further **defining the scope** of the risk assessment.
- Progressing the **combination of relevant elements** of privacy and cybersecurity risk assessment methodologies.
- **Refining our approach to mapping risk assessment** scope to risk factors via the 'Enhanced Five Safes Plus One'.
- Automatically assess privacy risks for representative use case using **cyber security risk modelling and simulation**

Table of Contents

Abbreviations

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

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

List of Tables

List of Figures

1. Introduction

A DARE UK Sprint Project. The 'Privacy Risk Assessment Methodology' ("DARE UK PRiAM project") project is 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 will complete in August 2022.

1.1 Motivation for DARE UK PRiAM project

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). Organisations responsible for carrying out and facilitating such research activities must ensure that reasonable and acceptable levels of privacy protection are in place so that individuals, groups of people and wider society are not put at risk of undue harm.[2]

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"**[3] 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).[4] 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".[5]

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

[3] A "problematic data action" is defined by NIST (n.d.) as "a data action that could cause an adverse effect for individuals".

[4] 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".

[5] Related to Recital 75 of the General Data Protection Regulation (GDPR).

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).

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).

1.2 Aim of the DARE UK PRiAM project

We aim to lay the foundations for a standard privacy risk assessment framework that can describe and automatically assess privacy risk for safe federations.[6] The objectives of the project are to:

1. Define **use cases and data patterns** for advanced analytics;
2. Identify **privacy risk factors**;
3. Define a **risk tier classification framework**;
4. **Assess privacy risks for use cases** (related to public health research and integrated care) using cyber security risk modelling and simulation; and
5. Develop, evaluate and disseminate the framework and lessons learnt through **engagement with experts and the public**.

The framework for comparative assessment of different privacy risks will provide a reference to enable organisations to assess the overall risk levels. We will then investigate how to extend ISO/IEC 27005 information security risk management concepts and processes to privacy risk management. We will explore:

- **Important types of privacy risk** from the framework (e.g., re-identification);
- **Threats** that can cause privacy risks (e.g., linking);
- **Patterns of assets** to identify threats (e.g., aggregation of datasets);
- **Environments** that affect the likelihood of privacy threats (e.g., environment affecting the risk of re-identification);
- **Adversarial conditions** (e.g., motivations, capabilities and opportunity); and

[6] 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).

- **Controls** (e.g., homomorphic encryption, parquet encryption, secure enclaves, contracts) that can lower the likelihood of threats occurring or mitigate the impact of the risk.

Three work packages (WPs) will address user needs, privacy risk framework and implementation. WP1 "Use Cases, Evaluation & Stakeholder Engagement" will analyse use cases, requirements, conduct evaluation and capture/disseminate lessons learnt to maximise impact. WP2 "Privacy Risk Framework Specification" will identify privacy risks factors and develop the risk tier classification framework. WP3 "Privacy Risk Modelling & Simulation" will model risk factors and assess use cases using the ISO/IEC 27005 information security risk management methodology.

1.3 Deliverable 1 Report: Objectives

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),[7] 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 4:** 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. This report further contributes to addressing the second project objective—to be further refined as part of WP2 research activities:

> **Project objective 2 of 4:** 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).

[7] For definitions of the terms 'research collaboration' and 'federated research network' see Section 2.4 and the glossary in Section 8.

1.4 Deliverable 1 Report: Overview

This D.1 report is divided into the following three parts:

- **Part A. Outlining the Context for Privacy Risk Assessment: Defining Use Cases and Data Usage Patterns for Advanced Analytics.** In Section 2, we provide a detailed overview of the three driver use cases: (i) 'Use Case A: Complex Hospital Discharge—PROCED Project'; (ii) 'Use Case B: Multi-morbidity Prevention—MELD-B Project'; and (iii) 'Use Case C: NHSx Wessex Federated TREs'. Then, in Section 3, we examine data usage patterns and emerging data sharing needs related to these three driver use cases. *Part A addresses the first project objective (above).*
- **Part B. Laying the Foundations for a Standard Privacy Risk Assessment Framework: Initial Conceptualisation of Risk Factors.** In Section 4, we examine the evolution of the Five Safes from best practice principles for safe research focused on a single facility or platform, to the need for enhancing these Five Safes to better align with emerging data usage patterns and needs of advanced analytics in cross council research networks. In Section 5, we propose an initial privacy risk assessment approach, combining ISO/IEC 27005 methodology for information security risk management and other key sources related to privacy risk assessment, to determine the factor types needed to identify privacy risks for safe federations. *Part B contributes to addressing the second project objective (above).*
- **Part C. Conclusions.** The report then concludes by summarising key points from Parts A and B; and outlines next steps for the project in relation to ongoing WP2 and WP3 research activities.

Note that a glossary of key terms is provided in Section 8 of this report. This D.1 is the first in series of three DARE UK PRiAM project reports to be delivered.[8]

[8] The following reports are due for submission at the end of the DARE UK PRiAM project (Month 8, 31 August 2022): the D.2 report—'Privacy Risk Framework' that will describe the privacy risk assessment framework; and the D.3 report—'Privacy Risk Framework Application Guide' that will outline example usage in risk modelling platform.

PART A. Outlining the Context for Privacy Risk Assessment: Defining Use Cases and Data Usage Patterns for Advanced Analytics

Our work on privacy risk assessment is driven via cross-council use cases revolving around health (MRC), social science and social care (ESRC), and computer science (EPSRC). We have 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 to drive the identification of factors and situations causing and affecting privacy risks; which will also serve as validation cases.[9] These three use cases are (note: more detail is provided in the following sub-sections):

- **'Use Case A: Complex Hospital Discharge—PROCED Project'** that focuses on how to: (i) proactively model patient discharge risks (e.g., elongated length of stay, readmission) and expected departure points; and (ii) schedule and optimise provision of ongoing community care services. Data linkage: 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.
- **'Use Case B: Multi-morbidity Prevention—MELD-B Project'** that centres on how to: safely deliver an artificial intelligence (AI) enhanced epidemiological analytic system in which optimal lifecourse time points and targets for prevention of early-onset, burdensome multimorbidity are identified through analysis of birth cohorts and electronic health and care records. Data linkage: requires both individual linking and federated learning between longitudinal birth cohort and routine data sets over the lifecourse.
- **'Use Case C: NHSx Wessex Federated TREs'** that aims 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. Data linkage: 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]) and safe federations. 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.

Part A of this report therefore addresses our first research objective: to analyse driver use cases and data usage patterns from health and social care research typical of future MRC, ESRC and EPSRC data sharing (WP1, Outcome: understanding of future unmet data sharing needs). Part A is divided into two parts. First, in section 2, we provide a more detailed overview of these three use cases and outline some emerging data sharing needs. Then, in section 3,

[9] For instance, these use cases will be further assessed in WP3 "Privacy Risk Modelling & Simulation" using a privacy risk assessment approach combining the ISO/IEC 27005 methodology for information security risk management with well-known principles for safe research—the Five Safes (Desai et al., 2016)—and other key sources related to privacy risk assessment.

we examine data usage patterns and emerging data sharing needs related to these three drivers use cases.

