DARE UK Privacy Risk Assessment Methodology Project (PRiAM) Project: D3 Report v1.1

Privacy Risk Framework Application Guide

UK Research and Innovation
HDRUK Health Data Research UK
ADRUK Data-driven change

DARE UK

Document Details

Date	20/09/22
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

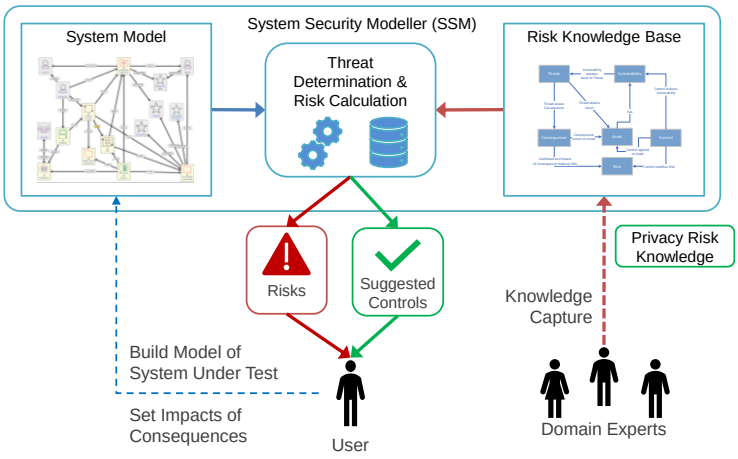
This report is independent research supported by the National Institute for Health and Care Research ARC Wessex. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

Further dissemination

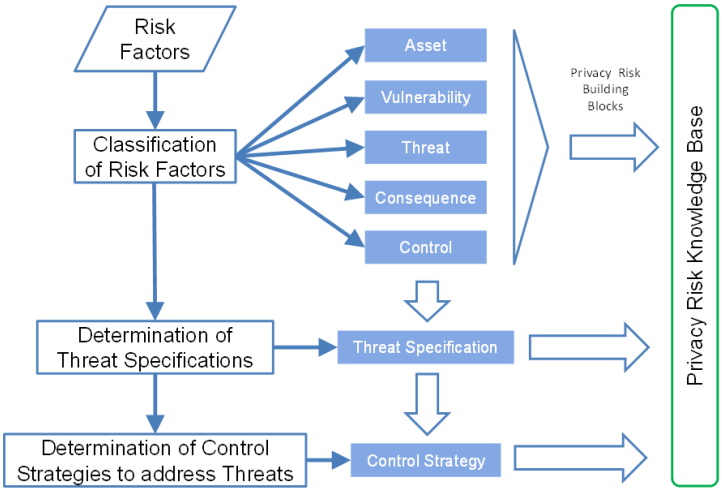
An overview of this work — entitled 'Towards a Socio-Technical Approach for Privacy Requirement Analysis for Next-Generation Trusted Research Environments' — was presented at the CADE 2022 Conference (Competitive Advantage in the Digital Economy) on 13 June 2022.

Executive Summary

This report is Deliverable 3 (D3) "PRiAM Privacy Risk Framework Application Guide" of the DARE UK PRiAM project. The report is one in a series of four project reports, which focus on working towards standardisation of privacy risk assessment for cross-domain access and re-use of sensitive data for research purposes.



This report describes how to automate privacy risk assessment by augmenting a pre-existing cybersecurity knowledgebase with privacy risk factors and then using the combined knowledge in an ISO 27005 risk assessment process using a System Security Modelling (SSM) platform. This approach allows data governance practitioners to construct a model of a system that can be used to explore threats, risks and consequences in a transparent, repeatable and efficient way. ISO 27005 is adopted as it is well established and integrating privacy risk management into a methodology that already supports cybersecurity risk management has considerable benefits. Traditionally, risk assessment is undertaken through communication and consultation with stakeholders and often requires significant expertise. Encoding privacy risk factors within a reusable knowledge base and providing a decision support tool implementing standard processes reduces the expertise needed by data governance practitioners.



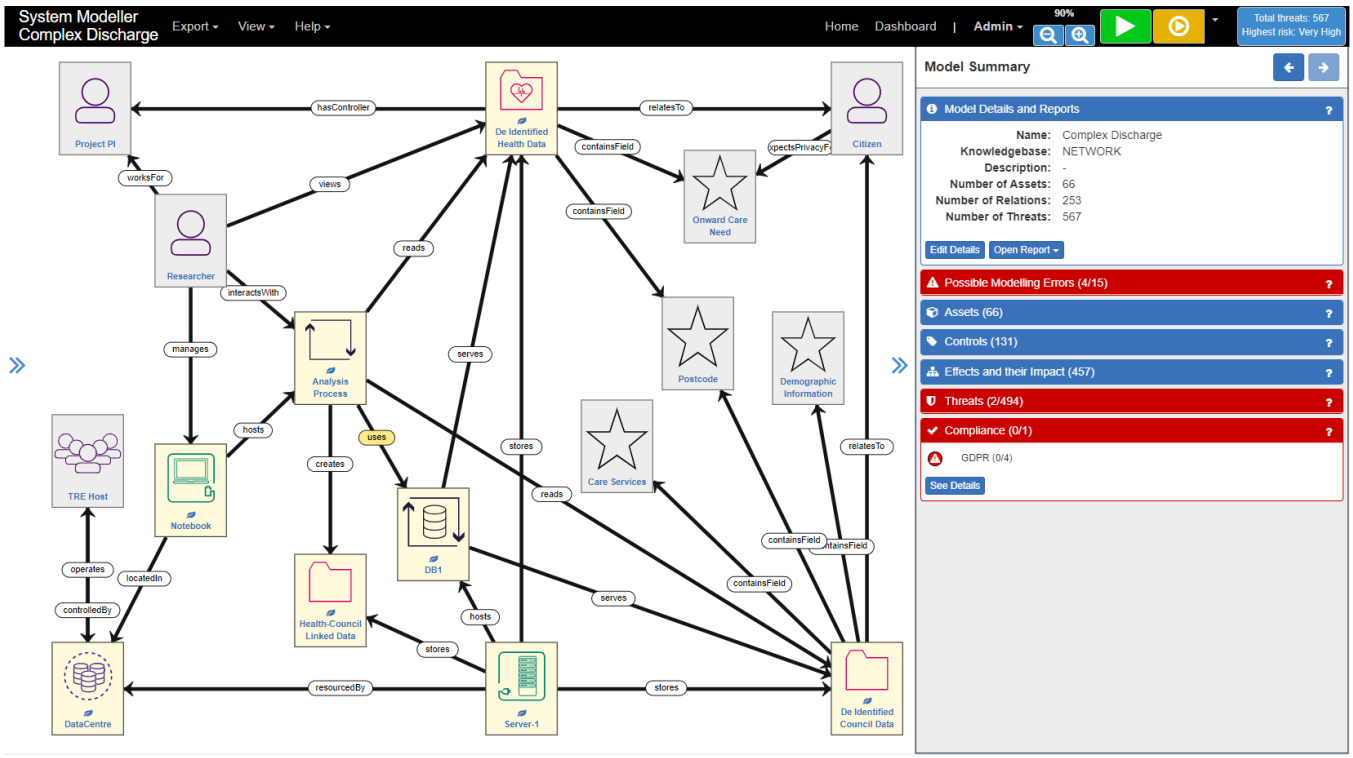
The process of knowledge capture and engineering is based on identifying and classifying the cause and effect relationships between the elements of risk. These elements include types of Assets, Vulnerabilities, Threats, Consequences and Controls that together define Threat Specifications and Control Strategies to address threats. New elements of each of these types have been determined specific to privacy protection from analysis of risk factors associated with the Five Safes framework. Example risk factors considered are those identified in the PRiAM Risk Tiers framework such as: Is Considered Sensitive, Presence of Direct Identifiers, Presence of Indirect Identifiers that can Single Out, More Data than Required for Project, Analytics Experience of Researchers, Activity Logging, Data Linkage Policy and Control, and Data Egress.

The overall risk assessment process is illustrated by example using a use case scenario for a research project studying complex hospital discharge that is hosted within a Trusted Research Environment. The use case depends on linking between multi-stakeholder datasets across an integrated care pathway (e.g., acute care provided in hospital and community care provided by a local authority). The data contains sensitive information for which there is an expectation of privacy on the part of data subjects with a risk of privacy violations even though the scenario has assumed that the data from both the hospital and council is de-identified and minimised as a starting condition. This example illustrates that federated data analysis and data linking can cause additional privacy risks that must be addressed. Two threat and control strategies are modelled and elaborated in detail within the use case:

- Presence of indirect identifiers that can single out and the threat from “Reidentification from Viewing Sensitive Data”.
- Data Linkage Policy and Control and the threat from “Linking of Sensitive Data”

The Threat Specifications described in this report are encoded into the SSM’s knowledge base for the purposes of demonstration and evaluation of the mapping between risk factors and ISO27005 concepts, but the concept of Threat Specification is useful in its own right because it ties together the real-world elements (e.g. data subjects, data, processing, operators, etc), their relationships with the causes (vulnerabilities and threats) and effects (consequences) of privacy risks and the controls to address threats.

The system model for the use case is defined in the SSM as shown below. The initial risk levels are calculated, threats explored, and control strategies implemented to reduce the residual risk in the system. We demonstrate how the overall risk level for the system risk is reduced from “Very High” to a “Low” by application of control strategies to address the risk of “Loss of Privacy” at the data subject.



The approach described is a first step towards the open curation of knowledge for privacy risks as a key underpinning element of standardisation. In future we would expect open communities of domain experts and the public to contribute to identification, curation and reuse of an open knowledge base for modelling, assessing and communicating risks in collaborative research networks. Sources of knowledge about privacy risk factors will continue to be identified from literature, expert communities and public consultation, whilst the DARE UK community and wider stakeholders continue to be a further valuable source of privacy risk factors.

A key recommendation from DARE UK PRiAM is to continue to build community expertise in analysis of risk factors and curation of privacy knowledge in both human and machine-readable formats to increase awareness of practitioners and to allow for development of transparent, repeatable and automated privacy risk assessment processes.

Table of Contents

List of Figures

List of Tables

Abbreviations

CNIL	Commission nationale de l'informatique et des libertés (France)
D.	Deliverable
DARE UK	Data and Analytics Research Environments UK
DARE UK PRiAM	DARE UK Privacy Risk Assessment Methodology
GDPR	General Data Protection Regulation
HDR UK	Health Data Research UK
ICT	Information Communication Technologies
ISO	International Organization for Standardization
KB	Knowledge Base
NHS	National Health Service
NIHR	National Institute for Health and Care Research (UK)
NIST	National Institute for Standards and Technology (USA)
NIST PRAM	NIST Privacy Risk Assessment Methodology
PI	Principal Investigator
PTSD	Post-Traumatic Stress Disorder
RFC	Request for Comments
SSM	System Security Modeller
TRE	Trusted Research Environment
UK	United Kingdom
UKRI	UK Research and Innovation
WP	Work Package

1. Introduction

1.1. Purpose

This report is Deliverable 3 (D3) *"PRiAM Privacy Risk Framework Application Guide"* of the DARE UK PRiAM project. The report is one in a series of four project reports, which together focus on working towards standardisation of privacy risk assessment for cross-domain access and re-use of sensitive data for research purposes.

1.2. About the DARE UK PRiAM project

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

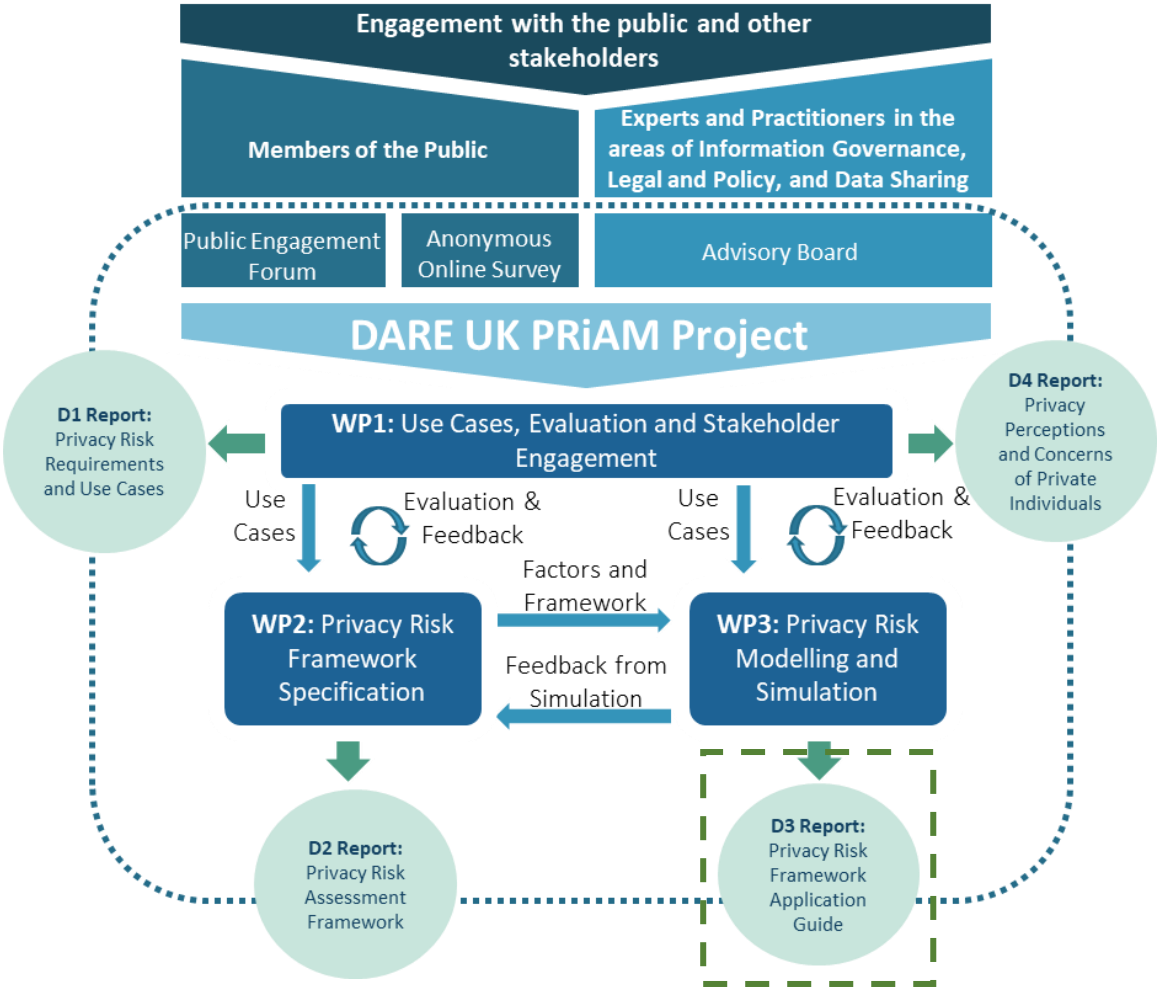


Figure 1: An Overview of the DARE UK PRiAM Project: Deliverables, Stakeholder Engagement and Work Packages

1.2.1. Motivation

Trustworthy and collaborative data sharing and re-usage for approved research purposes can help to advance public health and patient care. Data and analytics systems are changing and new ways to share and access data are emerging, including the potential for greater federation of resources and services. These changes are bringing about new and evolving risks. What remains vital is that people are protected from harms associated with data disclosure and re-use — and that public confidence and engagement in health and social care research are maintained. As such, the DARE UK PRiAM project aims to explore methods and tools that can support decision-makers, patients and the public to assess and manage privacy risk when considering emerging data access and re-usage scenarios, such as federation.

1.2.2. Project objectives

Our project objectives are as follows:

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

1.2.3. Project structure

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

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

1.2.4. Engagement with the public and other stakeholders

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

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

1.3. Scope of the D3 report

Work Package 3 (WP3): Privacy Risk Modelling & Simulation. This Deliverable 3 (D3) report focuses on modelling influencing factors on privacy risk identified in PRiAM D2 in terms of assets, threats and risks to privacy consistent with the risk assessment process of ISO27005 and encoding risk factors in a knowledge base. The knowledgebase is then used to simulate a use case scenario from those defined in PRiAM D1 report. This D4 report specifically concentrates on the following project objective:

> Assess privacy risks for use cases using a cyber security risk modelling and simulation platform, focusing on privacy risk (re-identification), threats (linking), adversarial conditions (motivations, capabilities and opportunity), controls (homomorphic encryption, parquet encryption).

To achieve this objective, this document describes the process of Privacy Risk Modelling, where ISO 27005 processes have been adapted to address the needs of privacy risk. In order to provide decision support for practitioners in this process, the knowledge on privacy risk factors (PRiAM D2 Report) have been encoded into a Privacy Risk Knowledge Base which is integrated into an automated toolkit for risk management. This document describes the process of encoding these factors into the Knowledge Base, and how this Knowledge Base is used to provide privacy risk decision support for practitioners.

This document is structured as follows:

- **Section 2** describes the overall risk modelling approach from the perspective of ISO 27005, and positions the Privacy Risk Knowledge Base in terms of automation of this process.
- **Section 3** describes how the Knowledge Base can be augmented with privacy risk factor knowledge, in terms of an abstract knowledge capture process, resulting in an enhanced knowledge base.the res the underlying principles of the risk modelling approach, its key elements and the process of mapping risk factors into the elements needed for the ISO/IEC 27005 methodology for information security risk management.
- **Section 4** is a worked example of the knowledge modelling approach, the augmentation of the knowledge base and the knowedge base's usage in a decision support siutation driven by a key PRiAM Use Case A from D1 ("Complex Discharge from Hospital"). The use case is a research project involving linking sensitive data from two different sources and indirect identifiers of people, with the associated risks of privacy harms. The modelling process is described and an illustration of the knowledge encoded into an automated risk management decision support tool illustrated to show how the knowledge can be used to reduce privacy risk levels.
- **Section 5** summarises the content of this report, draws conclusions and makes recommendations for future work.