2. Use Cases—related to Public Health Research and Integrated Care

We now provide a more detailed overview of our three driver use cases:

2.1 Use Case A: Complex Hospital Discharge—PROCED Project

"PROactive, Collaborative and Efficient complex Discharge" PROCED Project

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".

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. 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.

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.

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.

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.

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—MELD-B Project

"Multidisciplinary Ecosystem to study Lifecourse Determinants and Prevention of Early-onset Burdensome Multimorbidity" MELD-B Project

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).

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.[10]

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.

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)[11]

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.

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).

[10] 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.

[11] 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.

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.

2.3 Use Case C: NHSx Wessex Federated TREs

NHSx Wessex Federated TREs

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" (Jhutty & 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."

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).

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.

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

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.[12] 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).

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.[13] 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'[14] as

[12] 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).

[13] 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.

[14] 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.

"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).[15]

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, IT systems, governance practices, cultural attitudes to data sharing and privacy, etc.)" (Elliot et al., 2020).[16] 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'[17] 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).[18]

3. Data Usage Patterns in Safe Research Collaborations

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.[19] 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" (Jhutty & 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.[20]

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

[15] 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".

[16] 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). We have applied this principle to privacy risk assessment in relation to federated research networks.

[17] 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."

[18] For further discussion about data sharing in relation to 'direct' and 'indirect care' see: Information Governance Review (2013).

[19] 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."

[20] 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".

to better understand how privacy concerns, expectations and associated risks may develop and change as research collaborations become more federated—and the how this may further shape our conceptualisation of safe federation.

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)

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'[21] 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 in society 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.
- **Data linkage is established by data controllers**[22] 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.**

[21] 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).

[22] 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 […]".

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).

Figure 1 (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).

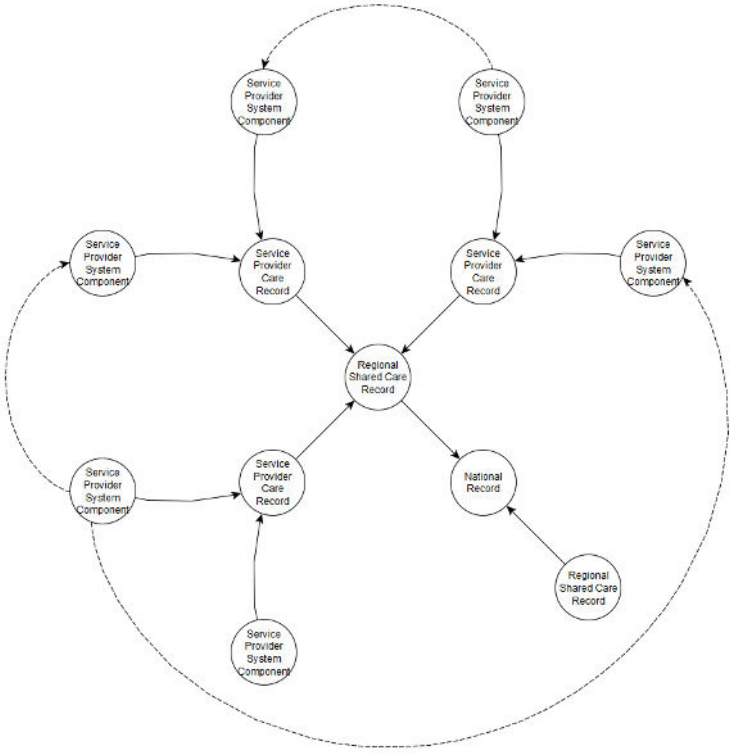


Figure 1: A Simplified Health Network including Service Integration and Data Aggregation

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 is 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.

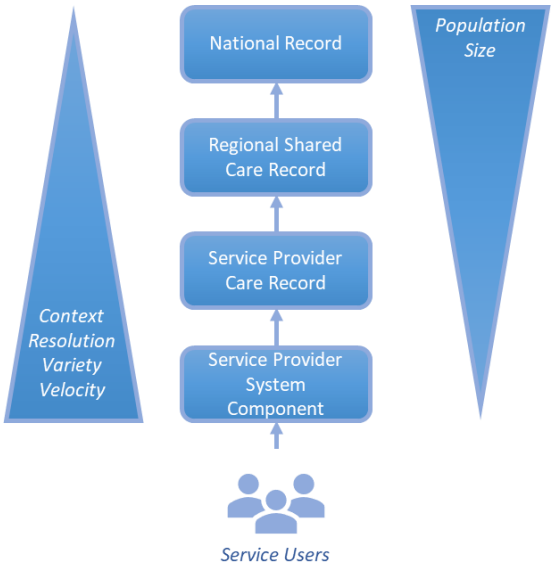


Figure 2: Distribution of Data Value and Impact of Data Aggregation

3.2 Trusted Research Environments in Health Care Systems

TREs have emerged to provide safe access to personal and 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. 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).[23] Three typical TRE deployment scenarios have emerged (see Table 1 on next page), which can be classified as:

[23] 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").

Table 1: Three Typical TRE Deployment Scenarios

THREE TYPICAL TRE DEPLOYMENT SCENARIOS			
	Data User TRE	**Service Provider TRE**	**Broker TRE**
Deployed at:	A research institution/company	A service provider	A legal entity operating a data marketplace
Data are ingested from:	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
Access to TRE:	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
Needed for situations e.g., where:	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

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 3 (below) provides an illustration of these three typical TRE deployment scenarios:

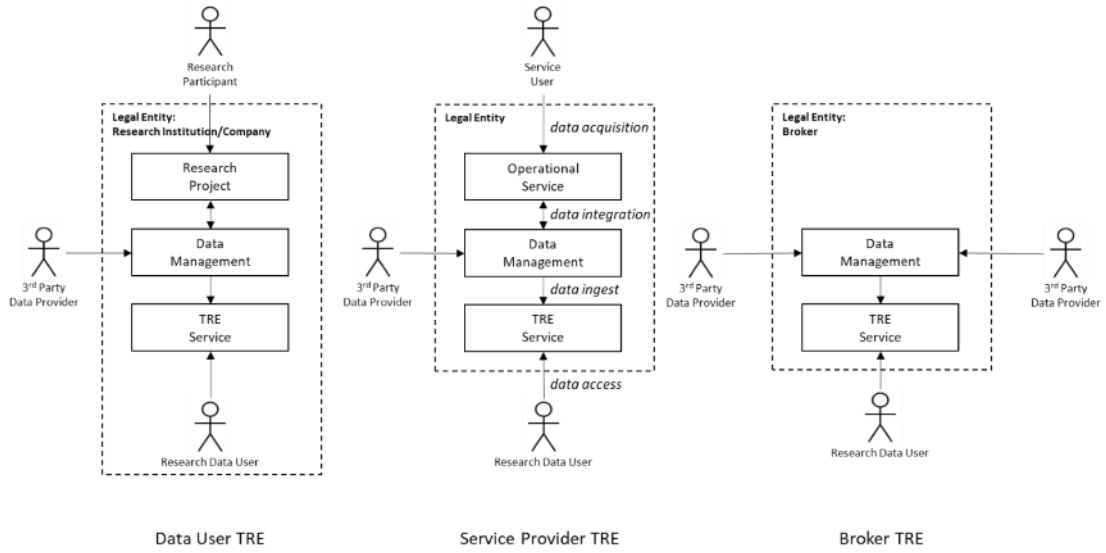


Figure 3: Typical TRE Deployment Scenarios

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 in different context 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 4 (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:

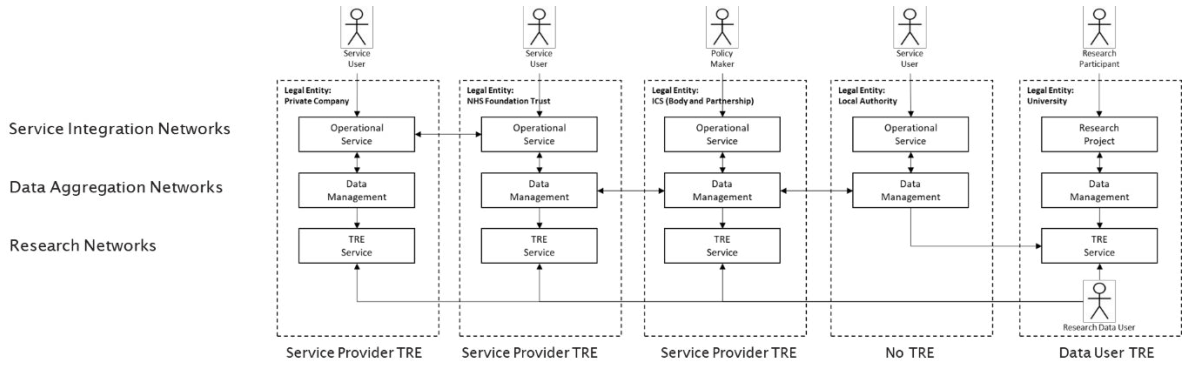


Figure 4: A Representative Example of a Research Network

We now explore these layers from this representative example of a research network (presented in Figure 4 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):

(i) 'An enhanced research experience' including support for 'advanced federated analysis' and 'distributed machine learning; and

(ii) 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 as a "a key concern" for researchers. 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[24]—such as for AI/ML—but also to be more connected with other TREs. Further, safe federation will also require greater

[24] For further discussion on the 'next generation capabilities' of TREs see: Kavianpour et al. (forthcoming/in press).

availability and interoperability[25] of quality data from service providers (e.g., for health care providers, social care providers) for research purposes.[26]

3.3.2 '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 (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.

Building greater capabilities for interaction with data

One area of interest for safe federations 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 safe federations can build greater capabilities for interaction with data, which empower data subjects to exercise their data-related rights[27] 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.

* * *

We now utilise our analysis of the use cases, data usage patterns and emerging data sharing needs (Sections 2 and 3) to help guide our initial conceptualisation of risk factors related to research collaborations in Part B of this report. We place particular emphasis on the ways in which the increasing federated nature of research collaborations is disrupting the particular assumptions and context on which best practice principles for safe research are based, such as the original Five Safes (Desai et al., 2016).

[25] 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".

[26] 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."

[27] 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).

PART B. Laying the Foundations for a Standard Privacy Risk Assessment Framework: Initial Conceptualisation of Risk Factors

The DARE UK PRiAM project aims to lay the foundations for a standard privacy risk assessment framework that can describe and automatically assess privacy risk for safe federations. This framework will combine well-known principles for safe research—the Five Safes (Desai et al., 2016)—and the ISO/IEC 27005 methodology for information security risk management to enable consistent, efficient and usable privacy assessment. In Part B of this report, we focus on conceptualising privacy risk factors and make an initial contribution towards our second research objective, to be fulfilled in WP2, which is: to identify key factors contributing to privacy risks within the Five Safes when datasets are linked as part of research collaborations (WP2, Outcome: understanding of privacy risk factors and consequences).

Part B is divided into two sections. First, in Section 4, we examine the evolution of the Five Safes from best practice principles for safe research focused on a single facility or platform, to the need for enhancing these Five Safes to meet emerging data usage patterns and needs for collaborative research. Then, in Section 5, we propose an initial privacy risk assessment approach, combining ISO/IEC 27005 methodology for information security risk management and other key sources related to privacy risk assessment, to determine the factor types needed to identify privacy risks for safe research collaborations. This initial privacy risk assessment approach will be further refined and developed in WP2 and WP3.

4. Evolution of the Five Safes

4.1 An Overview of the 'Five Safes Plus One'

The Five Safes (see Table 2 below) 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 Five Safes are also used as principles to describe best practices for safe research in relation to a particular data access facility or platform (see Table 2 below for example interpretations).

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 possibility 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). It is worthwhile to note that, in some cases, the Five Safes have been adapted—such as, to include "Safe Algorithm" on application to AI by ACS[28] (2018).

[28] The ACS is the professional association for the technology sector in Australia.

Table 2: Original Five Safe Questions together with Example Interpretations

	Original Five Safe Questions devised for the Office for National Statistics (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?"
Desai et al. (2016):	"[…] 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. […]"[29]	"[…] 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."
Health Data Research UK (HDR UK, n.d.):	"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"
Australian Institute of Health and Welfare (AIHW, 2021):	"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"
UK Data Service, SecureLab (2022):	"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"
Arbuckle & Ritchie (2019)—in the context of risk-based anonymisation:	"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"[30]	"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"

[29] 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.

[30] 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.

4.2 The Importance of Risk Communication

Privacy is a nebulous concept[31]—holding various meanings for people[32] as well as for different stakeholder groups and disciplines.[33] Furthermore, 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 it is being considered and can develop and change over time. Given our focus on identifying key factors contributing to privacy risks when datasets are linked as part of research collaborations, we are primarily examining those factors related to information privacy[34]—that is, the extent in which data flows across a safe federation can be controlled and can be interacted with (particularly by those people who are data subjects and/or co-designers).[35]

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).

Given privacy risk assessment for safe federations requires transdisciplinary, multi-stakeholder and cross-organisational attention, the Five Safes Plus One is an important aspect of a risk assessment methodology not only to structure such related discussions and decision-making, but also to help categorise risk factors in safe federations (see section 5 for more detail). It is important to highlight that Arbuckle & Ritchie (2019) have applied the Five Safes to risk-based anonymisation—using the example of "a health-care scientist who requests individual patient data" for research purposes—refer to Table 2 above for their interpretation of the Five Safes in this context. We highlight some key reasons for focusing on the Five Safes Plus One, as part of privacy risk assessment for safe federations, as follows:

- **Familiarity**—in that, many stakeholders are likely to have awareness of or previous experience of using the Five Safes.