Section 6 covers references and Section 7 has a Glossary of key terms.

2. Risk Modelling Methololology

PRiAM's automated risk assessment is founded upon the methodology proposed by ISO 27005 [ISO27005]. Other risk management methodologies exist, and the reasons for selection of ISO 27005 have been given in PRiAM Report D1. Briefly, ISO 27005 is well established and supports cybersecurity risk management, a closely related concern to privacy risk management. Integrating privacy risk management into a methodology that already supports cybersecurity risk management has considerable benefits. The overall process for Risk Management followed by ISO 27005 is illustrated in Figure 2.

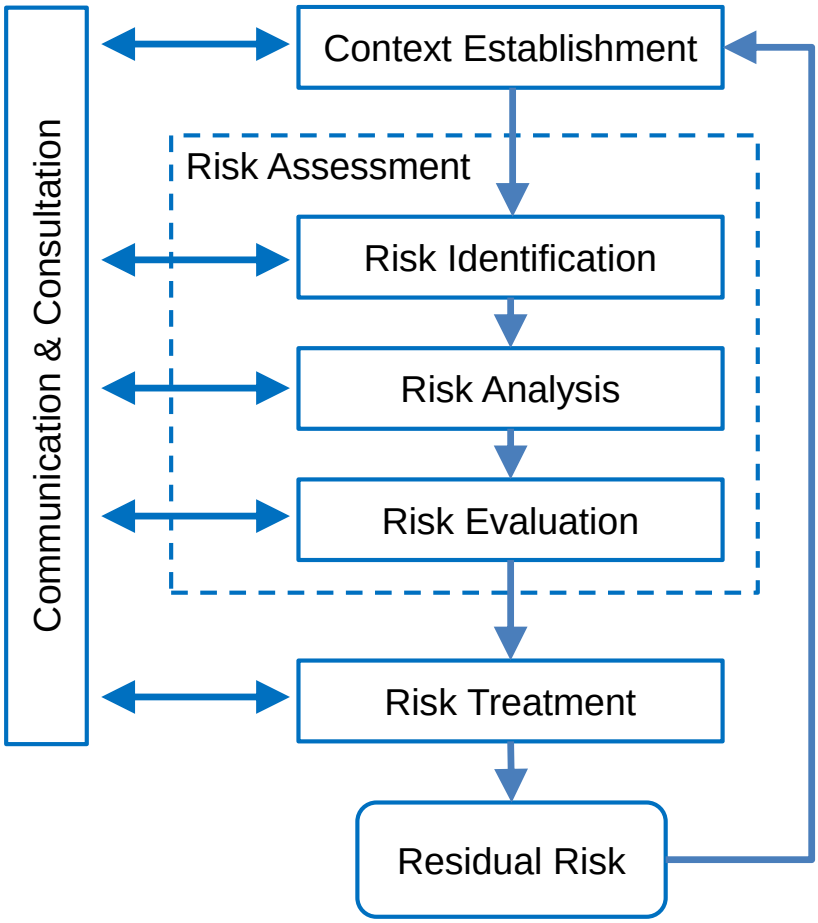


Figure 2: Risk Management Process (from ISO 27005)

Context Establishment defines:

- *Scope & Boundaries.* Determination of the physical and organisational scope of concern. ISO27005's primary focus is on a business (e.g., a company) but for collaborative research networks we should consider a research project as a key scoping entity. A project requires access to data, whilst a project may extend beyond a single stakeholder to multiple collaborating institutions. In terms of the Five Safes, this is the "Safe Project". The "Safe Project" helps scope the other Safes, i.e., a project has staff it employs ("Safe People"), the environments & places it operates within ("Safe Settings"), the data it uses ("Safe Data") and results it delivers ("Safe Outputs").
- *Purpose of Risk Management.* The purpose of privacy risk management is to enable the research project to achieve its goals without harming the privacy of the data subjects. Related aspects such as legal compliance and reputation protection within in the context of the scope and boundaries, are also valid purposes.

- *Criteria for Risk Management.* Determination of risk types that are in scope and their levels that are acceptable. Key risk types include losses of privacy on data subjects or losses of confidentiality integrity on personal data that can lead to losses of privacy.
- *Key Stakeholders.* The actors and roles that have interest in the risk management or are affected by the processing. This maps to the Five Safes' "Safe People". This can include individual or institutional roles, e.g. Trusted Research Environment (TRE), data providers, project Principal Investigators or researchers.

Risk Assessment involves:

- *Risk Identification*. Determination of the key assets of concern, their vulnerabilities, the threats that can affect them and the consequences arising from the threats that lead to risk.
- *Risk Analysis.* Determination of the likelihood and impacts of consequences of threats on assets to determine risk levels.
- *Risk Evaluation.* Assessment of risk levels and comparison against risk assessment criteria to determine if the overall risk level is acceptable.

Risk Treatment concerns determination of controls that can be applied to address the threats and therefore reduce the risk levels.

The output of Risk Treatment is Residual Risk - the remaining risk levels after identified controls have been applied. If the Residual Risk levels are acceptable, the process can stop. Otherwise the risk management process needs to be iterated, with additional controls identified to reduce the risk to an acceptable level.

Traditionally, this process is executed via communication and consultation with key stakeholders, but the challenge is that risk identification, analysis, evaluation and treatment often required significant expertise in risk modelling and the application domain of concern, and as such, ISO 27005 is typically implemented by domain experts either within an institution / company or bought in via consultancy.

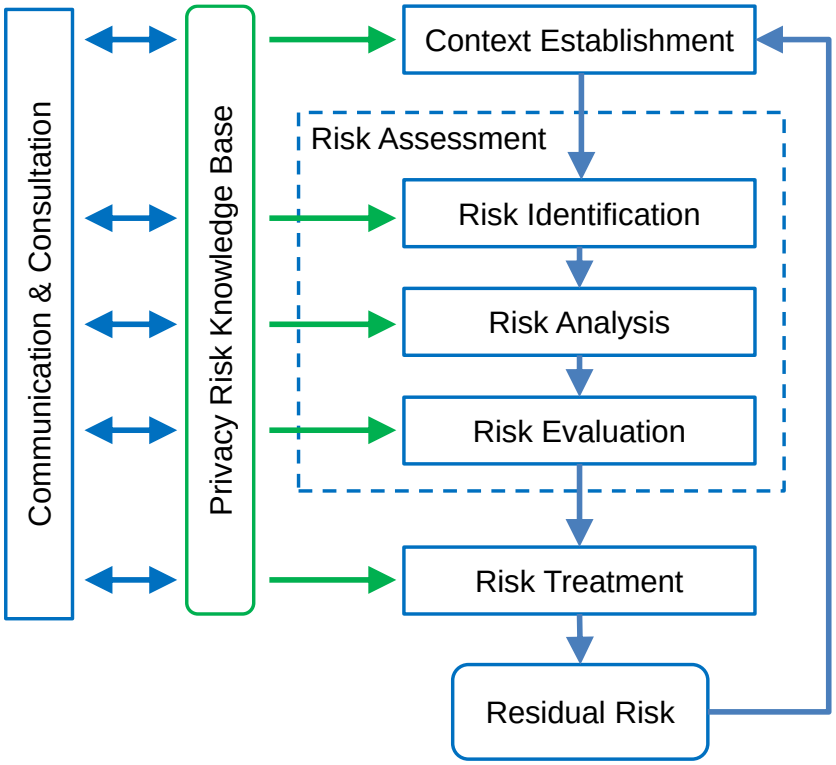


Figure 3: PRiAM Privacy Risk Knowledge within ISO27005 Risk Management

For DARE UK PRiAM, the ***domain of concern is privacy risk***, with privacy risk factors identified through the communication and consultation with stakeholders undertaken by WP1 (detailed in D1 and D4) and WP2 (detailed in D2). This report describes the process by which this knowledge of privacy risk factors may be encoded into a knowledge base that can provide decision support in the steps of the ISO 27005 risk assessment methodology. The approach provides decision support to compliance practitioners without the need for extensive human expertise or consultancy. We illustrate how privacy risk knowledge is incorporated into the ISO 27005 processes in Figure 3 (as marked in green). Once knowledge of privacy risk factors are encoded into a Privacy Risk Knowledge Base we are able to automate decision support to all phases of the ISO 27005 Risk Management Methodology.

The Privacy Risk Knowledge Base describes the relationships between the key elements of risk assessment: assets, their vulnerabilities, threats that can affect assets, consequences of threats on assets resulting in risks and controls that address the threats. The Privacy Risk Knowledge Base is aimed to be machine readable to enable automated decision support for privacy risk assessment, but it is useful in its own right to help communication because it is also human readable provides a taxonomy of risk assessment elements and their relationships.

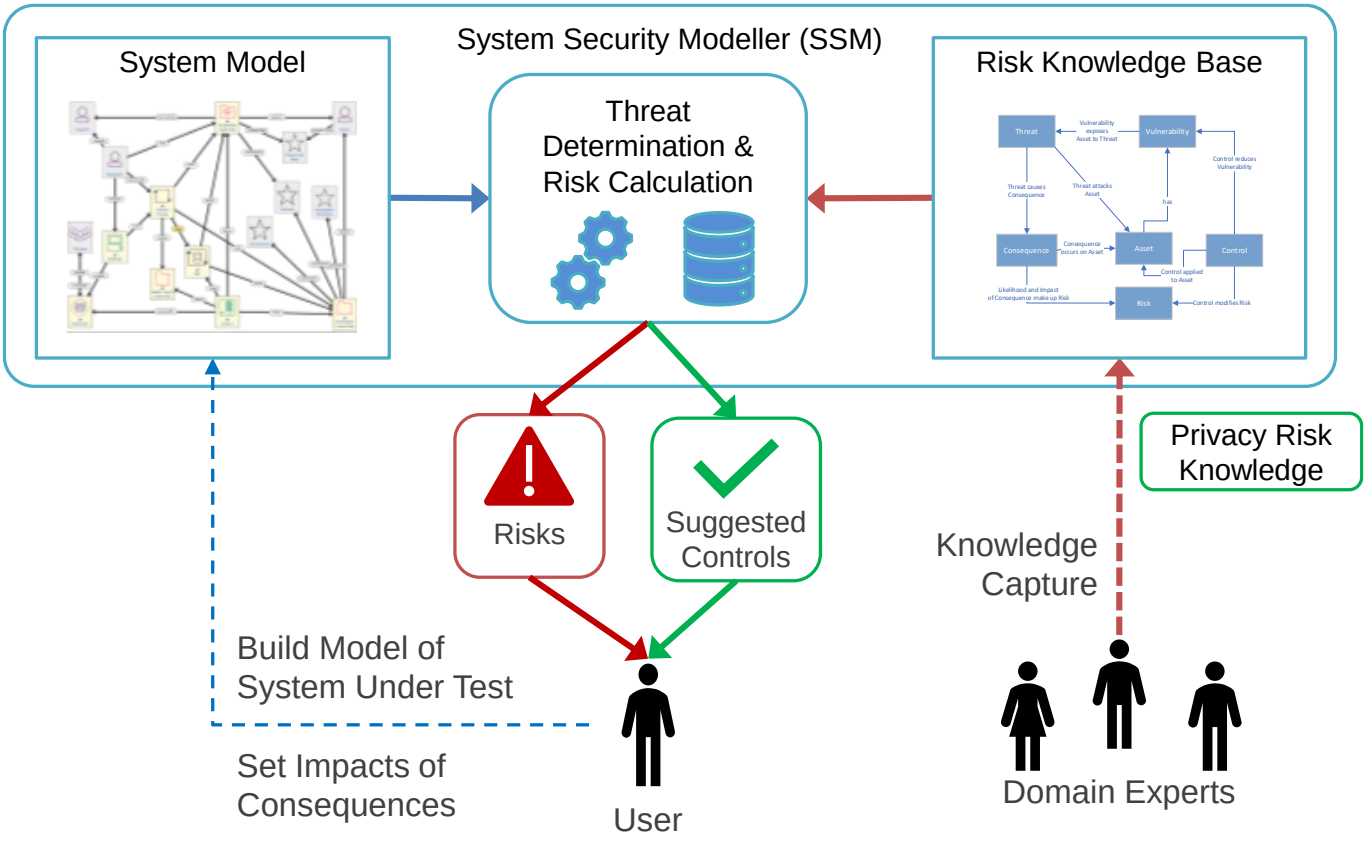


Figure 4: UoS System Security Modeller Concept

For the purposes of evaluation, the Privacy Risk Knowledge Base has been integrated into an automated risk assessment toolkit named the System Security Modeller (SSM), described in [Surridge 2018] and [Phillips 2022]. The SSM is a decision support simulator where expert knowledge is encoded into its Risk Knowledge Base regarding systemic assets, vulnerabilities, threats, consequences (termed "Effects" in the SSM) leading to risks; plus controls to address the threats. As shown in Figure 4, a user can construct a model of the system under examination (the System Model) and threats, risks and consequences are automatically detected in the System Model via utilisation of the expertise in the Risk Knowledge Base. The SSM began in the cybersecurity domain, was extended in [Taylor 2021] for regulatory compliance decision support using the GDPR as an exemplary regulation and is further

extended in PRiAM as will be described in the next sections to provide decision support for privacy risks via extension of its Risk Knowledge Base with privacy-specific risk knowledge.

Each threat in the SSM Risk Knowledge Base specifies asset types, the relationships between them and vulnerabilities that trigger a threat, plus consequences of the threat on affected assets. If a threat's specification of assets, relationships and vulnerabilities exists in a System Model (e.g., as shown in Figure 4), then the threat is automatically determined to be present in the system and the likelihood of threat's consequences (e.g. Loss of Confidentiality, Loss of Privacy, etc) can be calculated. The threat patterns may include the data flows and network paths, etc, that the SSM finds in the model. The threats are generic (regular updates are not required) and all threats are considered at once (i.e., there is no need to define the attacker or attack point). The threat coverage includes access and control privileges, software vulnerabilities, non-malicious threats, insider attacks, stolen devices, malicious attacks and compliance. Controls to bring a system into compliance include specifying policies such as gaining user consent or other lawful bases.

Risks are associated with consequences, i.e., the risk of the consequence occurring. The SSM uses an ordinal scale of risk levels, from "Very Low", through "Low", "Medium", "High" to "Very High". Risk levels are determined via a lookup table from the likelihood calculated by the SSM plus the impact (severity) specified by the modeller.

		Calculated **Likelihood**				
		Very Low	Low	Medium	High	Very High
Specified **Impact**	Very Low	Very Low	Very Low	Very Low	Low	Low
	Low	Very Low	Very Low	Low	Low	Medium
	Medium	Very Low	Low	Medium	High	High
	High	Low	Medium	High	Very High	Very High
	Very High	Low	Medium	High	Very High	Very High

Figure 5: SSM Risk Determination from Impact (Severity) and Likelihood

The Impact (or Severity) and Likelihood scales are defined as follows (Table 1). These definitions predate DARE UK PRiAM and focus on cybersecurity, but they represent the same concepts as a privacy-specific approach such as (CNIL 2018a). The SSM definitions have been preserved so that privacy risks can share the same scales as the related field of cybersecurity, as described above.

Table 1: Impact (severity) & Likelihood Scale and Definitions

Level of Consequence	**Impact Definition**	**Likelihood Definition**
Very high	Fatal to key interests. Very few effects will have this impact level, and those that do must be prevented at all costs.	Something will definitely go wrong if the possibility exists even only for a short time.
High	Has a serious impact on interests, and will be fatal if not addressed quickly.	Something is likely to go wrong if the possibility exists even only for a short time. Something will definitely go wrong if the possibility persists.

Medium	Can be tolerated for a short time, but will become serious if not addressed.	It is unlikely that anything will go wrong if the possibility exists only for a short time. Something is likely to go wrong if the possibility persists.
Low	Can be tolerated for a longer time, but it does degrade function or efficiency.	It is unlikely that anything will go wrong if the possibility exists only for a short time. Something is likely to go wrong if the possibility persists for a long time.
Very low	Can be tolerated for a long time, with limited impact.	It is unlikely that anything will go wrong even if the possibility persists for a long time.

Figure 6 shows the user interface of the SSM Toolkit with an example System Model loaded. The top right corner of Figure 6 indicates the worst case (highest) risk, which here is Very High. The section at the right hand side of the screen has collapsible panels covering different information useful to the modeller. There are panels for model details and possible errors in the modelling. Following this there is an Asset panel, which is an index of the assets the user has placed in the model. Next there is a Controls panel, which indicates the controls that are applied (or can be applied) at assets. Below this is the Effects panel, which describes the types of consequences at each asset along with their impact and likelihood. Threats are indexed next, followed by compliance issues. The Effects panel is a key focus of interest since the effects (consequences) determine the risks in the system being modelled, where the risk level is determined from the impact (severity)of the effect combined with its likelihood.