[31] 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).

[32] 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."

[33] 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).

[34] 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".

[35] 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).

- **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 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) related to safe federations. Such synthesis of diverse requirements is of vital importance for robust data governance for federated research networks (and beyond).

4.2.1 Project Engagement Activities

Risk communication is a crucial element in the ongoing development of a standard privacy risk assessment framework. At time of writing, stakeholder engagement with legal, ethics, regulatory and information experts and practitioners is ongoing through an advisory board—to elicit requirements from use cases and provide insight into privacy risk. Further, public engagement activities are in progress through the 'Privacy risk Assessment Forum (PrAF)'—to capture public perceptions of the privacy assessment framework and its ethical treatment of personal data, trust development and maintenance. For example, the DARE UK PRiAM PrAF has used three scenarios as the basis for discussion: (i) online shopping, (ii) health activity tracking, and (iii) a research project studying complex discharge.

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.

4.3 Expanding the 'Five Safes Plus One' for Safe Collaborative Research

As shown by Table 2, the original Five Safe questions together with some example interpretations focus on best practices for safe research within a single data access facility—e.g., where data access is provided at a supervised and secure physical data room—or through a single platform—e.g., where remote data access is available through a trusted research environment.

From our analysis of the driver use cases and data usage patterns (Part A), it appears that **safe federation has disrupted the particular assumptions and context on which the original Five Safes is based on** (e.g., data access through a single facility or platform)—and further changes the risk factors associated with each of these dimensions. We therefore explore how we might expand the Five Safes principles to better-suit the context of safe federations and related emerging data sharing needs.

4.3.1 Federated Five Safes for an Alliance TRE Ecosystem (UK HDRA & NHSx, 2021)

In the paper published by the UK HRDA & NHSX (2021) on principles and best practices for TREs, we begin to see an initial evolution of the Five Safes—where 'best practice principles' for a group of federated TREs (in this case, an Alliance TRE ecosystem) are outlined:

- **Best practice principle for 'Federated Safe Projects':** "a single streamlined data access request management API" for Alliance TREs (UK HRDA & NHSX, 2021)
- **Best practice principles for 'Federated Safe People':** 'federated identity management' for Alliance TREs with "digital assertions of accreditations to be linked to identity", "external

evidence of information governance training attached to an individual's digital identity" and "any disbar process […] flagged against an individual's identity, so it is visible across the ecosystem" — further, where possible, information relating to "declarations required for transparency and avoidance of conflicts of interest" should be collected in one place (e.g., "an individual's account on the Gateway") (UK HRDA & NHSX, 2021)

- **Best practice principle for 'Federated Safe Data':** 'progressive adoption' of "a Common Data Model, agreed not only across the Alliance but across the wider TRE ecosystem" together with the development of "processes to generate transformations of data assets into this form" (UK HRDA & NHSX, 2021)
- **Best practice principles for 'Federated Safe Settings':** "TRE providers should": "provide ingress and egress (where allowed) to transfer data and code securely between SAFE Settings"; "adopt common pseudo-identifier generation processes to enable safe linkage between SAFE Settings" and "provide services that allow individuals to remotely execute analysis workflows using TRE supplied tools or externally supplied research software" (UK HRDA & NHSX, 2021)
- **Best practice principle for 'Federated Safe Outputs':** "TRE providers should consider deploying Beacons within their safe setting to support the programmatic generation of safe outputs from standardised external queries, subject to airlock review processes to approve acceptable query types" (UK HRDA & NHSX, 2021)

By considering the types of mechanisms that can be used to support a safe federated TRE ecosystem, these best practice principles for federated safe research provides a crucial first step in the evolution of the Five Safes for safe federations. However, we now consider the possibility that the Five Safes principles themselves require expansion to better suit the context of safe federations and related emerging data sharing needs.

4.3.2 Expanding the Principle of 'Safe Projects' to 'Safe Collaborations'

From our analysis of the use cases (Section 2), research collaborations are becoming increasing federated, in terms of resources and services, to support multi-disciplinary research undertaking advanced analytics (AI/ML). By definition, the 'Safe Projects' dimension focuses on research projects, however this may not sufficiently capture the increasingly federated nature of research (e.g., at project-level), and emerging data usage patterns. Research collaborations not only include projects, but extend to e.g., programmes of work, other types of data sharing initiatives, the provision of long-term resources and services by professional bodies, federated research networks.

There is therefore a case for potentially expanding this dimension to 'Safe Collaborations' to better-reflect the diverse ways in which organisations are coming together to share resources and services for research in safe federations, as this is not just happening at project-level but also at programme and institutional levels. For illustration, we provide the following example principle for 'Safe Collaborations' (based on Table 2 sources):

'Safe Collaborations'. The use of resources and services as part of a research collaboration (involving two or more organisations) is lawful, ethical and aligned with stakeholder expectations. The research collaboration is for public benefit.

4.3.3 Expanding the Principle of 'Safe People' to 'Safe Stakeholders'

The 'Safe People' dimension mainly focuses on researchers and data analysts for a specified project. Yet, research collaborations involve a wide-range of stakeholders (as illustrated in Section 3), including organisations (e.g., supervisory bodies, data providers, TRE operators), that have "responsibility for" and "access to" data (Elliot et al., 2020) such as data stewards,

co-designers, and policy makers—who are likely to have connections with multiple projects, programmes and data institutions.[36] 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.[37] There is therefore a case for expanding this dimension to 'Safe Stakeholders' to better highlight both its organisational aspect and focus on a wider range of actors. For illustration, we provide the following example principle for 'Safe Stakeholders' (based on Table 2 sources):

'Safe Stakeholders'. All people and organisations with responsibility for, accesses and utilises resources and services as part of a research collaboration 'have the skills, knowledge and incentives to act in accordance with required standards of behaviour'.

4.3.4 Greater Emphasis on Data Flows: 'Safe Data' and 'Safe Settings'

Given safe federations involve the combination of multi-stakeholder data and aim to support emerging data sharing needs (e.g., federated TRE ecosystems for advanced federated analysis, distributed machine learning), there is a need to place greater emphasis on data flows as part of safe settings.[38] For illustration, we provide the following example principles for 'Safe Data' and 'Safe Settings' in the context of safe federation (based on Table 2 sources):

'Safe Data'. All data processed as part of a research collaboration are reviewed to ensure adequate standards of data quality, data completeness and data richness. Data are treated appropriately and effectively to minimise privacy risks to individuals, organisations, groups of people and wider society.

'Safe Settings'. All data processed as part of a research collaboration takes place within one or more specified environments; all environments and data flows in-between have effective and appropriate privacy and security controls.

4.3.5 On the relationship between 'Safe Data' and 'Safe Outputs'

The Original Five Safes make a clear distinction between inputs and outputs of data from a statistical disclosure control perspective. Data inputs are data made accessible via an individual facility or platform for (re)usage as part of a specified project, which are the focus of the 'Safe Data' dimension. Data outputs are the data insights generated from data analysis (e.g., derived data) that are authorised for release outside the specified facility or platform under consideration, usually as open or restricted publications, which are the focus of the 'Safe Outputs' dimension.

However, from the viewpoint of the UKAN Anonymisation Decision-Making Framework: "this is a distinction without a difference" as data inputs and outputs are both constitute flows of data moving between different environments (Elliot et. al., 2020). These different data

[36] 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).

[37] For instance, 'stakeholders' 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".

[38] 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.

situations—further described by Elliot et al. (2020) as "the relationship between some data and their environments, seen as a total system"—will require different types of controls. It is further interesting to note that the HDR UK (2021b) also pairs together 'Safe Data' and 'Safe Outputs' as part of its "open, federated and interoperable technology stack for trusted research environments" and "capability maturity model" for a "federated data analytics infrastructure".

Within a safe federation, there may be multiple flows of data, including ingestion and outputs across various TREs, for a single project. There is therefore a need to distinguish 'Safe Outputs'—i.e., a transfer of data outside the data sharing initiative, from those other outputs as data flows between environments within a research collaboration. For illustration, we provide the following example principles for 'Safe Outputs' (based on Table 2 sources) and 'Safe Return' (based on Hubbard et al. [2020]) in the context of safe federation:

'Safe Outputs'. Insights generated from a research collaboration will undergo appropriate checks to ensure any residual risks remain very low to individuals, organisations, groups of people and wider society.

'Safe Return'. The re-combination of outputs from a research collaboration with other data at the 'clinical setting that originated the data' can only take place where permitted and consented by the data subjects concerned.

5. Risk Factors, Scope of Risk Assessment Privacy and Protection Goals in Safe Research Collaborations

This section aims to pave the way for subsequent work in WP2 and WP3 by identifying the types of factors and privacy goals needed for privacy risk assessment. It further provides a first draft approach for mapping these factors to the 'Expanded Five Safes Plus One' for safe research collaborations (identified in Section 4), as a method to classify the elements needed to identify privacy risk—an approach that will be developed and expanded as further work.

5.1 Privacy Risk Assessment Factors

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 outlined in Part A). We now explore a selection of these methodologies in more detail.

5.1.1 Sources

The factors are identified by taking a baseline of cybersecurity risk assessment, a closely related (but distinct) field to privacy risk assessment (see Sub-Section 5.3 for more detail), plus existing privacy risk assessment methodologies themselves; and identifying common factors they use. 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[39] e.g., assets, consequences, controls, threats, vulnerabilities—as outlined by ISO/IEC 27005. The consideration of key risk factors

[39] 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).

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:

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 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.

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).

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).

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"[40] (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).

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 5 (below). The ISO/IEC 27005 methodology helps organisations to identify systemic assets together with vulnerabilities and threats that can exploit them. An

[40] See section 1.1 of this report for further information.

"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.

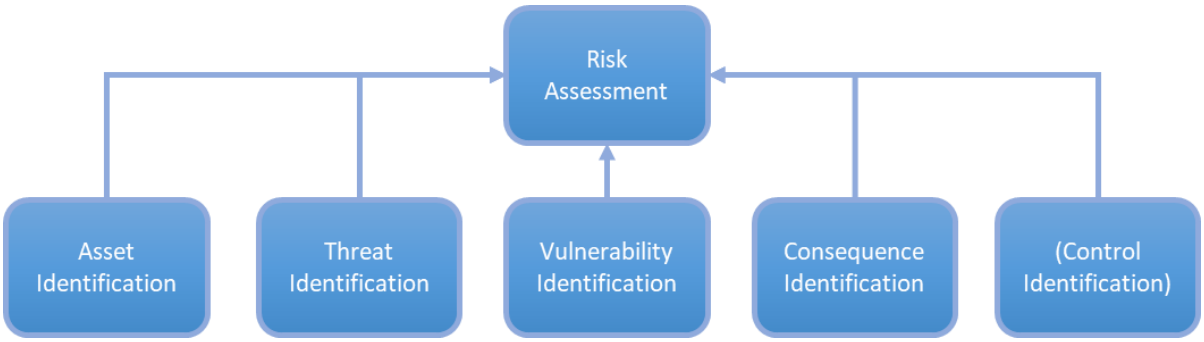


Figure 5: ISO 27005 Risk Assessment Identification Activities

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.

5.1.2 Risk Factor Types

In order to understand the types of risk factor 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 sources. The first draft of this ontology is shown in Figure 6 (below) and followed by the definitions and their mapping to the sources.