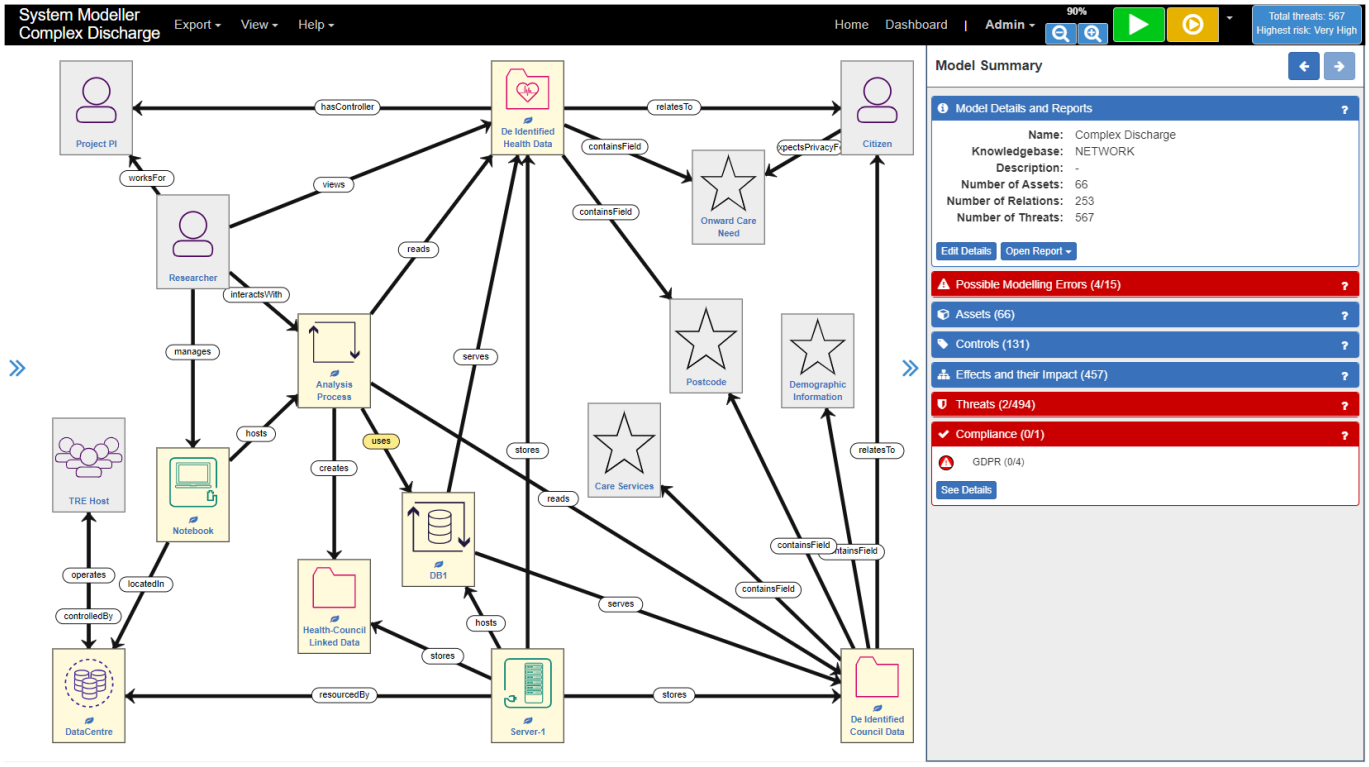


Figure 6: System Security Modeller Toolkit Environment

The SSM supports risk assessment and treatment in the ISO 27005 pattern:

1. *Risk Identification*. The Knowledge Base specifies the assets from which the user can construct a System Model, and also has threats, vulnerabilities and consequences associated with assets. The user builds

System Model from assets specified in the KB, and threats and their consequences on assets are automatically detected from the configuration of assets in the System Model.

2. *Risk Analysis.* As part of building a System Model, the user specifies the impacts (severity) of consequences based on their priorities and concerns. For example, the user may regard a "loss of confidentiality" consequence on personal data as "High" impact and a "loss of availability" on the same data as "Medium" impact. The SSM has an engine that automatically determines the likelihood of consequences of threats on assets. The combination of risk and likelihood determines a risk level for each consequence.
3. *Risk Evaluation.* The SSM displays the consequences with their associated risk levels (highest risk level first), so the user can judge whether the individual risks and the overall worst case risk level is acceptable.
4. *Risk Treatment.* The SSM enables the user to explore the threats that lead to consequences and recommends controls that can be applied to address the threats.

The remainder of this document describes how the Risk Knowledge Base is enriched with privacy risk knowledge and how this knowledge is employed in the SSM tool to provide automated decision support on privacy risks. The next section, 3, describes the knowledge modelling process by which the Risk Knowledge Base is created or augmented by mapping privacy risk factors identified in PRiAM Report D2 into the relevant elements needed for ISO 27005 risk management. This is followed by a worked example that illustrates how the Risk Knowledge Base is extended to accommodate privacy risks in section 4, motivated by a real-world use case from PRiAM Report D1 (Use Case A: Complex hospital discharge — "PROactive, Collaborative and Efficient complex Discharge" Research Project) and then describes how this knowledge is used in the SSM. Section 4 first describes the motivating use case. It then illustrates how the knowledge modelling process is employed for two threats to privacy are encoded into the knowledge base of the risk management toolkit[1]. Finally, it illustrates how the ISO 27005 process can be followed in the toolkit using the example case and the new knowledge of the privacy threats.

[1] It is important to note that although this worked example is driven by an exemplary use case, the knowledge encoded in the knowledge base is applicable to many other use cases.

3. Privacy Risk Knowledge Modelling

This section describes the concepts and process of knowledge modelling. This is the encoding of privacy risk factors (determined as described in PRiAM Report D2) into the elements needed for ISO27005 risk assessment and treatment, so as to create a computational description of privacy knowledge to enable automation of risk management.

3.1. Risk Modelling Concepts

In order to encode the privacy risk factors into elements needed for ISO27005 risk assessment and treatment, the elements required for ISO27005 risk assessment need to be defined, and Table 2 provides these definitions. This table is an expanded version of the table "Mapping Risk Management Concepts to Privacy Risk Assessment" in PRiAM Report D1. D1 defined the concepts in its "Risk Management Upper Ontology", and in this document, the Risk Management Upper ontology has been expanded to provide more detail (Figure 7), and the updates to Table 2 reflect this greater level of detail.

Table 2: Risk Modelling Concept Definitions

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.
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	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.,

Risk Management Concepts	Definition(s) from Cybersecurity Risk Assessment	Mapping these Risk Management Concepts to Privacy Risk Assessment
	basic types are 'unauthorized disclosure', 'deception', 'disruption' and 'usurpation'" — as defined by RFC 4949 (Shirey, 2007). 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.	Privacy Harms. These two concepts are defined as follows: **Feared Event:** 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:** 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

Risk Management Concepts	Definition(s) from Cybersecurity Risk Assessment	Mapping these Risk Management Concepts to Privacy Risk Assessment
		autonomy", "loss of liberty" and "physical harm"; and "loss of trust".
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 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).	**Privacy Control:**[2] "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 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).
Impact Criteria	The degree of damage or costs to the organization caused by an information security event considering: the level of classification of the impacted information asset, breaches of	The impact criteria regards the impact / severity of a consequence relating to a loss of privacy. The loss of privacy is experienced by the data subject but may have further related

[2] 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: Agencia Española de Protección de Datos (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. (2019) classify controls as "corrective controls", "detective controls", "directive controls" and "preventative controls".

Risk Management Concepts	Definition(s) from Cybersecurity Risk Assessment	Mapping these Risk Management Concepts to Privacy Risk Assessment
	information security (e.g. losses of confidentiality, integrity and availability), impaired operations (internal or 3rd party), loss of business and financial value, disruption of plans and deadlines, and damage of reputation. [ISO 27005]	consequences. The impact / severity level will at least depend on the type of consequence and the actor involved. The data subject may suffer privacy harms such as unwanted exposure or prejudice resulting from the loss of privacy. Any responsible party for the processing of the personal data may suffer prosecution or reputation damage resulting from failing to comply with regulations or publicity from a data breach that leads to the subject's loss of privacy. For representation, PRiAM are using an ordinal scale of risk levels, from "Very Low", through "Low", "Medium", "High" to "Very High", as defined in Table 1.
Likelihood	Chance of something happening. In ISO 27000 this is specifically the likelihood of 'consequences' of an event. [ISO 27000]	For representation, PRiAM are using an ordinal scale of risk levels, from "Very Low", through "Low", "Medium", "High" to "Very High", as defined in Table 1.
Risk	"Effect of uncertainty on objectives" [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).
Risk Level	Magnitude of a risk expressed in terms of the combination of consequences and their likelihood. Called 'level of risk' in ISO 27000. [ISO 27000]	PRiAM uses an ordinal scale, calculated from Impact and Likelihood as defined in Figure 7 and Table 1.
System	Set of applications, services, information technology assets, or other information-handling components. [ISO 27000] Synonym for "information system" [RFC 4949]: An organized assembly of computing and communication resources and procedures --	For the DARE UK PRiAM project, systemic modelling concerns a Cyber-Physical System: "a system that comprises of interacting digital, analog, physical, and human components engineered for function through integrated physics and logic" as defined by

Risk Management Concepts	Definition(s) from Cybersecurity Risk Assessment	Mapping these Risk Management Concepts to Privacy Risk Assessment
	i.e., equipment and services, together with their supporting infrastructure, facilities, and personnel -- that create, collect, record, process, store, transport, retrieve, display, disseminate, control, or dispose of information to accomplish a specified set of functions. (See: system entity, system resource. Compare: computer platform.) [RFC4949]	the NIST Framework for Cyber Physical Systems: Volume 1, Overview (Griffor et al., 2017).
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". **Problematic Data Action is used by NIST PRAM rather than Threat and Vulnerabilities:** "A data action that causes an adverse effect, or problem, for individuals" (Brooks et al., 2017). A key point to note regarding Threat for the purposes of PRiAM is that the Threat action may not necessarily be hostile, and it may be within the System of concern as opposed to external. For example, the normal processing of personal data may be considered a Threat if it has the potential to lead to adverse Consequences.
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	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 Management Concepts	Definition(s) from Cybersecurity Risk Assessment	Mapping these Risk Management Concepts to Privacy Risk Assessment
	including ICT hardware, computer software, networking, places, people and governance to reflect weaknesses that may increase the likelihood of their being affected by threats.	

The risk elements are organised into a concept structure as shown in Figure 7.

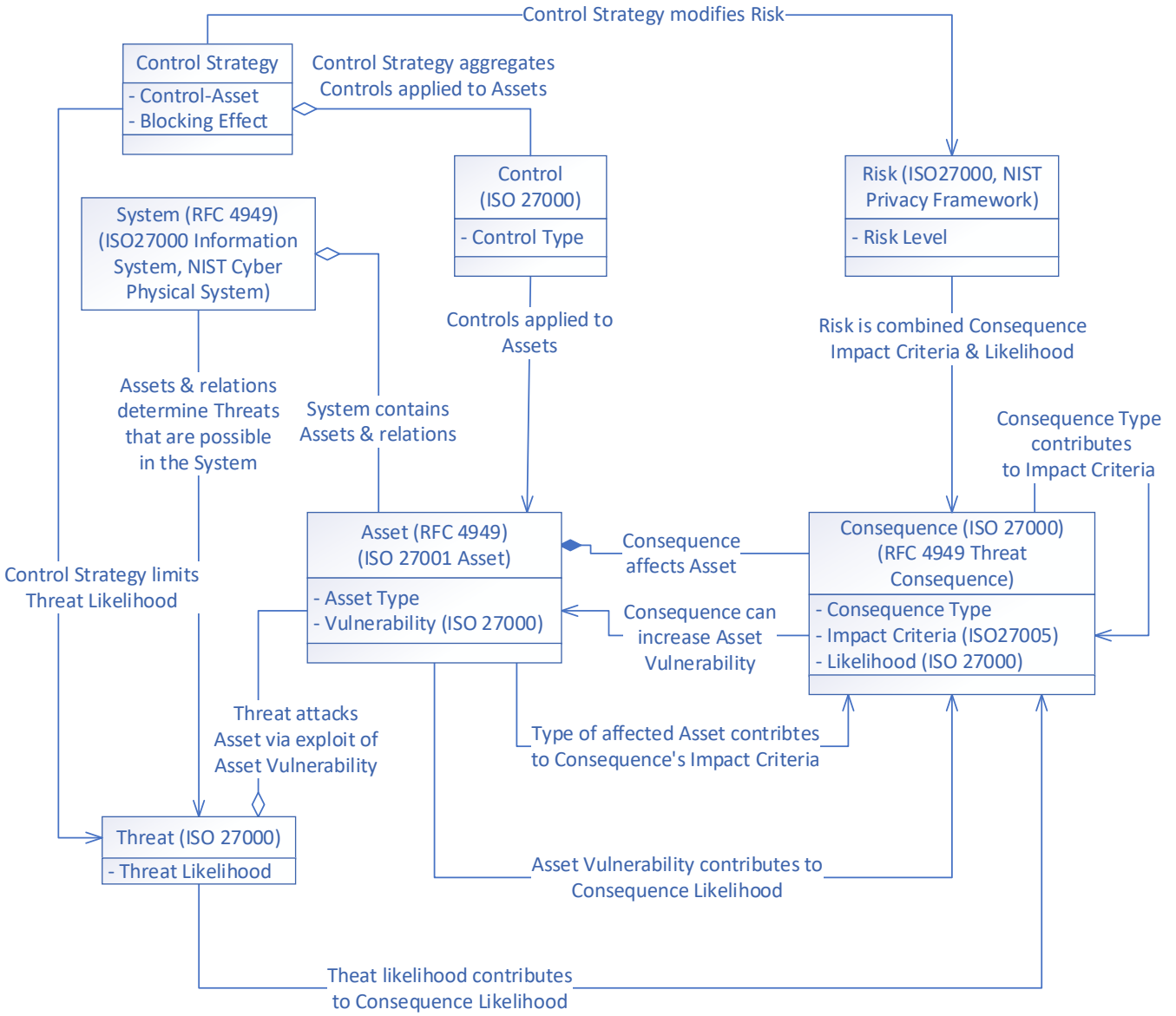


Figure 7: PRiAM Risk Modelling Concept Structure

Beginning from the System and working anti-clockwise, the key properties that determine the cause and effect of Threats on Risks are as follows.

- The System under examination is a cyber-physical system of different types of Assets and their relationships.

- Assets have properties that can be represented as resilience or weakness on the same scale, and Vulnerabilities describe weaknesses. These properties can be of different types, representing concepts such as trustworthiness[3] of actors, security of resources, reliability of components, or other desirable properties such as privacy, confidentiality, availability and integrity. In some cases, these properties represent resilience or defensive capabilities against Threats. When these properties have low values (e.g. low trustworthiness of actors), they are vulnerable; and when they have high values, they are resilient.
- Threats attack Assets by exploiting Vulnerabilities. The configuration of Assets and relationships in the System determines the Threats that are possible in the System. Each Threat has a specification of Assets and relations necessary for the Threat to be triggered (known as the Threat's "matching pattern").
- A Threat has a Likelihood determined by intrinsic factors such as its inherent difficulty, and extrinsic factors such as the motivations of actors to attack via this Threat.
- Assets have Consequences, which represent undesirable effects of Threat attacks on Assets. Examples of Consequences include Loss of Confidentiality or Loss of Privacy. Consequences may increase Vulnerabilities and expose the Asset to other Threats.
- A Consequence has a Likelihood, which is determined by the causing Threat Likelihood combined with the Vulnerabilities of the attacked Asset that are exploited by the Threat, considering the presence or absence of defensive capabilities on the Asset that lower or raise its Vulnerabilities respectively. Many Threats may lead to the same Consequence on the same Asset, in which case the highest (worse case) Threat Likelihood determines the resulting Consequence's Likelihood.
- A Consequence has an Impact Criteria level that represents the severity of the type of Consequence on the affected Asset. Impact levels are usually determined by people responsible for a System and the impact reflects the damage a Consequence would cause to their System's objectives. For example, a Loss of Confidentiality of a Personal Data Asset is likely to be regarded as "high" impact.
- A specific Consequence on an Asset has an associated Risk Level. The Risk Level is determined by the Consequence's Impact Criteria (determined by judgement) combined with its Likelihood (determined by the likelihood of its causing Threats).
- Controls modify Risk levels by introducing defensive measures to Assets. In many cases, multiple Controls are needed simultaneously on different Assets, so our approach includes the concept of a Control Strategy, which is a set of Control-Asset pairs that all need to be applied to provide the necessary resilience to block a Threat.
- Control Strategies block Threats by limiting the Threat Likelihood. A Control Strategy has a Blocking Effect, which is expressed as a level that indicates the combined effectiveness of all its Controls at blocking Threats, and the Blocking Effect determines the limits on the Threat Likelihood. The higher the Blocking Effect, the lower the Likelihood of the Threats it addresses - for example a Control Strategy with a "High" Blocking Effect puts an upper limit of "Low" Likelihood on any addressed Threats.

In summary, Threats link Vulnerabilities to Consequences (which determines Risk) and are addressed by Controls. It is often helpful to determine a cause and effect chain by investigating three types of cause-effect relationships: Asset Vulnerabilities enable Threats; Threats cause Asset Consequences (which leads to Risk); and Controls applied to Assets block Threats.

[3] Note the term "trustworthy" is defined by Dictionary.com (n.d.) as "Worthy of trust or confidence; reliable, dependable". Further, the term "trustworthy system" is described by RFC 4949 as "A system that not only is trusted, but also warrants that trust because the system's behavior can be validated in some convincing way, such as through formal analysis or code review" (Shirey, 2007).

3.2. Risk Knowledge Modelling Process

A modelling process is used to capture the knowledge regarding privacy Risks and to translate this knowledge into the suitable ISO27005 concepts, as described in Figure 8. The process of knowledge capture and engineering is based around identifying and classifying the cause and effect relationships between the elements of Risk, determining Threat Specifications under which circumstances a Threat is valid, and determining Control Strategies (collections of simultaneously applied Controls) to address Threats.