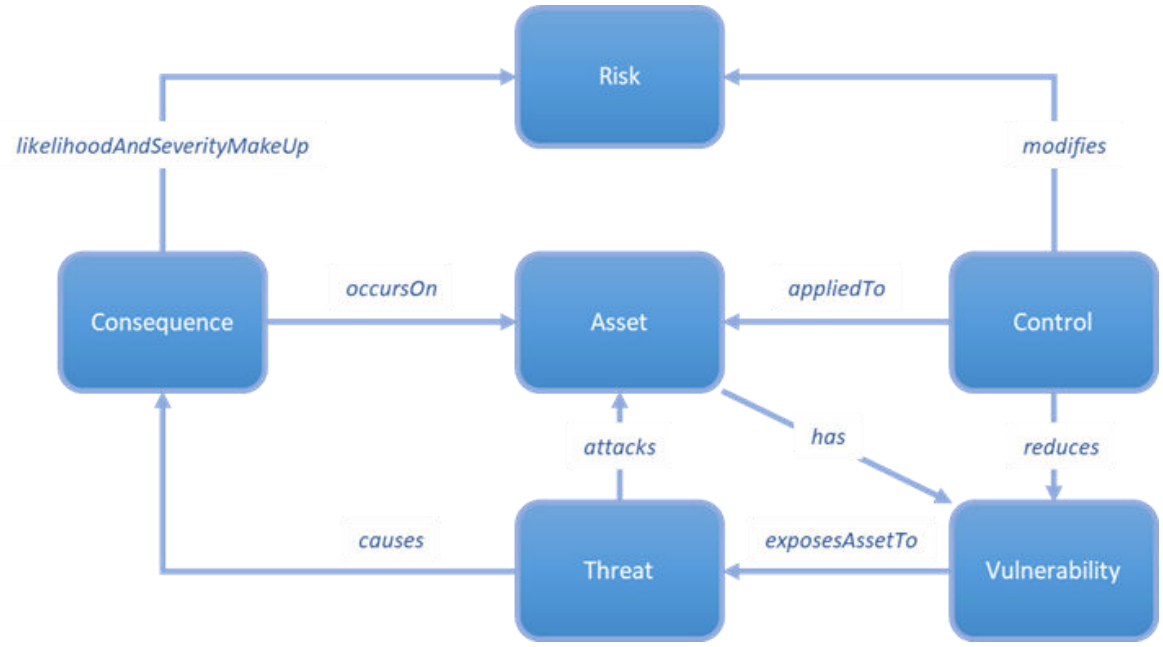


Figure 6: Risk Management Upper Ontology

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

Table 3: 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
Asset	• "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). • "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.	• 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." • Inria Privacy Risk Analysis Methodology also considers **Supporting Asset** as "such as hardware, applications, data stores, software environment, etc." (De & Le Métayer, 2016). Also, **Data Actions, Data** and **Relevant Contextual Factors**. • 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). The main focus of NIST PRAM therefore is on data actions rather than assets.
Threat	• "Potential cause of an unwanted incident, which may result in harm to a system or organisation"—as defined by ISO 27000. • "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).	• CNIL PIA (2018a) defines Threat as "Procedure comprising one or more individual actions on data supporting assets". • Note **Problematic Data Action** in NISTIR 8062 is used rather than Threat and Vulnerabilities**:** "A data action that causes an adverse effect, or problem, for individuals" (Brooks et al., 2017).
Consequence	• "Outcome of an event affecting objectives"—as defined by ISO 27000. • Also, **Threat Consequence: "**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).	• 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: **Feared Event:**

	• 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.	• CNIL PIA (2018a) defines Feared Event as "Potential data breach likely to have impacts on data subjects' privacy". • 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 & Le Métayer, 2016). **Privacy Harm:**[41] • 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 & Le Métayer, 2016). • 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). • **Privacy Harms** can be considered a specific type of Consequence. Also, **Problems:** • 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).

[41] Also, refer to Section 1 of this report for further examples of potentially harmful activities and undue harm related to privacy.

Vulnerability	• "Weakness of an asset or control that can be exploited by one or more threats"—as defined by ISO 27000. • "(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). • 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.	• CNIL PIA (2018a) refers to the "the level of vulnerabilities of personal data supporting assets". • As a "more general term than vulnerabilities", Inria Privacy Risk Analysis Methodology utilises the term **Privacy Weakness**: "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 & Le Métayer, 2016). • Again, note **Problematic Data Action** in NISTIR 8062 is used rather than Threat and Vulnerabilities (Brooks et al., 2017).
Risk	• "Effect of uncertainty on objectives"—as defined by ISO 27000.	• 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).
Control	• "Measure that is modifying risk. May include any process, policy, device, practice or other action. Controls may not always exert the intended or assumed modifying effect"—as defined by ISO 27000. • Also, **Security Control**: "The management, operational, and technical controls (safeguards or countermeasures) prescribed for an	• **Privacy Control**:[42] "The administrative, technical, and physical safeguards employed within an agency to ensure compliance with applicable privacy requirements and manage privacy risks"—as defined by NISTIR 8062 (Brooks et al., 2017). • CNIL PIA defines Control as "Action to be taken. [/] Note: this may be technical or organisational and may

[42] 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".

	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).	entail putting fundamental principles into practice or avoiding, reducing, transferring or assuming all or part of the risks". • 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 & 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 & Le Métayer, 2016).

5.2 Scope of Risk Assessment

The boundaries and scope of the risk assessment needs to be identified. These will determine the risk factors above that are of concern, specifically including those in the control of key stakeholders, and the outside factors that influence the risk assessment but cannot be controlled by them. 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).

ISO27005 has a **Context Establishment** setup step in its methodology (Figure 7 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).

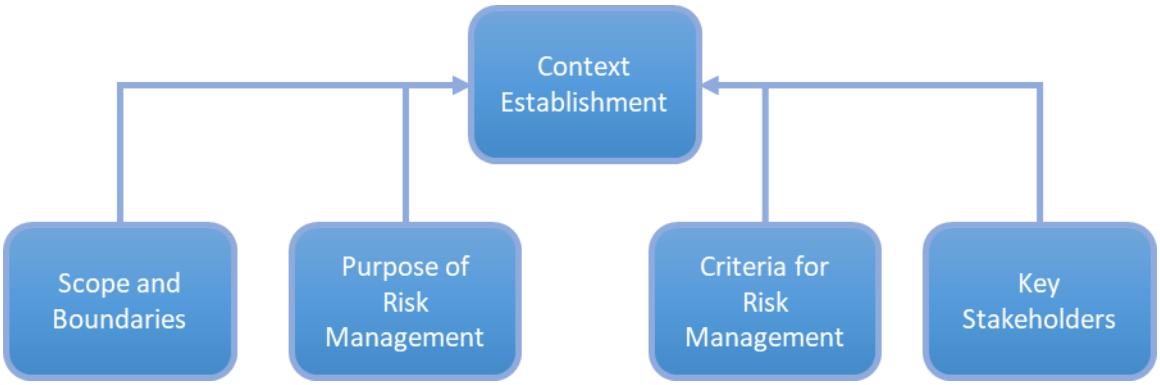


Figure 7: ISO 27005 Context Establishment

It is important to note that ISO/IEC 27005 focuses on a system that is traditionally assumed to be from the perspective of a business (e.g., a company) seeking to understand their systems and the vulnerabilities, threats and risks associated with them. While traditional risk assessments often focus on a fixed scope (e.g., the operations of a company), **a key distinguishing aspect of emerging data usage patterns for research collaborations is in their fluidity**. E.g., 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. This is a key point that will need extension for the purposes of the federated contexts that are the subject of the DARE UK PRiAM project. Determination of the risk assessment scope (and therefore the scope of the 'socio-technical system' referred to above) is a key item of further work. It is clearly related to the federated aspect referred to in Sections 2, 3 and the extension of the Five Safes to accommodate federations in Section 4. The relationship between the risk assessment scope and these aspects will be undertaken in WP2 and WP3 combined and will be reported in D.2 and D.3 from the perspectives of risk tiers and risk modelling respectively.

5.3 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).[43] 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.,

[43] 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.

2015).[44] The Standard Data Protection Model (Conference of the Independent Data Protection 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:[45]

- **"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"[46]**—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).[47] These measures have been "tried and tested in the data protection audit practices of several data protection supervisory authorities for many years" (CIDPSAFL, 2020).

[44] 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).

[45] 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.

[46] 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).

[47] 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."

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:

1. **Confidentiality and availability**—e.g., where restrictions on data accessibility may conflict with access to data without delay (Hansen et al., 2015).
2. **Integrity and intervenability**—e.g., where the need to modify data may conflict with a requirement to keep data intact (Hansen et al., 2015).
3. **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.[48] This balancing exercise therefore requires employment of robust risk communication mechanisms, including meaningful stakeholder involvement (Section 4), and readiness of appropriate and effective technical and organisational measures to minimise existing privacy risks.