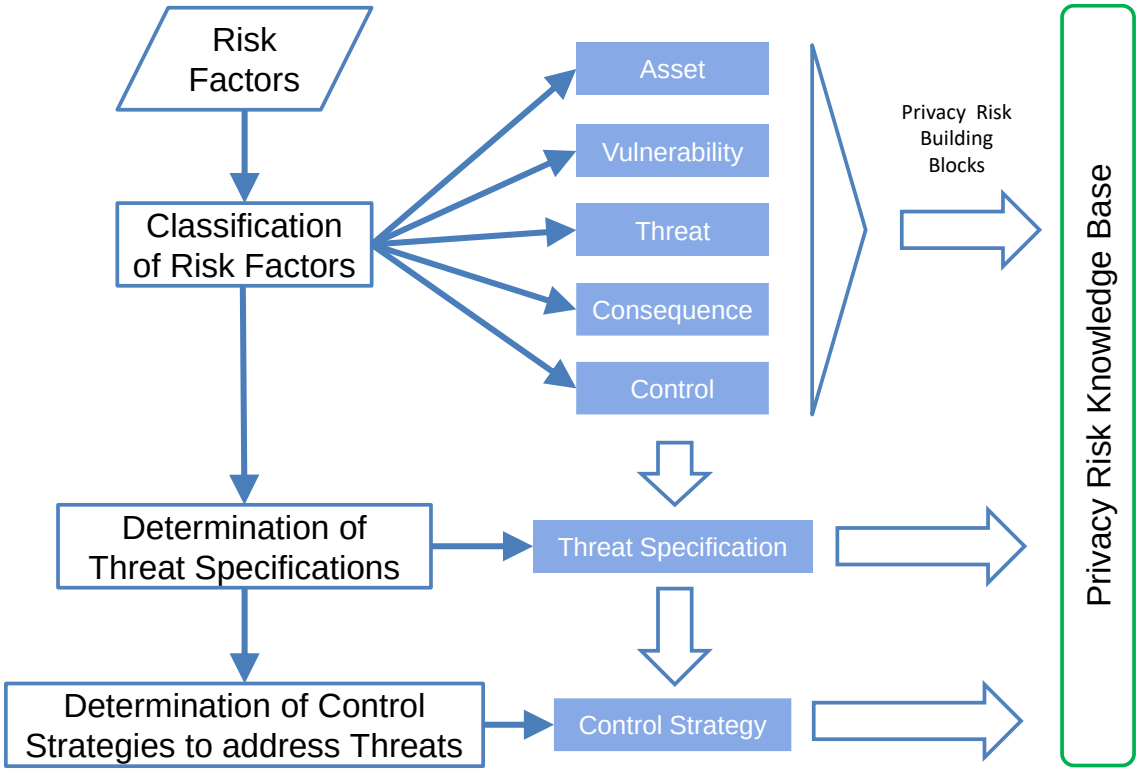


Figure 8: Privacy Risk Knowledge Modelling Process

3.2.1. Classification of Privacy Risk Factors

The first step is to take the privacy risk factors and questions, as described in the questionnaire provided as Appendix A in D2, and to identify Assets, Vulnerabilities, Threats, Consequences and Controls from them. Some heuristics have evolved through this work to help with this task.

- Data types, human roles, processing types, computer hardware, computer networks and physical spaces owned by a TRE are all common types of Asset. Data Subject is a key human role - i.e., the person whose rights will be violated if their privacy is compromised.
- In the worked example following in Section 4, possible answers to the questions in the exemplar Example Risk Factors Questionnaire (see Appendix A of PRiAM Report D2) determine candidates for Controls. Many of the questions in the questionnaire are phrased in terms of policies, procedures, techniques, technology for the specific purpose of increasing defensive capability. Other questions are phrased in terms of identifying Vulnerabilities, i.e., weaknesses that can expose an asset to a Threat.
- A key Consequence is Loss of Privacy at a Data Subject. This is because many Threats lead to Loss of Privacy, but also Loss of Privacy can lead to other knock-on consequences for the data subject, e.g., unwanted exposure or prejudicial actions against them (typical types of privacy harm). Loss of Privacy on the Data Subject can also lead to consequences for other actors, e.g., a Project Principal Investigator may suffer

losses of reputation or prosecution if their project leaks data that causes a loss of privacy in a subject whose data they are processing. Hence, because of its key positioning, the Consequence of Loss of Privacy is a key focus of this document.

For communication purposes and consistency with established privacy management methodologies, it is beneficial to classify the identified Assets, Threats, Vulnerabilities, Consequences and Controls in terms of the Five Safes. A key heuristic for this classification is that the Assets concerned in each relationship also determine to which of the Five Safes the Threats, Consequences and Controls belong:

- Data and its processing belong to Safe Data;
- People and their management belong to Safe People;
- ICT hardware, software, physical spaces and environments belong to Safe Settings; and
- egress of Data belongs to Safe Outputs.

Given the stated scoping of DARE UK PRiAM towards a data analysis project, the ISO 27005 context of scope and boundaries also can be determined by the Safe Projects. This aspect collects together concerns related to the project, including its assets, activities, governance, policy and legal status. Finally, it is worth noting that some Threats and Controls span more than one of the Five Safes because they can affect or be applied on different Assets.

We derive Assets, Vulnerabilities, Threats, Consequences and Controls from the exemplary privacy risk factor questionnaire and the associated Risk Tiers, both provided as Appendices in PRiAM Report D2. Table 3 maps the information given (the risk factors, the questions and possible answers) to Assets, Vulnerabilities, Threats, Consequences and Controls that arise from analysis of the risk factors, questions and answers. Examples of this table in use are given later.

Table 3: Risk Factor Classification Table Template

ID	Safe	Risk Factor	Express as Question	Possible Answers	Assets	Vulnera-bilities	Threats	Consequences	Controls
ID	Relevant Safe(s) of the 5 Safes	Risk Factor Name	Express the risk factor as a question	List possible answers to question	Assets identified from factor and questions	Names of Asset Properties that represent Vulnerabilities	Names of Threats that exploit Vulnerabilities and lead to Consequences	Consequences arising from Threats	Controls that can address Threats

3.2.2. Determination of Threat Specifications

Once the elements are identified, they need to be organised into Threat Specifications, which specify the Asset types concerned and their relationships, the Vulnerabilities on Assets that cause the Threats and Consequences that arise. Diagrams of these relationships have been used to capture this analysis in this document, and the notation used is shown in Figure 9.

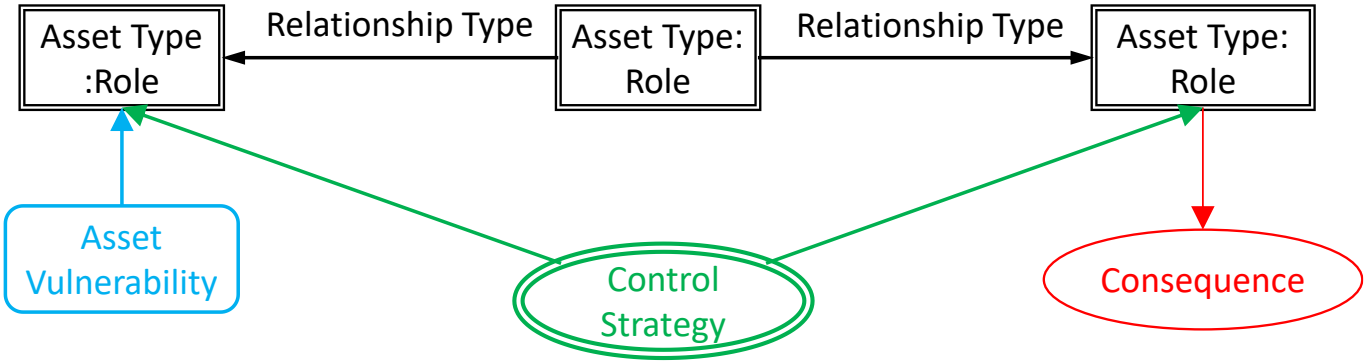


Figure 9: Threat Specification Notation

The overall picture represents one Threat. Black boxes indicate Assets and the arrows between them indicate relationships that need to be in the System Model for the Threat to be triggered. This is the so-called "matching pattern" used in the Risk Knowledge Base, named because it determines which assets and relations in the System Model need to be matched for the threat to be determined present in the system. Blue boxes indicate Asset Vulnerabilities, and Consequences are represented by red ovals. Control Strategies (collections of Controls applied at different assets, discussed next) are represented by green ovals.

3.2.3. Determination of Control Strategies to Block Threats

The final step is to map risk factors to Controls. In many cases, combinations of controls are needed to block a Threat, and for this, we use the notion of a Control Strategy, which is a set of Control-Asset pairs, indicating that different controls are applied simultaneously to different Assets. Control Strategies can also be specified using a tabular format, shown below (Table 4). Examples of this format's use are given later.

Table 4: Control Strategy Template

Safe	Threat	Control Strategy	Control	Controlled Asset	Blocking Effect	Description
Relevant Safe(s) of the 5 Safes	Threat Name	Control Strategy Name	Control 1	Asset 1	Very High to Very Low	Descriptive text of Control Strategy
			Control 2	Asset 2		
			Control 3	Asset 2		

The next section illustrates this process in action resulting in an augmented knowledge base, along with an illustration of the use of the knowledge base.

4. Worked Example of Risk Modelling

This section describes a worked example of the Privacy Risk Modelling based on PRiAM Use Case A - "Research Project Studying Complex Discharge from Hospital". We show how the Privacy Risk Knowledge Base is enhanced with privacy risk knowledge, and the how the Privacy Risk Knowledge Base is subsequently used for automated assessment in the SSM.

4.1. Use Case A Description - "Research Project Studying Complex Discharge from Hospital"

This section describes one of the use cases from PRiAM Report D1, concerning a research project studying patient discharge risks and expected departure points from hospital. The purpose of the project is to optimise patient care provision for Complex Discharge from hospital, where after discharge from hospital, the patient needs ongoing care that is provided by the community. The project depends on linking between multi-stakeholder datasets across an integrated care pathway (e.g., acute care, community care, local authority). Of the three use cases from D1, Complex Discharge is the most illustrative of the Privacy Risk Modelling. The sources of these datasets are typically from the hospital and local authority (i.e. council), as shown in Figure 10. It is important to note that the datasets used in the illustration of this document are not actual real data. Knowledge of the structure of the real datasets has informed the illustration in this document but any values are fictitious.

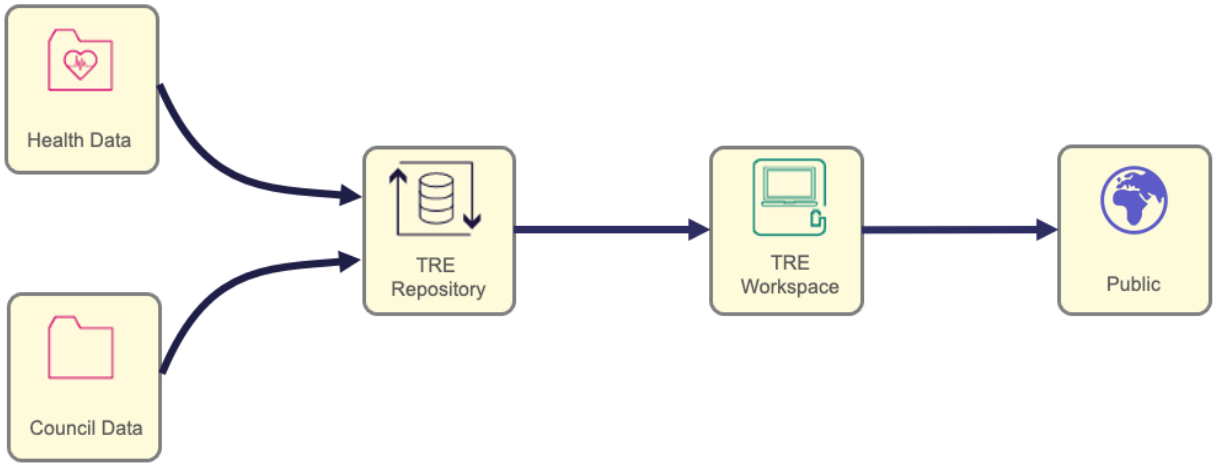


Figure 10: Complex Discharge Data Linking

Each dataset is specified through a research protocol with agreed inclusion and exclusion criteria ensuring minimisation (data is limited to that only necessary for the purpose), whilst each data provider is responsible for de-identification (removal of direct identifiers). The datasets to be linked consists of the following:

- De-identified hospital data (hospital episodes, onward care needs assessment, discharge destination postcode)
- De-identified council data (community care service usage, demographics, residency postcode)

The System Model for the scenario is shown in Figure 11.

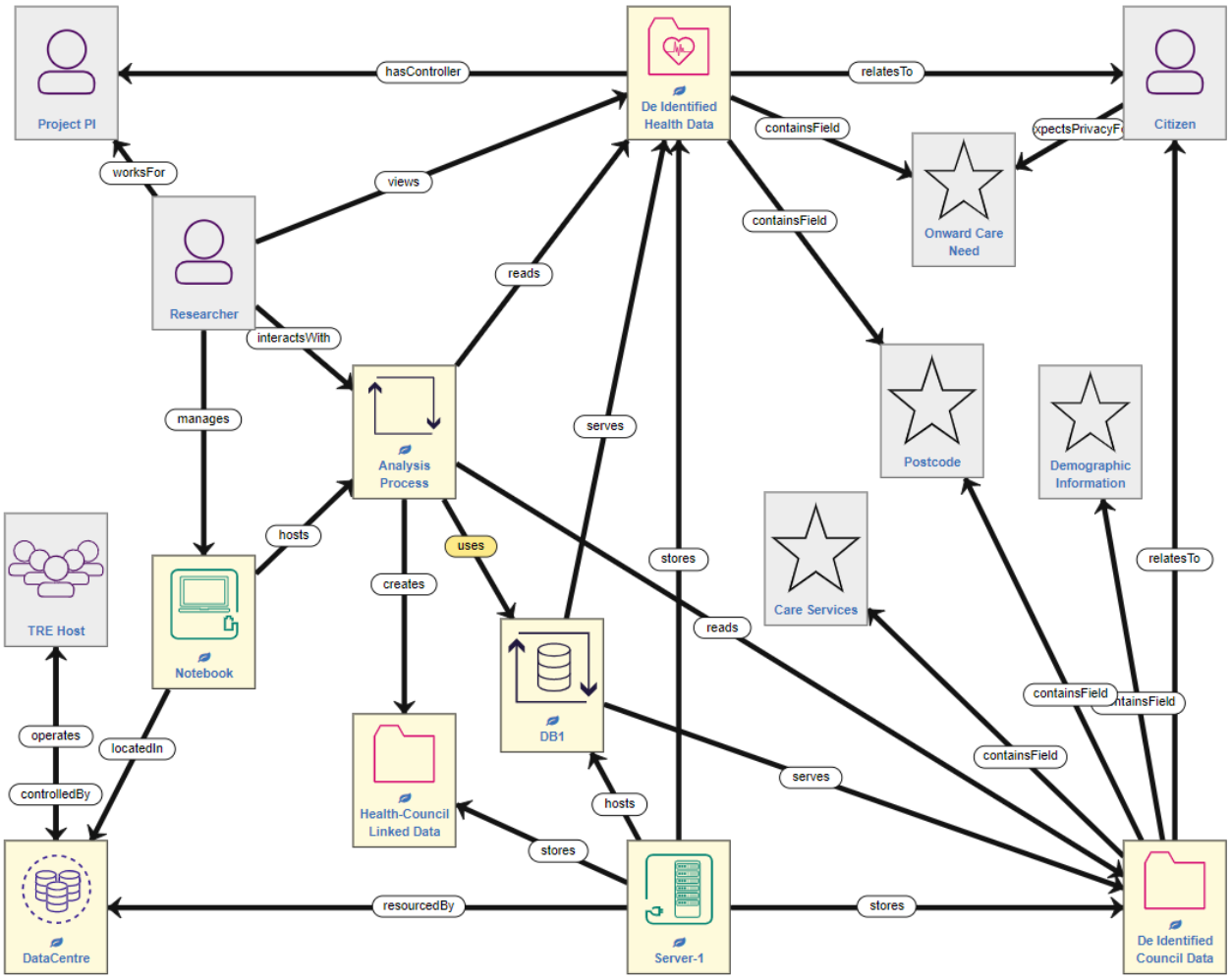


Figure 11: Complex Discharge System Model

There are two datasets, De-Identified Health Data (top centre) and De-Identified Council Data (bottom right). For the purposes of this discussion, we assume that the relevant permissions and licenses to use both datasets has been acquired, and that all conditions for the licenses are satisfied as a pre-requisite to any kind of processing for these datasets.

In the System Model, the datasets are connected to the data subject, the Patient (top right) via the "relatesTo" relation. This means that the datasets either contain personally identifiable information or derive from personal information about the subject. Here, the datasets are de-identified so do not contain personally identifiable information but still may contain indirect identifiers. The Health Data contains two Data Fields (denoted by a star): Onward Care Need and Postcode (Postcode is an indirect identifier). The Onward Care Need is sensitive information and there is an expectation of privacy on the part of the subject for this field. The Council Data contains three fields, Postcode, Care Services and Demographics.

The TRE is hosted in a DataCentre, which is operated by the TRE Host. The datasets are stored in DB1, which is hosted on Server-1 inside the DataCentre. A Researcher interacts with a software process named Analysis Process that links the data. The Analysis Process is hosted on a Notebook within the physical space of a TRE DataCentre. The Researcher works for the Project PI (Principal Investigator), who has overall responsibility for the data linking & analysis.

The data is linked via the postcode - this field is common to both datasets and enables the study of optimisations across the care pathway by mapping the onward care needs to the care services and demographics. Linking will

enable studying the association between care services and care demand via the postcode. This enables understanding of the care needs vs the already-provided care services in a local area, along with the types of people needing the care. This insight can be used to identify gaps in local care provision and to support prioritisation of resources.

While the datasets are de-identified, the data is not aggregated, so each row represents one person. The identity of the person is not explicit due to the prior removal of direct identifiers and data minimisation before ingress into the TRE but still contains indirect identifiers in the form of the *postcode* and *demographic information*, which are needed for the purposes of the study. Examples are given below of the two datasets in Table 5 and Table 6.

Table 5: Example Onward Care Needs

Postcode	**Onward Care Need [Sensitive]**
SO16 7NS	Type 1 Diabetic
SO16 7NS	Wheelchair Access
SO14 3FU	PTSD
SO18 6YS	Hepatitis Type C
SO14 7TB	Hypertension

Table 6: Example Care Services

Postcode	**Care Services**	**Demographic Information**
SO16 7NS	Disability House Modifications	Male, 75-80 years, Caribbean
SO14 3FU	Counselling	Male, 30-35 years, English
SO18 6YS	Home Help	Female, 75-80 years, Chinese
SO14 7TB	Diet Management	Female, 35-40 years, Pakistani

When the datasets are joined by the *postcode*, although it is not possible to be certain that the same individual is represented in both datasets, it is possible to infer from the combined information whether the information represents the same person, thus re-identifying them from the indirect identifiers. The specificity of the *postcode* is an important factor to this re-identification. In the UK, a full Postcode typically represents 15 properties[4], and if we assume 3 residents per property, this totals 45 people who could be identified. Applying demographic information such as age range, ethnic group and gender can quickly reduce this number to the point where the chances of re-identification are strong. If this information is linked with sensitive information for which privacy would be expected, such as the onward care need, then not only has re-identification occurred from indirect identifiers, but a privacy violation has occurred.

For example, in SO14 3FU, there is a patient who needs support for Post-Traumatic Stress Disorder (PTSD). There is also a record in care services for Counselling at the same Postcode, so there is a possibility that these refer to the

[4] As stated by IdealPostcodes (Kurdi, 2021).

same person because the Counselling services may be for the PTSD. Combining the Postcode with the demographic information narrows the field of candidate individuals - i.e. an English male aged between 30 and 35 years living at SO14 3FU, who has been provided with counselling services. Given this combined information, it is possible to reasonably easily single out an individual and link them with the sensitive information that they have PTSD.

4.2. Risk Knowledge Modelling

4.2.1. Privacy Risk Factor Classification

Several privacy risk factors are encoded into a questionnaire for the Risk Tiers (reported in PRiAM Report D2). This section discusses how to map a subset of these risk factors into the elements needed for risk modelling (Assets, Vulnerabilities, Threats, Consequences, Controls). These are shown in Table 7, Table 8 and Table 9 on the following pages. These tables are merely illustrative examples relevant to the scenario and are by no means exhaustive. The Threats, Consequences and Controls are also classified in terms of the Five Safes as per the process described in Section 3.

Is Considered Sensitive (Safe Data): concerns sensitive information in the form of the onward care need, is present in the scenario, so this presents a risk of unauthorised processing, with severe consequences of privacy loss for the data subject and a potential breach of data protection law for the project Principal Investigator. Options for controlling the risk are presented in Table 7. The processing is necessary and given our assumption above that the necessary permissions for the data has been acquired, the appropriate controls are “Confirmation of Necessity to Process” and “Relevant Permissions”. There is an alternative Control - “Cessation of Processing & Deletion of Data” covering the case where processing is stopped immediately, and data deleted.

Presence of Direct Identifiers (Safe Data): concerns the presence of direct identifiers in the data. For the scenario case, it is assumed that the data is already de-identified by removal of direct identifiers before ingress into the TRE, and appropriate controls are presented in row 2.

Presence of Indirect-Identifiers that can Single Out (Safe Data): covers the presence of indirect-identifiers and Threat of reidentification arising from them in the data, where the indirect-identifiers in the scenario are the Postcode and the demographic information. The exemplar threat shown is “Reidentification via Viewing Data”, which can lead to a loss of privacy of the subject by the researcher browsing the data and interpreting the information within it. This threat may be addressed by controls such as full de-identification (i.e. removal of any indirect identifiers) or tokenisation (obfuscating data fields so their meaning is obscured, but still permitting linking if the tokenisation uses the same key for all datasets to be linked).

More Data than Required for Project (Safe Projects): covers the checking and minimising of data so that only data necessary for the justified purposes of the project are loaded into the TRE. Controls of proportionality assessments and data minimisation are suggested to addresses the Threat of excessive data for the purpose of the project. For the Complex Discharge case, we have assumed that the data is minimised for the purposes of identifying gaps in local council care provision, as per the examples in Table 5 and Table 6. Row 4 in Table 7 is classed as “Safe Projects” because even though it concerns checking and minimising data, the minimisation needs to be determined by the purpose and needs of the project and consistent with it.

The factors above are data ingress factors concerning respectively: the assessment of data for sensitive information, direct identifiers, indirect identifiers and the minimisation of data. These are ingress factors because they should be checked or applied as necessary before the data is loaded into the TRE.

Analytics Experience of Researchers (Safe People): covers inappropriate activity by researchers due to lack of training or awareness that constitutes experience of data analytics, that can result in inadvertent or accidental breaches of privacy for the subject due to the lack of experience on the part of the researcher. Controls include

appropriate training, simply not using an inexperienced researcher (deselection), or limiting the researcher's access to personal or sensitive data (which crosses over into Safe Data).

Activity Logging (Safe Settings): covers audit logging of researchers' activity and describes numerous controls in the form of the types of activity that should be logged. This addresses inappropriate researcher activity, via encouragement of researchers to behave appropriately.

Data Linkage Policy and Control (Safe Settings): covers the linking policy and controls of the TRE. This is directly relevant to the Complex Discharge scenario, as it relies on linking the Health Data concerning onward care needs from the hospital and the Council Data concerning provided care services to citizens in the local community. Here, linking is via the shared field of Postcode. The threat is actually the activity of linking, which can lead to reidentification and privacy violation even if the datasets contain no direct identifiers, and in the Complex Discharge scenario there are indirect-identifiers (notably the shared Postcode, which is important to making the link between the datasets). Suggested controls include generalisation of data fields and researcher training to make them aware of the risks of reidentification and the consequent privacy losses for the data subject.

Data Egress (Safe Outputs): covers the data egress policy of the TRE, which has the specific intention of avoiding the release of PII and the consequent loss of privacy on the data subject. Numerous controls are suggested in the possible answers, which have different levels of safety, ranging from "no restriction" (least safe) upwards.

Table 7: Risk Factors - ISO27005 Element Mapping - Safe People, Safe Data, Safe Projects

ID	Safe	Risk Factor	Express as Question	Possible Answers	Assets	Vulnerabilities	Threats	Consequences	Controls
1	Safe Data	Is Considered Sensitive	Does the data have any information that could be regarded as sensitive?	Yes; No	Data	Sensitive Data	Unauthorised Processing of Sensitive Data	High Impact Loss of Privacy for Data Subject; Regulatory Violation for Project PI	Confirmation of Necessity to Process; Applicable Permissions; Cessation of Processing & Deletion of Data
2	Safe Data	Presence of direct identifiers	Are all direct identifiers removed before or during data ingress?	Yes; No	Data, Data Field	Anonymity of Direct Identifiers	Unauthorised processing of personal data	Loss of Privacy for Subject; Regulatory Violation for Project PI	De-Identification (e.g. via redaction or tokenisation); Confirmation of de-identified status; Cessation of processing & deletion of data
3	Safe Data	Presence of indirect-identifiers that can single out	Does the data have indirect information that could contribute to identification? (e.g. demographic information, Postcode, etc)	Yes; No	Data, Data Field	Anonymity of Indirect Identifiers	Reidentification from Viewing Sensitive Data	Loss of Privacy for Data Subject	De-Identification of Data; Redaction of Data Fields; Tokenisation of Data Fields; Cessation of Processing & Deletion of Data
4	Safe Projects	More Data than Required for Project	Does the data have more information than is strictly needed for the project?	Yes; No	Data	Dimension	Excessive Data for Purpose	Loss of Privacy for Subject; Regulatory Violation for Project PI	Proportionality Assessment Before Usage; Data Minimisation; Cessation of Processing & Deletion of Data
5	Safe People	Analytics Experience	Does the researcher have experience working with noisy real data?	Yes; No	Researcher	Researcher Experience	Untrained or Unaware Researcher	Unsafe Researchers; Loss of Confidentiality on Data; Loss of Privacy to Data Subject	Deselect Researcher; Train Researcher; Limit Researcher Access

Table 8: Risk Factors - ISO27005 Element Mapping - Safe Settings

ID	Safe	Risk Factor	Express as Question	Possible Answers	Assets	Vulnerabilities	Threats	Consequences	Controls
6	Safe Settings	Activity logging	What information about the researcher's activity is logged when they access data/environment?	All queries to the data; All keystrokes & mouse activity; Surveillance of users in the setting; Access to workspace - where from IP, location etc; Data brought in, if any; Data taken out; Compute deployed; Environment configuration; All code run against the data	Researcher	Researcher Trustworthiness	Inappropriate Researcher Behaviour	Unsafe Researchers	Activity logging (per answers)
7	Safe Settings	Data Linkage Policy & Control	How is linking with other datasets controlled?	No linkage allowed; Linking only approved datasets; Legal contracts to prevent re-id/reuse/linkage; Technical Controls for re-id prevention; No restrictions on linkage	TRE, Data	Anonymity of linking fields & data	Linking of Sensitive Data	Loss of Privacy to Data Subject	Generalisation of Data; Researcher Training; (plus possible answers)

Table 9: Risk Factors - ISO27005 Element Mapping - Safe Outputs

ID	Safe	Risk Factor	Express as Question	Possible Answers	Assets	Vulnerabilities	Threats	Consequences	Controls
8	Safe Outputs	Data Egress	What are the controls and processes for taking data outside the environment?	Approved function outputs can be directly taken out; All statistics are automatically made differentially private with a reasonable epsilon; No row level information is allowed, only aggregate information after manual review; All output needs manual review; No restrictions on output; Researchers never access data but they get the outputs (opensfely); Only certain approved & vetted users can take the data out; Airlock and manual approvals; PII scanning	Data	Anonymity	Release of PII	Loss of Privacy to Data Subject; Regulatory Violation for Project PI	Egress Controls (as per answers)

4.2.2. Threat and Control Strategy Specification

In the DARE UK PRiAM project, we are extending the SSM Risk Knowledge Base to include threats that may result in "Loss of Privacy to Data Subject" by modelling privacy risk factors identified within the Five Safes. The selected risk factors and associated threats are relevant to the use case scenario but also widely applicable to projects depending on cross council research datasets. These include:

- ***Presence of indirect identifiers that can single out*** and the threat from "Reidentification from Viewing Sensitive Data". NIST PRAM (2019) considers 're-identification' as a problematic data action that could result in "problems such as discrimination, loss of trust, or dignity losses".
- ***Data Linkage Policy and Control*** and the threat from "Linking of Sensitive Data"

These have been modelled as Threat Specifications, and each threat is described next. The Threat Specification includes a matching pattern (the configuration of assets and relations), along with the asset vulnerabilities that trigger a threat, that lead to consequences on affected assets. The Threat Specification also includes Control Strategies that may be applied to block the threat.

4.2.2.1. "Loss of Privacy to Data Subject" due to "Reidentification from Viewing Sensitive Data"

The threat specification for "Loss of Privacy to Data Subject" due to "Reidentification from Viewing Sensitive" threat is shown in Figure 12.

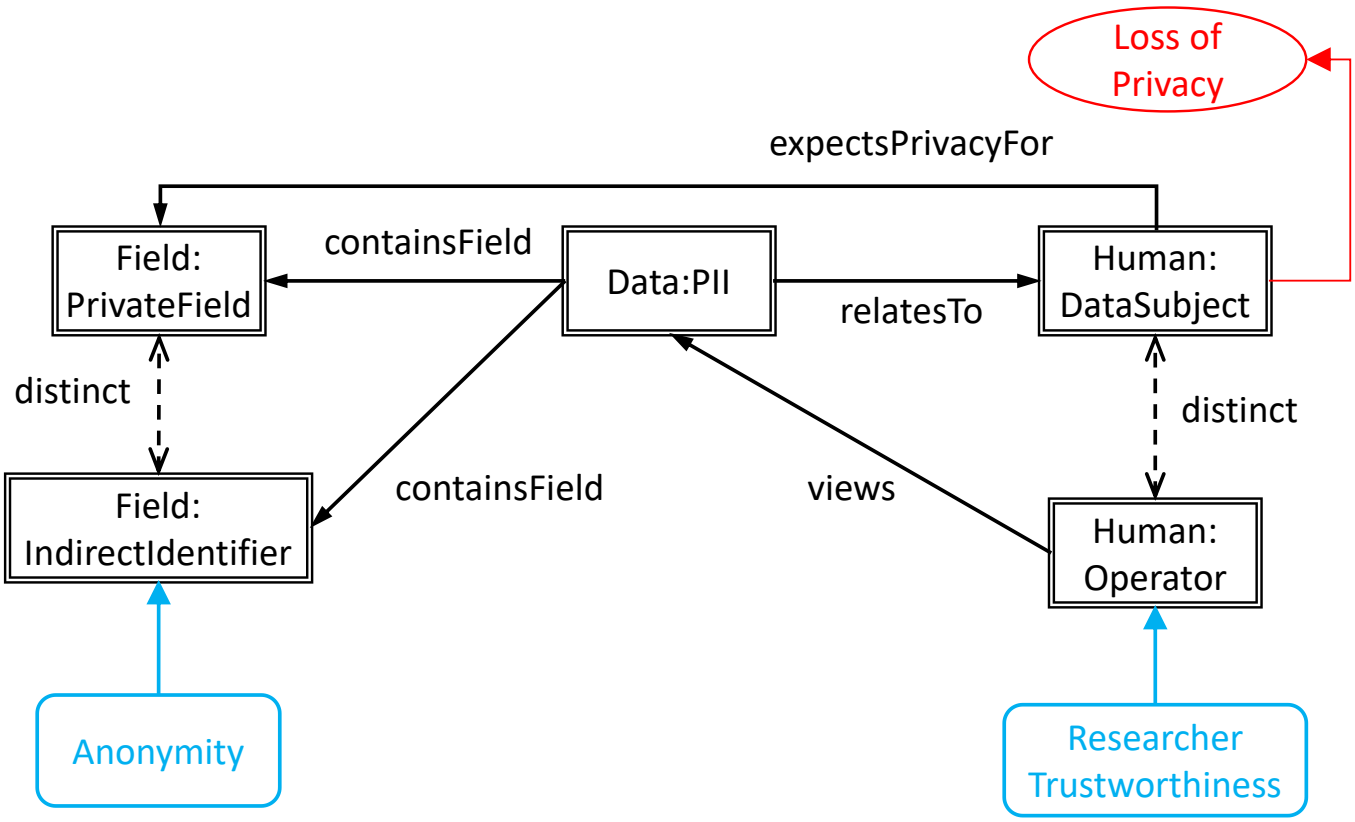


Figure 12: Threat Specification for "Loss of Privacy to Data Subject" due to "Reidentification from Viewing Sensitive Data"

The specification denotes personal data (Data:PII) for which there are two fields, a sensitive field (Field:PrivateData) for which there is an expectation of privacy on the part of the data subject; and an indirect identifier (Field:Indirect Identifier) that may have anonymity vulnerabilities. A researcher (Human:Operator) can view the data and by seeing both the indirect identifier and the sensitive field may be able to reidentify a data subject. The specification is described as follows:

- There are two Human Asset types in the model, one in the role of "Data Subject" and another in the role of "Researcher". There is a "distinct" relationship between the two Humans, indicating that the same element on the System Model cannot match both the Researcher and the Data Subject.
- The consequence of the threat is denoted by the red oval leading out of the Data Subject. This refers to the Loss of Privacy that occurs on a Data Subject if the threat is triggered.
- There is a Data Asset that is Personally Identifiable Information (PII)[5]. This is denoted as personal data by the relation "relatesTo" with a Data Subject.
- The PII data contains two Fields, PrivateData and Indirect Identifier. PrivateData and Indirect Identifier are distinct, i.e. the same element on the System Model cannot match PrivateData and Indirect Identifier.
- PrivateData corresponds to the Onward Care Need in the scenario, and the Data Subject expects privacy for this field. The Onward Care Need is therefore sensitive information.
- Indirect Identifier corresponds to the Postcode. The blue box Anonymity leading into Indirect Identifier is a vulnerability that contributes to the likelihood of the Threat. If Anonymity of the Postcode is low, then there is a high vulnerability of exposure leading to the Loss of Data Subject Privacy.
- The Human Operator views the PII Data. The Research has a vulnerability "Researcher Trustworthiness" - i.e. the level upon which the Researcher can be relied upon to act appropriately.

The Control Strategies that have been derived from the controls identified in Section 4.2 are described in Table 10 and shown in Figure 13. There are five Control Strategies, corresponding to the major rows in Table 10.

Table 10: Control Strategies for "Reidentification from Viewing Sensitive Data" Threat

Safe	Threat	Control Strategy	Control	Controlled Asset	Blocking Effect	Description
Safe People	Re-ID from Viewing Sensitive Data	Privacy Aware Practice	Reidentification Awareness Training	Operator	High	Operators who can access potentially sensitive and potentially personally identifiable data are trained to be aware of the possibility for reidentification given indirect identifiers and encouraged to follow good practices via audit logging of their activities in handling the data.
			Activity Logging	Operator		
Safe Data	Re-ID from Viewing Sensitive Data	Indirect Identifier Tokenisation	Tokenisation	Data Field	High	Tokenisation of indirect identifiers. Replacing their values with other values that have no identifiable meaning but can be mapped back to the original values with additional information.

[5] This also termed 'personal data' as defined by Article 4(1) of the GDPR.

Safe	Threat	Control Strategy	Control	Controlled Asset	Blocking Effect	Description
Safe Data / Safe Settings	Re-ID from Viewing Sensitive Data	Disable View Access	Disable View Access	Data	High	All view access to data is disabled for Data and Operator, preventing them from viewing the data.
Safe Data / Safe Settings	Re-ID from Viewing Sensitive Data	Display Redaction	Display Redaction	Data Field	Medium	Removal of sensitive data elements from displayed versions of data.
Safe Data / Safe People	Re-ID from Viewing Sensitive Data	Privacy Aware Practice And Tokenisation	Reidentification Awareness Training	Operator	Very High	Combined Control Strategy of Tokenisation of data fields, Privacy Awareness Training and activity logging on Researchers accessing the data.
			Activity Logging	Operator		
			Tokenisation	Data Field		

Each Control Strategy has one or more pairs of a Control and a Controlled Asset - this denotes the type of control and the asset to which it should be applied - for example the Control Strategy "Privacy Aware Practice" has two pairs - "Reidentification Awareness Training" applied to the Operator (in the scenario the Operator is the Researcher) and "Activity Logging" also applied to the Operator.