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 modelling—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.

[48] 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).

5.4 On the Relationship between Information Security and Information Privacy Risk Assessment Methodologies

As a pointer to further work in WP2 and WP3, it is important to note that while concerns associated with information security and information privacy are conceptually related,[49] each area offers a distinct focus (Bambauer, 2013; Brookes et al., 2017; Kuner et al., 2017; IAPP, n.d.).[50] We therefore highlight three key observations:

Consideration of impacts from (un)authorised processing

A key observation concerning this distinction between information security and privacy risk assessment methodologies is that "[i]n the security risk model, concerns focus on unauthorized activity that causes a loss of confidentiality, integrity or availability of information or systems" (Brooks et al., 2017). Whereas, in the privacy risk model "while some privacy concerns arise from unauthorized activity, privacy concerns also can arise from authorized processing of information about individuals" (Brooks et al., 2017).[51]

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).

Focus on impacts for a wider range of stakeholders

Breaches of privacy affect data subjects, groups of people and wider society so the risks extend not only to the operator or stakeholders of the federated system under examination, but also to people who may have no direct connection with it. Therefore, there are often additional third party roles that are affected by Consequences to be considered in privacy risk assessment over those typically involved in cybersecurity risk assessment, and avoidance of harm to these third parties is a security goal in its own right but may also cause knock-on Consequences for the operators or stakeholders of the system under examination (for instance reputation damage to data controllers resulting from privacy harms to data subjects). This issue is also related to the question of acceptability by groups representing data subjects and the general public discussed in Section 4.2.

5.5 Mapping Scope of Risk Assessment to Risk Factors via an 'Enhanced Five Safes Plus One'

The contribution and interplay between aspects covered in this section will need to be considered and incorporated into the 'DARE UK PRiAM Privacy Risk Assessment Methodology'—the key outcome of this project. A first draft approach to addressing this

[49] For instance, Section 2 of the GDPR specifically concerns "Security of Personal Data". Article 32 of the GDPR focuses on the 'security of the processing' and places an obligation on controllers and processors to "implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk" by considering "the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons". Further, Recital 39 of the GDPR states: "[…] Personal data should be processed in a manner that ensures appropriate security and confidentiality of the personal data, including for preventing unauthorised access to or use of personal data and the equipment used for the processing."

[50] In the words of the International Association of Privacy Professionals (IAPP, n.d.), "While security is necessary for protecting data, it's not sufficient for addressing privacy." Note that there is also an ISO/IEC 27701:2019 for Privacy Information Management.

[51] 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."

interplay is to combine the 'Enhanced Five Safes Plus One' (see Section 4) with the Types of Risk Factors to provide a classification scheme in terms of both axes. An example is shown below—note that Figure 8 is not yet populated, as this approach will be developed and expanded as further work in WP2 and WP3.

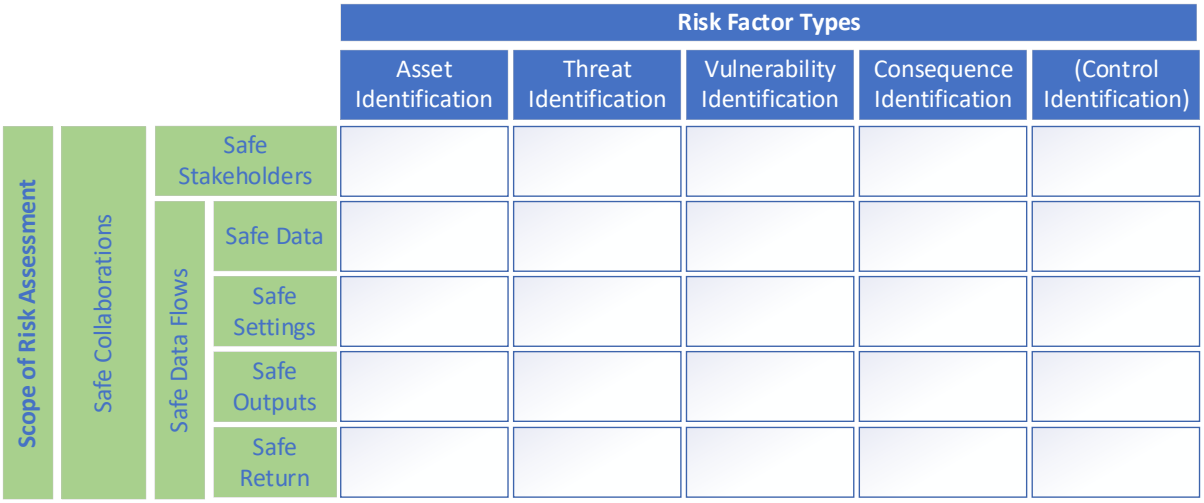


Figure 8: Mapping Risk Factor Types to Scope of Risk Assessment

The expected usage is that the scope of the risk assessment is determined by the Safe Collaboration(s), which in turn specifies the Stakeholders and Data Flows involved. The Data Flows concern Data assets that flow between different Settings, are Output or Returned to their sources. Each of these Safes can either identify, affect or be affected by one or more of the Risk Factor types. As a brief example, the Data or Stakeholders can be specified in the Assets column, the Data Flows between two Settings may present a Threat of Data Leakage resulting in a Consequence of Loss of Confidentiality for that data.

This mapping will be further developed in the risk assessment methodology work of WP2 and WP3 and will be used as a means to relate the risk assessment scope in terms of the federated situations in question, to identify the relevant factors of each type.

PART C. Conclusions

6. Summary of Key Findings

The three real-world use cases outlined in this report (Section 2)—a project related to complex hospital discharge; a project concerned with multiple long-term condition multimorbidity prevention; and a pilot scheme for developing a sub-national federation of TREs—provide exemplars of multi-disciplinary research collaborations.

From our analysis of these use cases, it is clear that data usage patterns related to a TRE, or otherwise federation of TREs, should be considered in the context of the system they are established to study (Section 3). Understanding 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. We therefore explored the operational context of health data networks, including the role of TREs, and provided a representative example of a research collaboration as part of a federated data usage scenario.

Key finding A. From this analysis of data usage patterns, it is clear that **a TRE offers a partial view of a complex data network**—where different TREs will have multiple views of a complex data network.

We then provided an overview of emerging data sharing needs in this context (e.g., advanced federated research analysis, distributed machine learning, greater opportunities for co-design and interaction with data) to better understand how privacy concerns, expectations and associated risks may develop and change as research collaborations become more federated—and the how this may further shape our conceptualisation of safe federation.

Key finding B. From this analysis of emerging data sharing needs, **the increasing federated nature of research collaborations is disrupting the particular assumptions and context on which best practice principles for safe research are based**, such as the original Five Safes—and further changes the risk factors associated with each of these dimensions.