Any of the five Control Strategies will limit the likelihood of the threat's consequences. The combined effects of all the controls on assets in the Control Strategy results in the Blocking Effect, which acts as a limiter to the likelihood of the threat's consequences occurring - higher blocking effects result in lower consequence likelihoods.

Some of the Control Strategies are combinations of others. For example, the last Control Strategy in Table 10 is "Privacy Aware Practice and Tokenisation" - a Strategy that includes all the Control-Asset pairs of two other Control Strategies earlier in the table. Because there are more controls applied that affect the likelihood of the Threat in different ways, the Blocking Effect of the combined Control Strategy is higher than that of the individual Control Strategies.

The Control Strategies for the "Reidentification from Viewing of Sensitive Data" threat are shown in and are summarised for the scenario as follows.

- Display Redaction applies to PrivateData - the sensitive field describing Onward Care Needs
- Tokenisation applies to Indirect Identifier - the Postcode, obfuscating its content to a viewer (the Researcher playing the role of Operator)
- Disable View Access applies to Data - the PII data, disabling any Operator from viewing it.
- Privacy Aware Practice applies to the Operator (the Researcher), giving them training regarding reidentification risk and logging their activity.
- Privacy Aware Practice and Tokenisation applies to both Indirect Identifier (Postcode) and the Operator (Researcher), applying tokenisation to the Postcode and training and activity logging to the Researcher. This

is the most effective Control Strategy as it combines different controls on different assets that together provide defence in depth to the threat and therefore lowers the likelihood of its consequences.

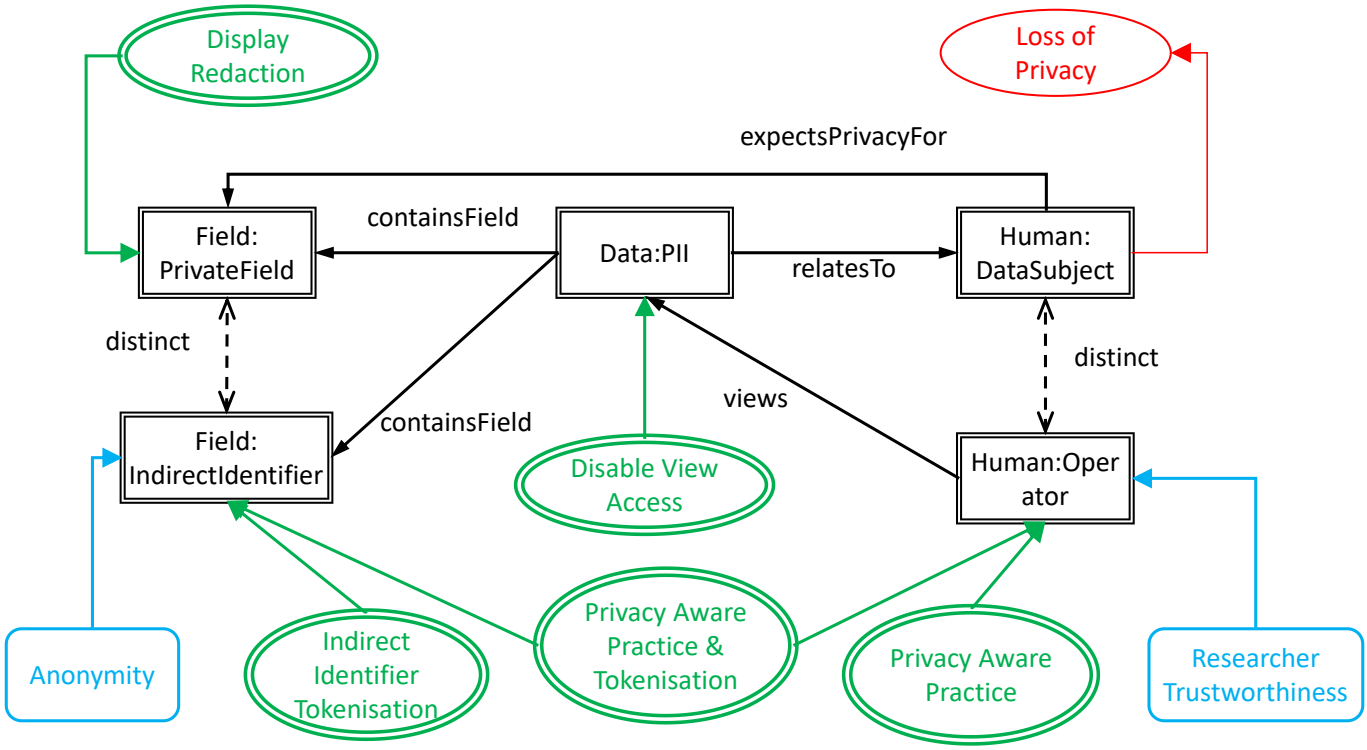


Figure 13: Control Strategies for "Privacy Loss via Viewing of Sensitive Data" Threat

4.2.2.2. Threat & Control Specification for "Reidentification & Privacy Loss via Linking Data "

The "Reidentification & Privacy Loss via Linking Data" Threats Specification is shown in Figure 14. Its matching pattern is an expanded version of the specification for the threat from Figure 12, and the additional assets are highlighted in light grey.

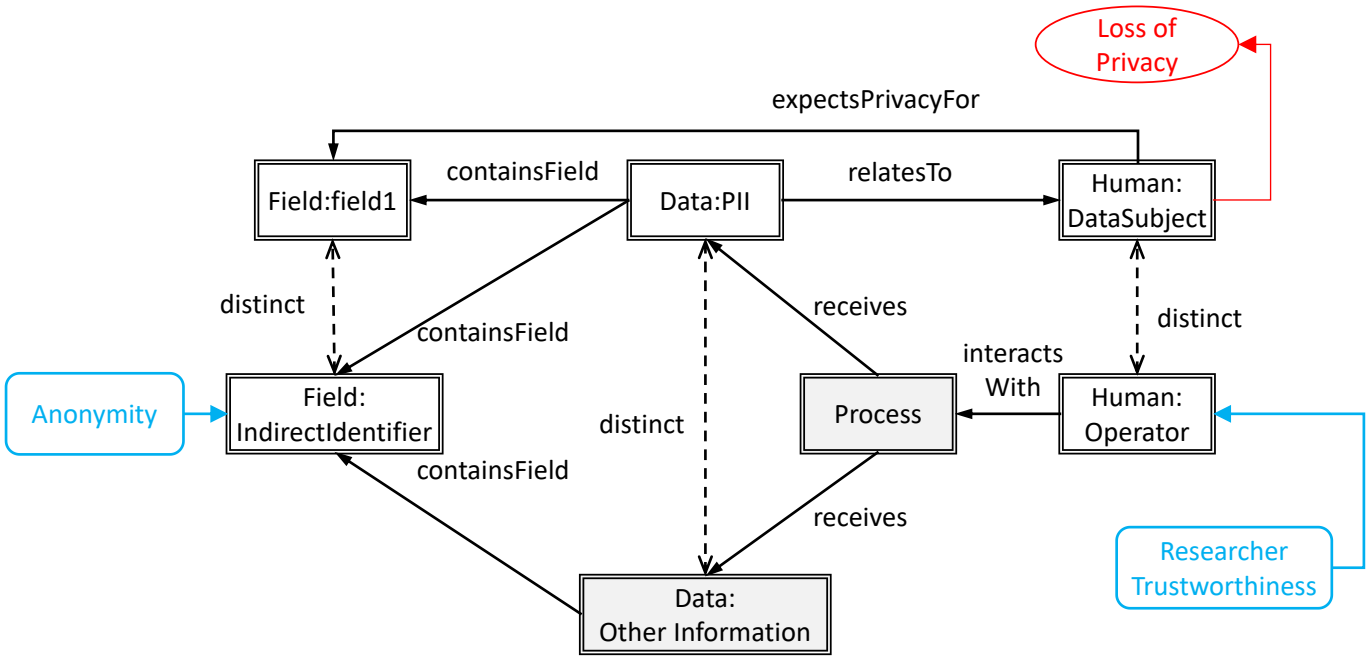


Figure 14: "Reidentification & Privacy Loss via Linking Data" Threat Specification

The additional assets denote a process which receives two distinct datasets (PII and Other Information) that both share the same field (Indirect Identifier), and because they share the same field, the two datasets may be linked. Though this linkage, the content of Other Information may be used to identify data subjects and violate their privacy via the link to PrivateData (the sensitive field with an expectation of privacy).

Two Control Strategies derived from the controls suggested in Table 7, Table 8 and Table 9 are described below in Table 11.

Table 11: Control Strategies for "Linking Data" Threat

Safe	Threat	Control Strategy	Control	Controlled Asset	Blocking Effect	Description
Safe Data	Linking Data	Data Generalisation	Generalisation	Data Field	High	Application of generalisation on field so its values are less specific. Typically used in privacy protection, so the data covers a larger set of potential people, resulting in lower risk of identification of a single person. Examples include truncating a Postcode or substituting a numerical age value with bands of age ranges.
Safe Data / Safe Settings / Safe People	Linking Data	Generalisation And Privacy Awareness And Logging	Generalisation	Data Field	Very High	Combined Control Strategy of Generalisation of data fields, Privacy Awareness Training and activity logging on Researchers accessing the data.
			Reidentification Awareness Training	Operator		
			Activity Logging	Operator		

As previously, any of the two Control Strategies will affect the likelihood of the threat's consequences. The first Control Strategy has a single control-asset pair and involves Generalisation of the common field shared between the two datasets. The second Control Strategy ("Generalisation And Privacy Awareness And Logging") is a combined Control Strategy that includes a Generalisation control, plus those described previously of "Reidentification Awareness Training" and "Activity Logging". As previously, the combined Control Strategy has a higher blocking effect than the simpler Control Strategies. The Control Strategies are shown pictorially in Figure 15.

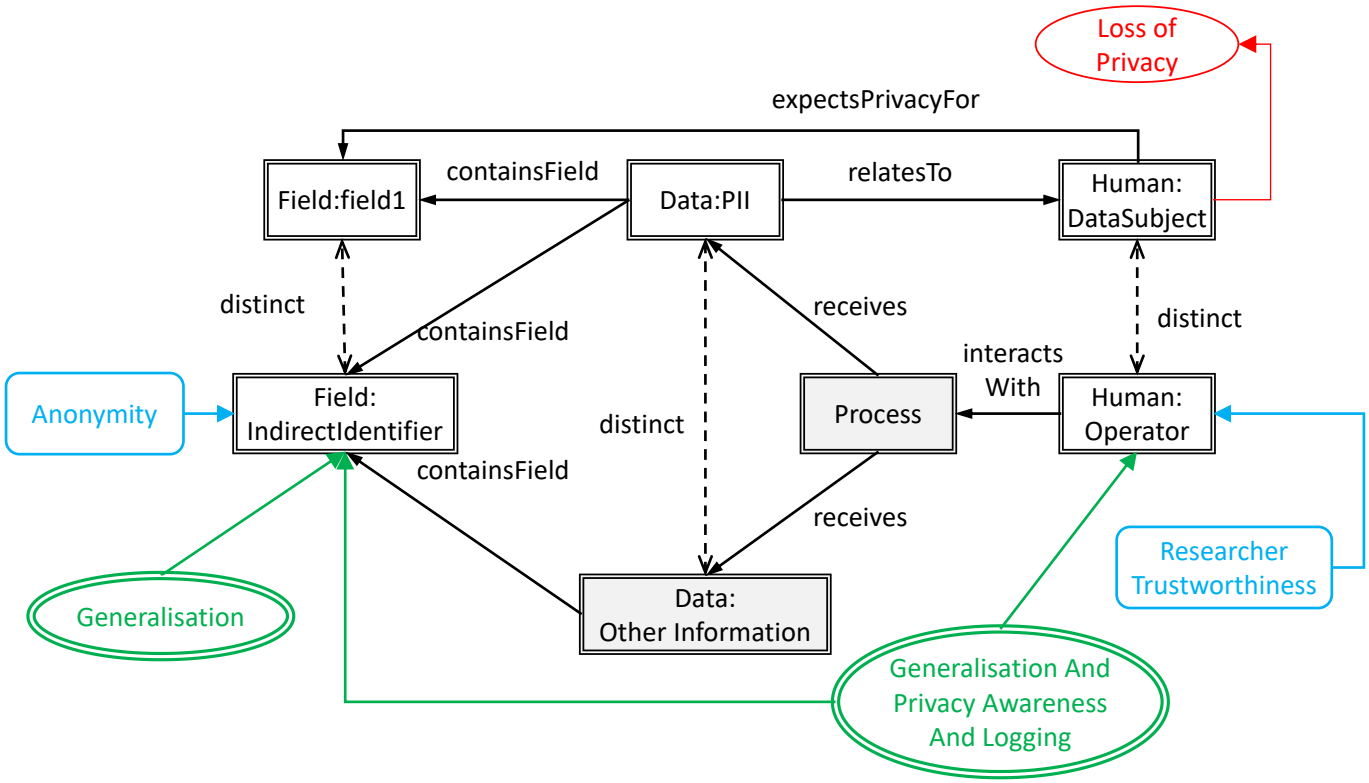


Figure 15: Control Strategies for "Linking Data" Threat

In the Use Case scenario, the PII Data is the De-identified Health Data that contains the patient's Postcode (Indirect Identifier) and the sensitive information of their Onward Care need in PrivateData. Other Information is the De-Identified Council Data, which contains the shared Postcode in Indirect Identifier, as well as any other information (e.g. the demographic information). The demographic information is not needed as an additional field for this pattern the presence of the Other Information data encapsulates this.

The Control Strategies' application in the Complex Discharge Scenario is briefly summarised below.

- Generalisation applies to Indirect Identifier - the Postcode data. As discussed above, a full 7-character Postcode identifies approximately 15 properties. If this Postcode is generalised to use the only first 4 characters (known as the Postal District), it identifies in the order of 8000 properties, hence the data is much less specific with a consequent reduction in the likelihood of reidentification, even given that there may be other specific data in the "Other Information".
- "Privacy Awareness" and "Logging" apply to the Operator (the Researcher) as in the previous threat.

4.3. Risk Assessment

The assets, threats, risks and controls have been encoded into the knowledge base of the System Security Modeller (SSM) Toolkit including the threat and control specifications identified from analysis of risk factors above (See section 4.2.2). The knowledge base operates on a System Model to provide automated ISO27005 risk assessment. This subsection describes an illustration of the new knowledge applied to a System Model describing the Complex Discharge use case, and we use screenshots of the SSM with text describing the risk assessment process to demonstrate how a TRE operator can interactively assess risk and apply controls to reduce risk to an acceptable level.

4.3.1. Defining the System Model

The first step is to define a System Model for the use case scenario by placing assets on the canvas and specifying relationships between them. System models are considered as covering four main aspects:

- Application layer: Data assets, Processes handling the data including data storage services and client processes to and from which data is communicated, plus data subjects.
- Network layer: the physical and virtual devices providing the communication and process execution environment, and users involved in managing them.
- Physical layer: locations of physical devices, which may be accessed by users, and the stakeholders who manage physical security of those locations.
- Regulatory layer: defines who is responsible for the operation of different elements of the system and provides links to the jurisdiction (i.e. body of regulations) under which they operate.

The 'layers' are not encoded in any way and so are not in any sense 'enforced' by SSM. They just provide a convenient conceptual way that is often helpful when trying to decide how best to model a given system.

In our use case scenario, we model a research analysis process that links two datasets hosted on a server within a data centre. We consider application layer (data assets, process, storage, data subject), Network Layer (server, notebook) and Physical Layer (stakeholders, data centre space), and Regulatory Layer (data controller). The System Model is simplified in this example to focus on privacy risk assessment by co-locating all network infrastructure, data and processes. In practice the network and physical layer would include application assets (data and process) within a distribution computing environment (i.e., access to a remote cloud TRE by a researcher). This is supported by the SSM but largely relates to analysis of security related risks necessary to establish a Safe Setting.

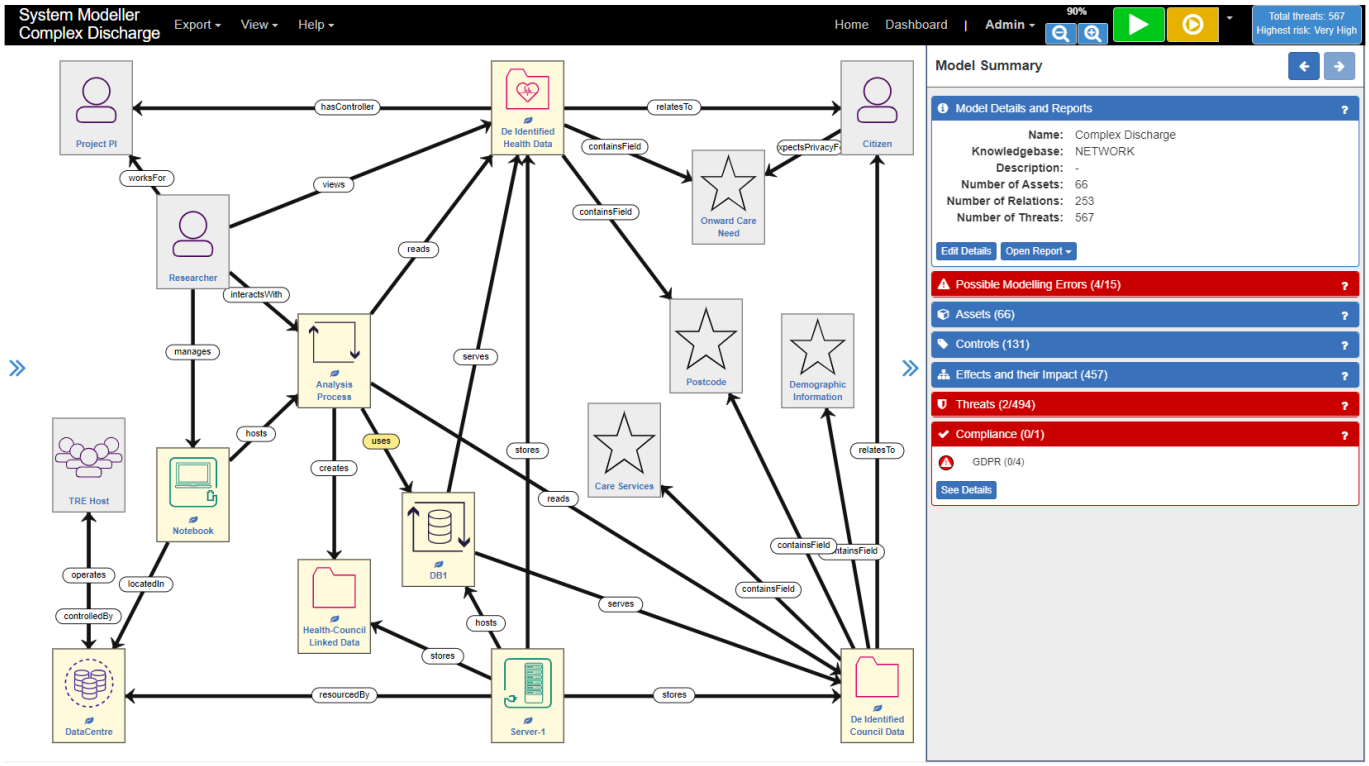


Figure 16: The Complex Discharge Project SSM System Model

Figure 16 shows the System model for scenario modelled in the SSM including the following assets:

- Application (Citizen, De-identified Health Data, De-identified Council Data, Health-Council Linked Data, Analysis Process, DB1)
- Network (Server-1, Notebook)
- Physical (DataCentre, TRE Host, Researcher, Project PI)

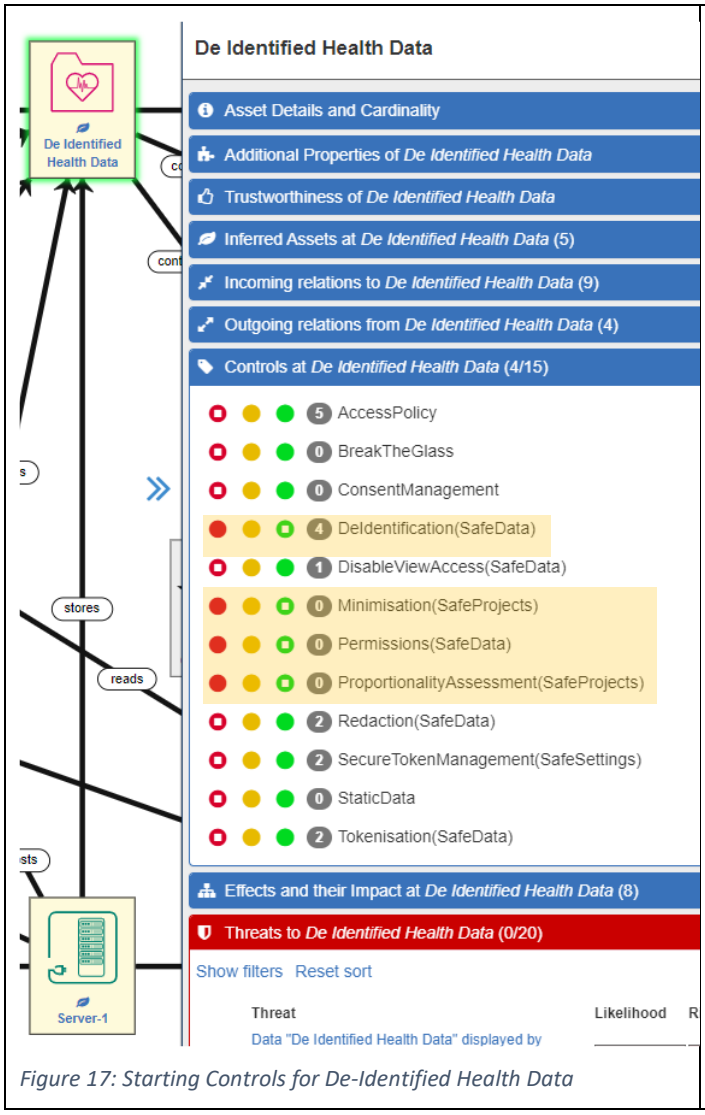


Figure 17: Starting Controls for De-Identified Health Data

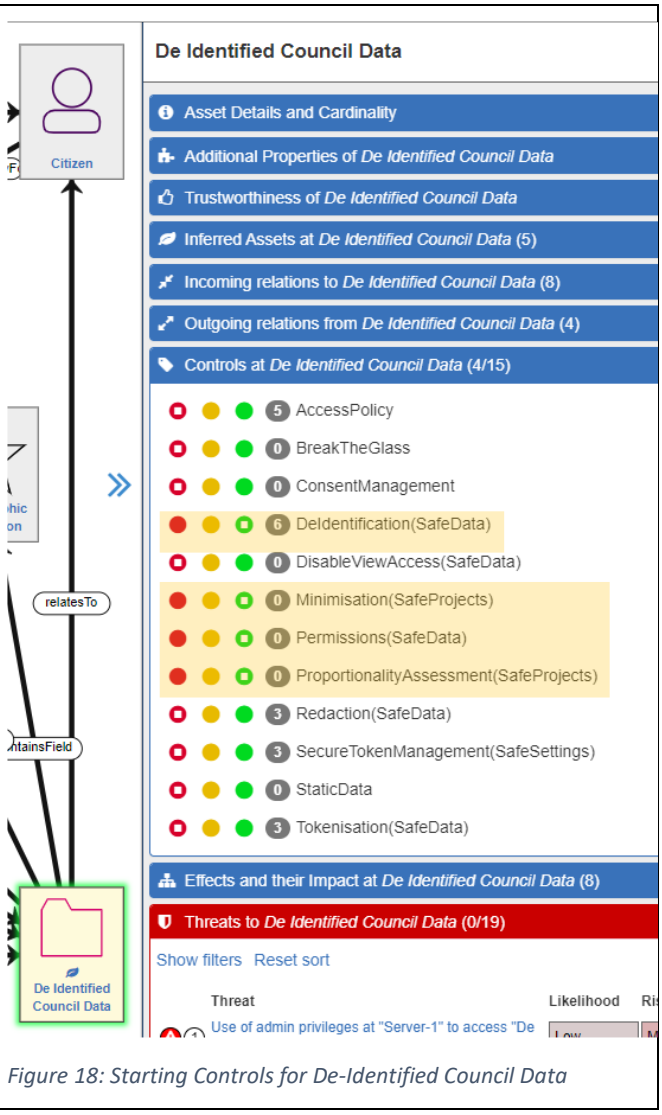


Figure 18: Starting Controls for De-Identified Council Data

Initially the data ingress policy is specified by applying controls to data assets ("De-Identified Health Data" and "De-Identified Council Data"). In our scenario the data is de-identified and minimised following a proportionality assessment to determine the minimisation scope. The risk analyst selects an asset and opens the asset Controls panel. Figure 17 and Figure 18 show the controls for the Health Data and the Council Data respectively. The controls displayed are from the Privacy-enhanced SSM Risk Knowledge Base. The KB also includes controls modelled earlier during Threat and Control specification (See Section 4.2.2), whose use will be illustrated in later sections.

4.3.2. Calculating Initial Risk Level

Once the system model has been created the risk analyst assesses the initial risk level in the system. The risk assessment is run automatically by pressing the orange button in Figure 16. This action matches threat specifications from the SSM Risk Knowledge Base to assets and relations that appear within the System Model and

then calculates likelihood and risk of Consequences. The results are displayed in the Effects[6] panel, as shown in Figure 19. In the scenario, the SSM identified a total of 567 applicable threats and calculated the overall (worst case) system risk to be Very High.

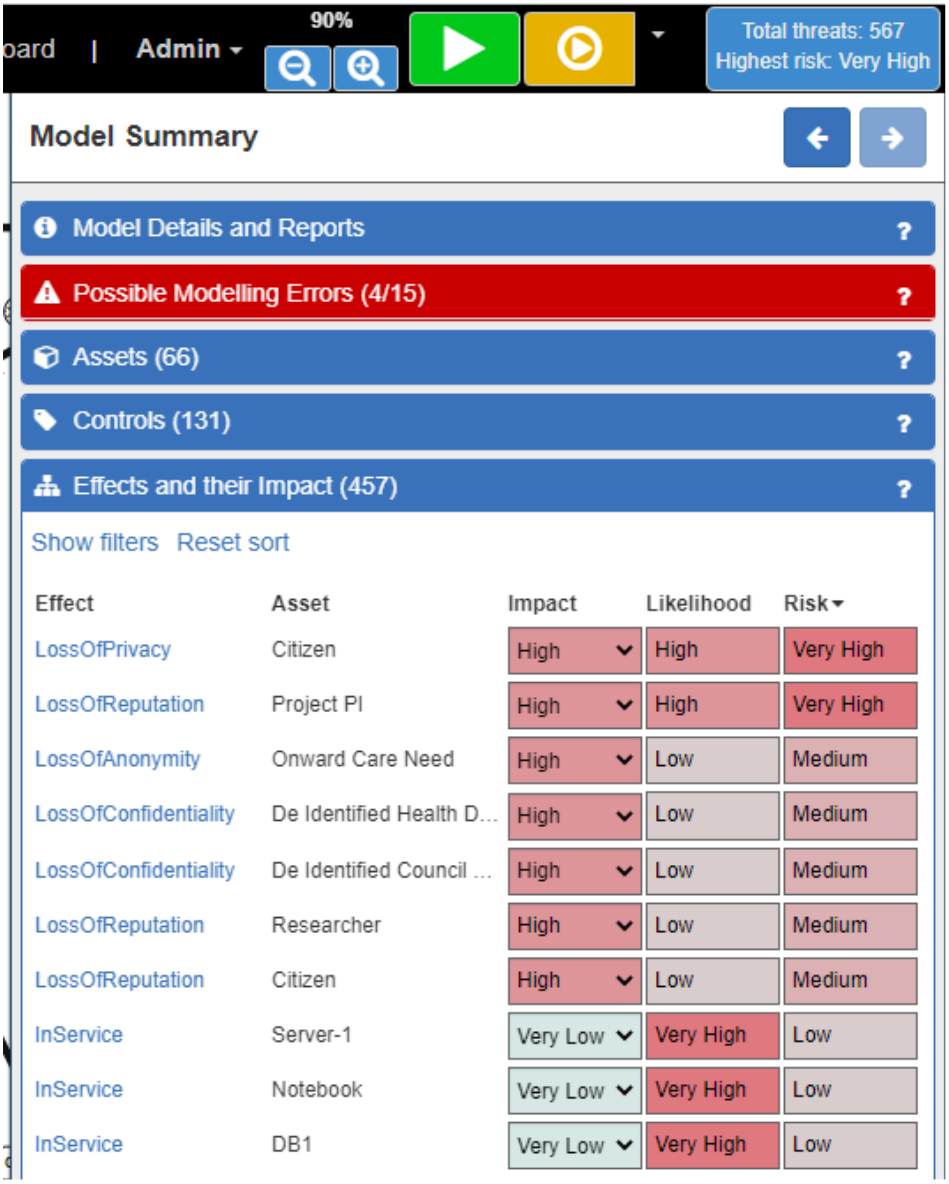


Figure 19: Complex Discharge Scenario - Starting Worst Case Risks

Analysing the situation further, we can see that the worst-case consequence in terms of risk level is a Loss of Privacy at the Patient (highlighted in green). There is also a Loss of Reputation of the Project PI - this is cascading effect of the Loss of Privacy at the Patient, i.e., the Project PI has ultimate responsibility for the project and if Data Subjects suffer harms associated with Losses of Privacy, then Loss of Reputation will be inevitably experienced by the Project PI. If the Loss of Privacy of the Patient consequence is addressed, the Loss of Reputation of the Project PI consequence will automatically be addressed.

[6] Consequences (ISO27000) are known as Effects in the SSM, further work on harmonising terminology is ongoing with the tool.

4.3.3. Threat Exploration and Control

We can now explore the Loss of Privacy consequence at the Patient by clicking on this consequence in the Effects panel. The Effects Explorer (Figure 20) is displayed which shows two threats that are direct causes of this consequence - "Viewing of Sensitive Data" and "Linking Data" and the calculated likelihoods. Both threats are High likelihood leading to a Very High risk level. The risk level is Very High as it is calculated from the threat likelihood at the asset (High) and impact of the consequence (i.e., Loss of Privacy consequence at the Data Subject has a High impact level, see Figure 19.

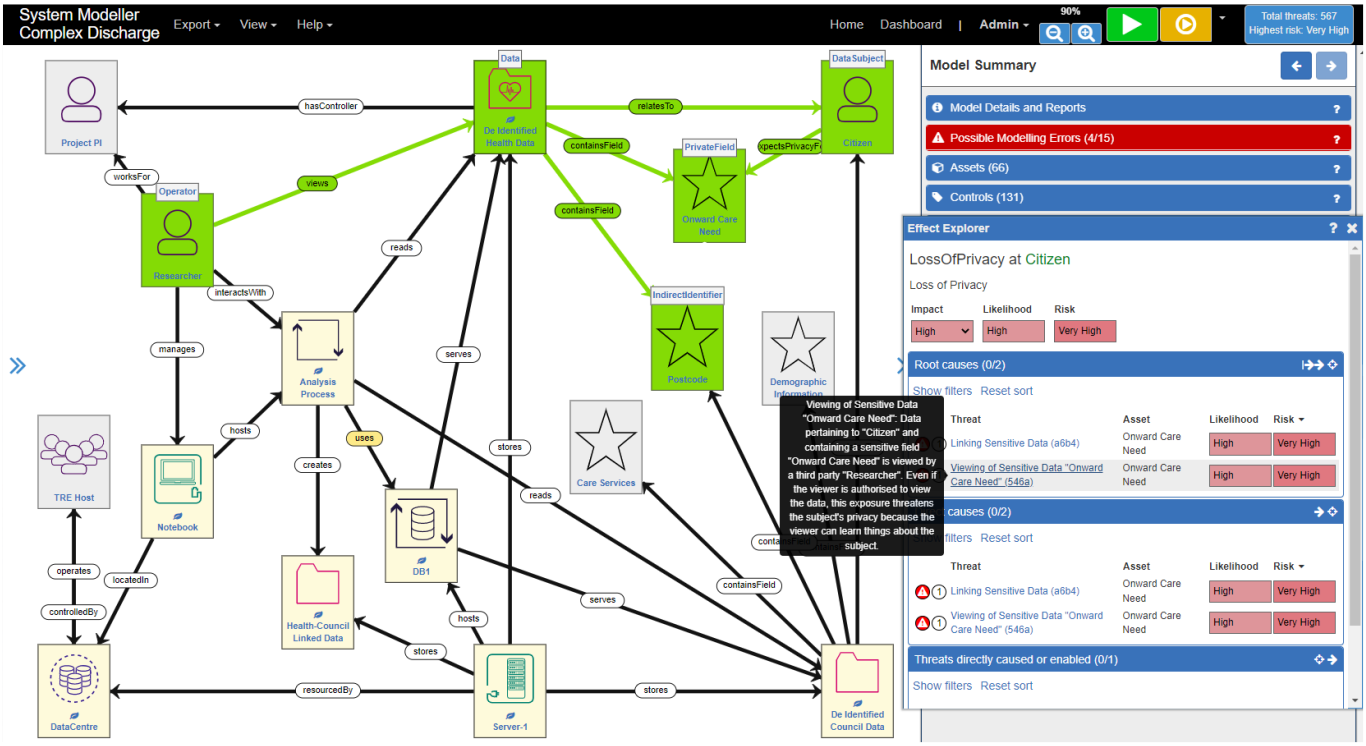


Figure 20: Exploration of Loss of Privacy Consequence at Patient

We can now explore each threat further to determine root cause and controls that can be applied to mitigate the threat. Firstly, as shown in Figure 20, by selecting the threat "Viewing of Sensitive Data", further threat description can be displayed using a tooltip and the assets in the system model that match the threat's matching pattern (as described in Figure 12 are visually highlighted (in green). This interactive exploration allows a risk analyst to learn about threats from the SSM knowledgebase, and how they occur in the system being modelled.

We can then explore each threat further using the Threat Explorer which is displayed by clicking a threat in the Effect Explorer. The Threat Explorer includes a causal analysis, showing direct and indirect effects of the threat; and also allows a risk analyst to select applicable control strategies to reduce the likelihood of the effects.

4.3.3.1. "Viewing of Sensitive Data" Threat & Controls

Figure 21 and Figure 22 show the Threat Explorer for "Viewing of Sensitive Data". Two screenshots are shown as the Threat Explorer includes more than can be displayed on a single screen.

The direct causes have been identified as "Anonymity at Postcode" and "DefaultTW at Researcher" (default trustworthiness of the researcher). Both Anonymity" and "DefaultTW" are asset vulnerabilities[7] that increase the

[7] Asset Vulnerability (ISO27000) is known as Trustworthiness Attribute in the SSM, further work is ongoing to harmonise terminology within the SSM tool.

likelihood of the threat occurring, as described in its Threat Specification. The Threat Explorer also shows the consequence of the threat - here Loss of Privacy on the Patient (Data Subject).

The Threat Explorer then recommendations control strategies to address the threat. Each Control Strategy is an option, and any one of these will address the threat by implementing blocking effects on the asset. The Control Strategies correspond to those discussed above in Section 4.2.2.1 and include controls that can be applied at the assets in the threat specification.