Key finding C. A further key finding is that the **emerging data sharing needs** of researchers and data analysts should not only be considered, but also those of **all key stakeholders** (e.g., data subjects, co-designers), such as how **TREs can offer a view on networked data by providing an interface for increased interaction.**

In Part B of this report, we started to lay the foundations for a standard privacy risk assessment framework that can describe and automatically assess privacy risk for safe research collaborations. This methodology will combine well-known principles for safe research—the Five Safes (Desai et al., 2016)—and the ISO/IEC 27005 methodology for information security risk management to enable consistent, efficient and usable privacy assessment.

We therefore explored how we might enrich and expand the Five Safes Plus One principles (Section 4) to better suit risk communication in relation to the emerging data usage patterns and needs of advanced analytics in cross council research networks. **We have therefore enhanced the Five Safes Plus One by:**

- **Expanding the 'Safe Projects' dimension to 'Safe Collaborations'** to better-reflect the diverse ways in which organisations are coming together to share resources and services for collaborative research, as this is not just happening at project-level but also at scale in

cross council research networks—e.g., programme, institutional levels (such as part of federated TRE ecosystems).

- **Extending the 'Safe People' dimension to 'Safe Stakeholders'** to better-highlight the wide range of people and organisations that have responsibility for, access to and influence over resources and services (such as, co-designers, data providers, data subjects, data stewards, TRE operators) as part of collaborative research in cross council research networks. (Given the 'Safe People' dimension is typically concerned with researchers and data analysts.)
- **Emphasising data flows in the 'Safe Stakeholders' and 'Safe Data'** to draw greater 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). (Given the 'Safe Settings' dimension typically concentrates on a single data access facility or platform.)

Key finding D. From discussion concerning the expansion of the Five Safes Plus One, it is clear that **safe federation is not a standalone principle** to be added in addition to 'Safe People/Stakeholders', 'Safe Projects/Collaborations', 'Safe Data', 'Safe Settings' and 'Safe Outputs'. Rather, safe federation needs to be understood in relation to best practice for safe research—and is conceived therefore as **an overarching, relational concept that permeates and augments each of these principles to better suit emerging data usage patterns and data sharing needs** for multi-disciplinary research collaborations utilising advanced analytics (AI/ML).

We then undertook a conceptual mapping exercise (Section 5)—by identifying common risk factors used by the ISO/IEC 27005 methodology for information security risk management and other selected privacy risk assessment methodologies; and highlighted key data protection goals for safe federations. As part of this exercise, we identified several key types of Risk Factors for risk assessment that will determine the types of information needed in order to assess the privacy risk for research collaborations and federated situations. These are Assets (which can be ICT components, software, data, and socio-technical aspects such as places and stakeholders specifically including data subjects), the Threats that can affect Assets, the 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—security or other defensive measures.

Key finding E. While traditional risk assessments often focus on a fixed scope (e.g., the operations of a company), **a key distinguishing aspect of emerging data usage patterns for research collaborations is in their fluidity.** E.g., where combinations of federated resources (e.g., multi-source data) and services (e.g., provided by a group of federated TREs) are used and re-purposed for multiple types of users for specific (connected) projects as well as within and across different programmes of work.

We also proposed an initial mapping between the federated context, the Enhanced Five Safe Plus One (Section 4) and the Risk Factors (Section 5), as a matrix to provide a classification scheme for the distinct types of information needed to determine the privacy risks in these federated scenarios. This will be evaluated and developed as necessary in further work in WP2 and WP3.

6.1 Next Steps

We continue our project research activities as part of WP2 “Privacy Risk Framework Specification” and WP3 “Privacy Risk Modelling & Simulation”. The use cases will further drive the identification of factors and situations causing and affecting privacy risks (WP2); which will also serve as validation cases (WP3). Further, the key findings we have identified in this report will help to inform the design of the risk tier classification framework in WP2; and contribute to the modelling of risk factors and further assessment of use cases in WP3. Key areas for further work include:

- To further define the scope of the risk assessment.
- To progress the combination of relevant elements of privacy and cybersecurity risk assessments—note that a key part of this work been initiated in this report in terms of the privacy goals and guiding principles that need to be respected and balanced with what is permissible (e.g., given legal, ethical, and cyber-security requirements) and practicable (e.g., given the objectives of the research collaboration under assessment).
- To refine our approach to mapping risk assessment scope to risk factors via the Federated Five Safes Plus One.

We will deliver two further reports at the end of the project: the D.2 report—that will describe the ‘Privacy Risk Framework’ and the D.3 report—that will outline the ‘Privacy Risk Framework Application Guide’.

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.

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/.

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., Lefkovitz, 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.

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.

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.

De, S.J., & Le Métayer, D. (2016). PRIAM: A Privacy Risk Analysis Methodology. Research Report, RR-8876, Inria - Research Centre Grenoble – Rhône-Alpes, ffhal-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. 2nd 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-pressures-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.

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.

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.

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). (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.

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.

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.

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

Jhutty, 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-Benssassi, 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/ipx009.

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.

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.

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.

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 EWelfare, 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.

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 IEFT 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.

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., & Zanfir-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.

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.

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.

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/.

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.

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.

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:

Complex discharge	A patient who requires "more specialised care after leaving hospital"—as defined by NHS (2019).
Controller	"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.
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).
Data Ecosystem	"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).
Data Flow	"The movement or transfer of data through a system, describing who has responsibility for and access to them, and the contexts in which it is held"—as defined by the UKAN ADF (Elliot et al., 2020).
Data Subject	An "identified or identifiable natural person" to whom personal data relates"—as defined by Article 4(1) of the GDPR.
Data Value Chain	"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).
Enhanced Five Safes Plus One	Refers to the expansion and enrichment of the Five Safes Plus One (see definitions below), as part of the DARE UK PRiAM project to better suit emerging data usage patterns and data sharing needs of research collaborations in relation to cross-council research networks.
Federated Research Network	"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).
Five Safes	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).
Five Safes Plus One	Refers to the addition of 'Safe Return' to the Five Safes by the HDRA UK (Hubbard et al., 2020).
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"—as defined by the World Health Organization (WHO, 2007).
Intervenability	"the data subject's capacity for intervention and control in the processing"—as defined by Agencia Española de Protección de Datos (AEPD, 2019).
Multimorbidity	The co-occurrence of two or more long-term health conditions.

Personal Data	"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.
Privacy Protection Goals	"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).
Privacy Risk Assessment	"A privacy risk management sub-process for identifying and evaluating specific privacy risks"—as defined by NIST Privacy Framework (NIST, 2020).
Processing	"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.
Research Collaboration	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 through federated research networks.
Risk Communication	"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).
Risk Factor	"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).
Service Integration	Connectivity between services to create business processes and care pathways (e.g., referral and discharge).
Trusted Research Environment (TRE)	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').