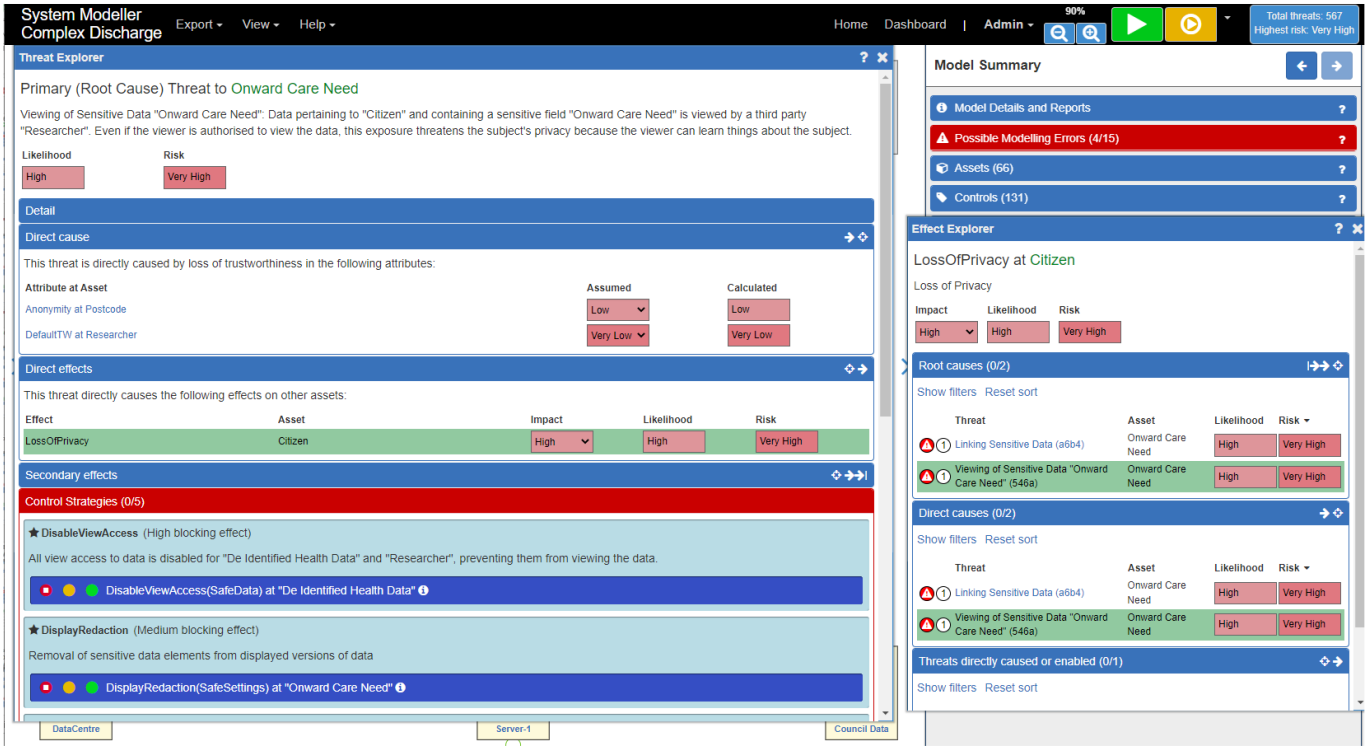


Figure 21: Threat Explorer for "Viewing of Sensitive Data" Threat (top)

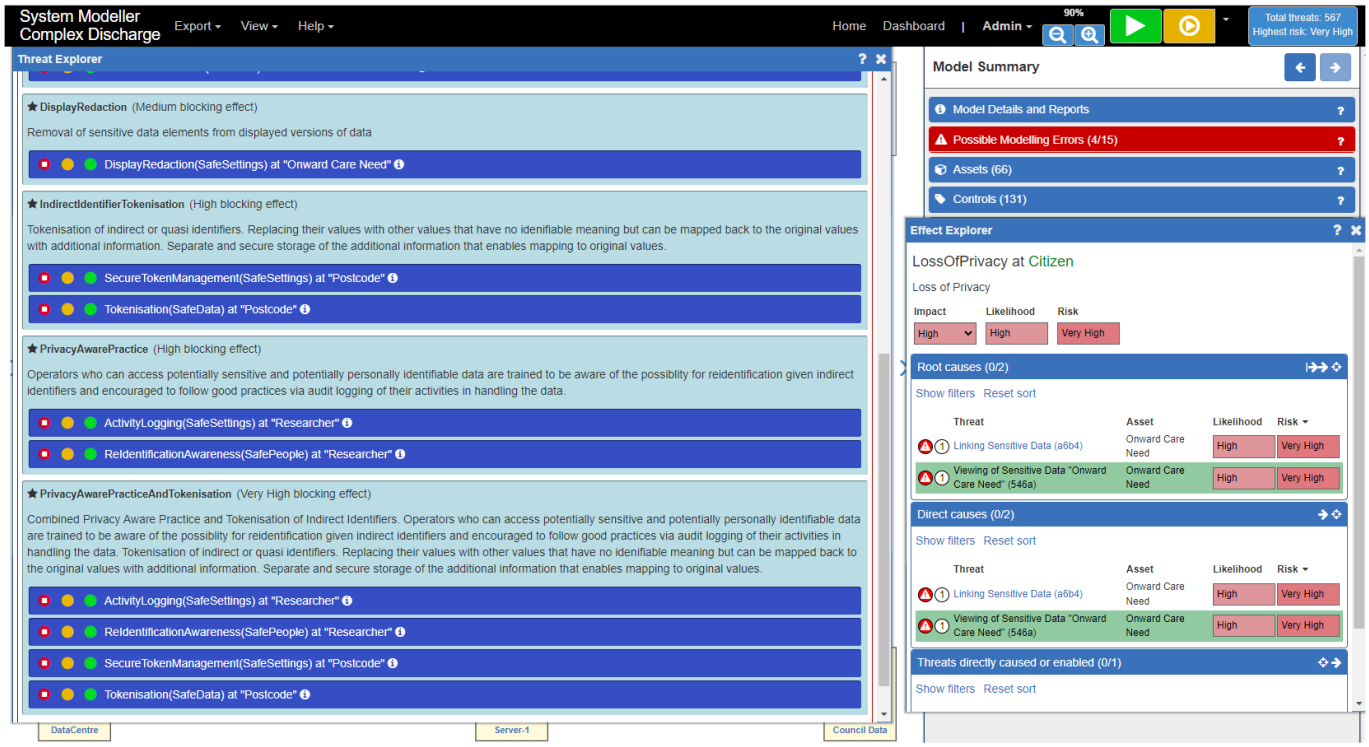


Figure 22: Threat Explorer for "Viewing of Sensitive Data" Threat (bottom)

To apply a control to an asset, we now click on the green radio button next to the control, as shown for "Indirect IdentifierTokenisation" and "PrivacyAwarePractice" in Figure 23. When all the controls are selected for a Control Strategy, the whole strategy turns green.

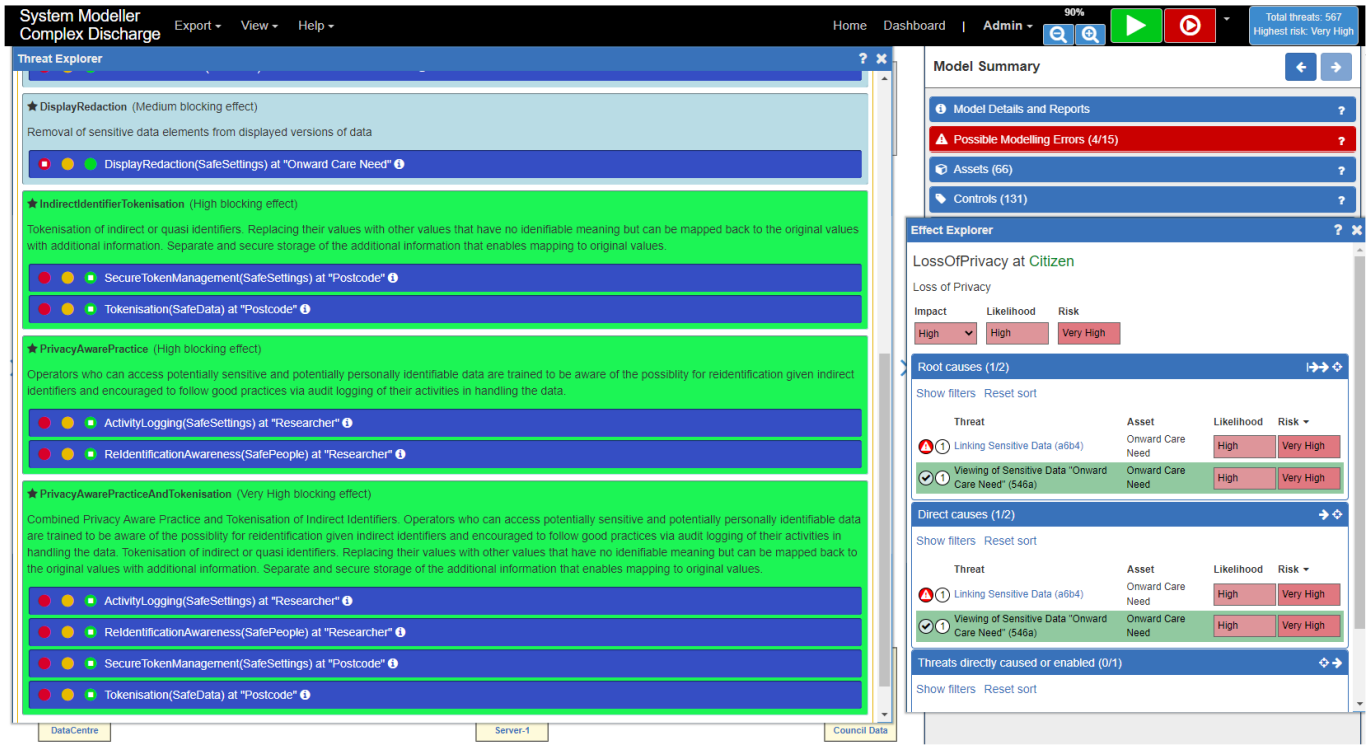


Figure 23: Threat Explorer for "Viewing of Sensitive Data" Threat (bottom, with controls selected)

The combined Control Strategy of "Privacy Aware Practice and Tokenisation" (bottom) has all its Controls selected, and because this strategy is a combination of the two simpler strategies these strategies are also green. Figure 23

also shows that the “Viewing of Sensitive Data” threat is addressed by showing a tick next to its name in the Threat Explorer (centre right).

Applying controls invalidates the risk level and the risk calculation must be rerun to see any changes in risk level. The invalid risk level is indicated by the red button at the top right of Figure 23. We then rerun the risk calculation with the result is shown in Figure 24. We can see in the Effect Explorer that the “Viewing of Sensitive Data” threat for “Onward Care Needs” data asset has a likelihood reduced from High (as previously shown in Figure 20) to Very Low (Figure 24). The reduction in likelihood is the result of the “Very High” blocking effect of the Privacy Aware Practice and Tokenisation control strategies.

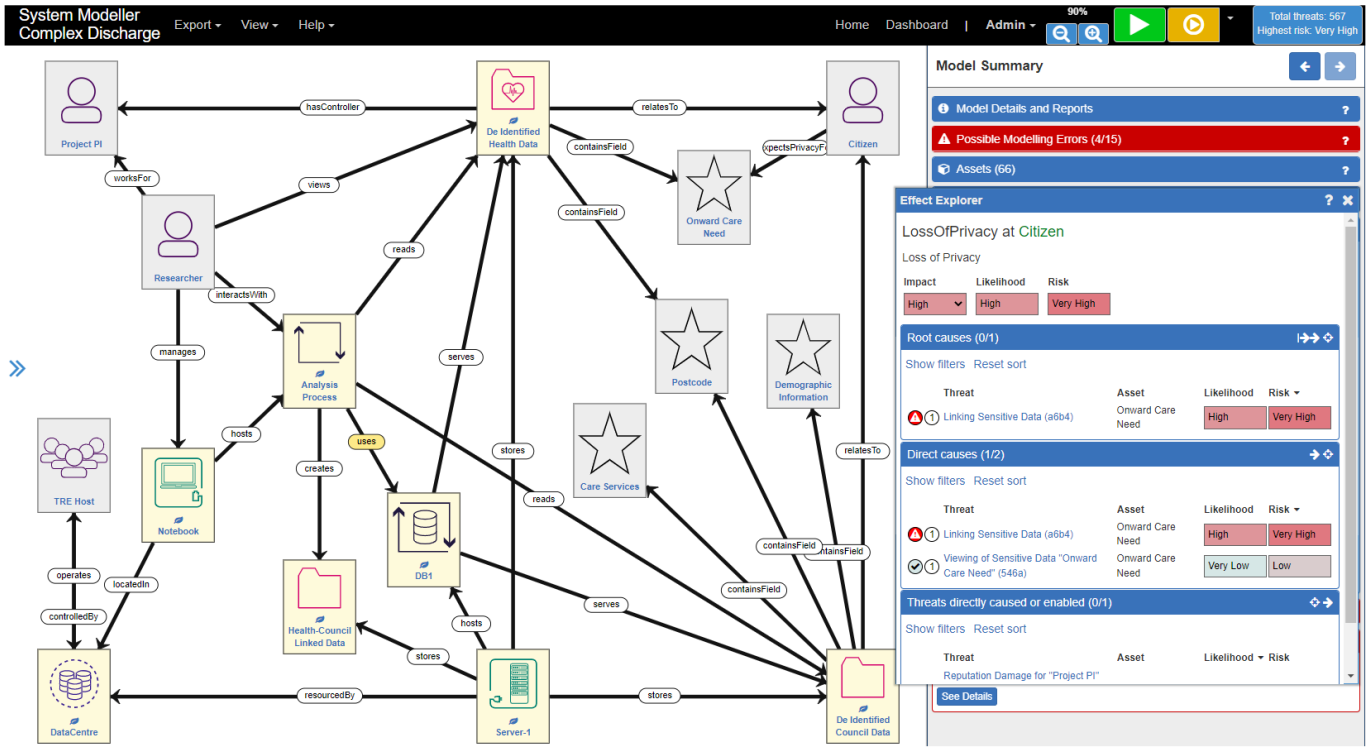


Figure 24: Result of Addressing "Viewing of Sensitive Data" Threat

4.3.3.2. “Linking Sensitive Data” Threat & Controls

Even though the likelihood of “Viewing of Sensitive Data” threat is reduced, the overall risk level for the system is still Very High (see the worst-case risk in the top right corner of Figure 24). This is because the “Linking Sensitive Data” threat is also causing the Loss of Privacy at the Patient with a High likelihood. Selecting the threat highlights the relevant assets and relations, as well as the threat description in a tooltip (Figure 25).

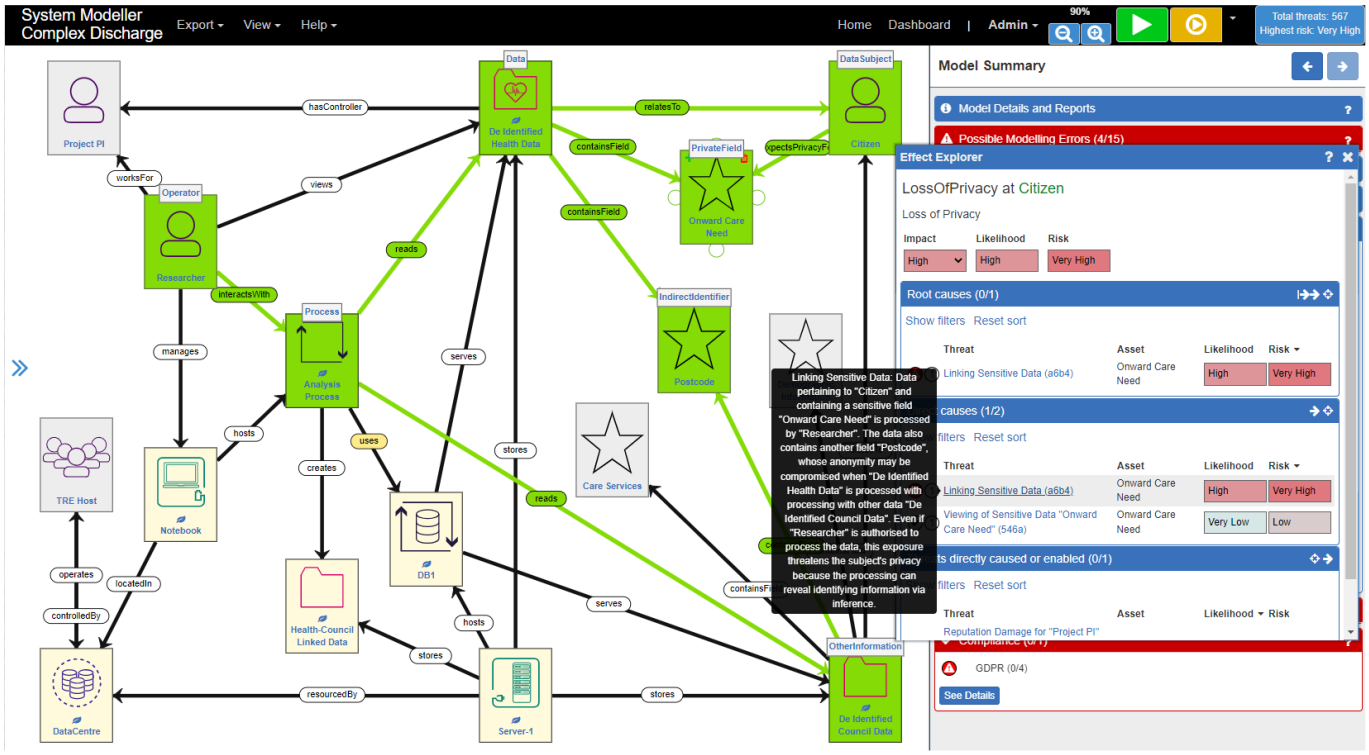


Figure 25: Assets Affected by Linking of Sensitive Data Threat

As previously, to find the Control Strategies for this threat, the Threat Explorer needs to be opened as shown in Figure 26 and Figure 27.

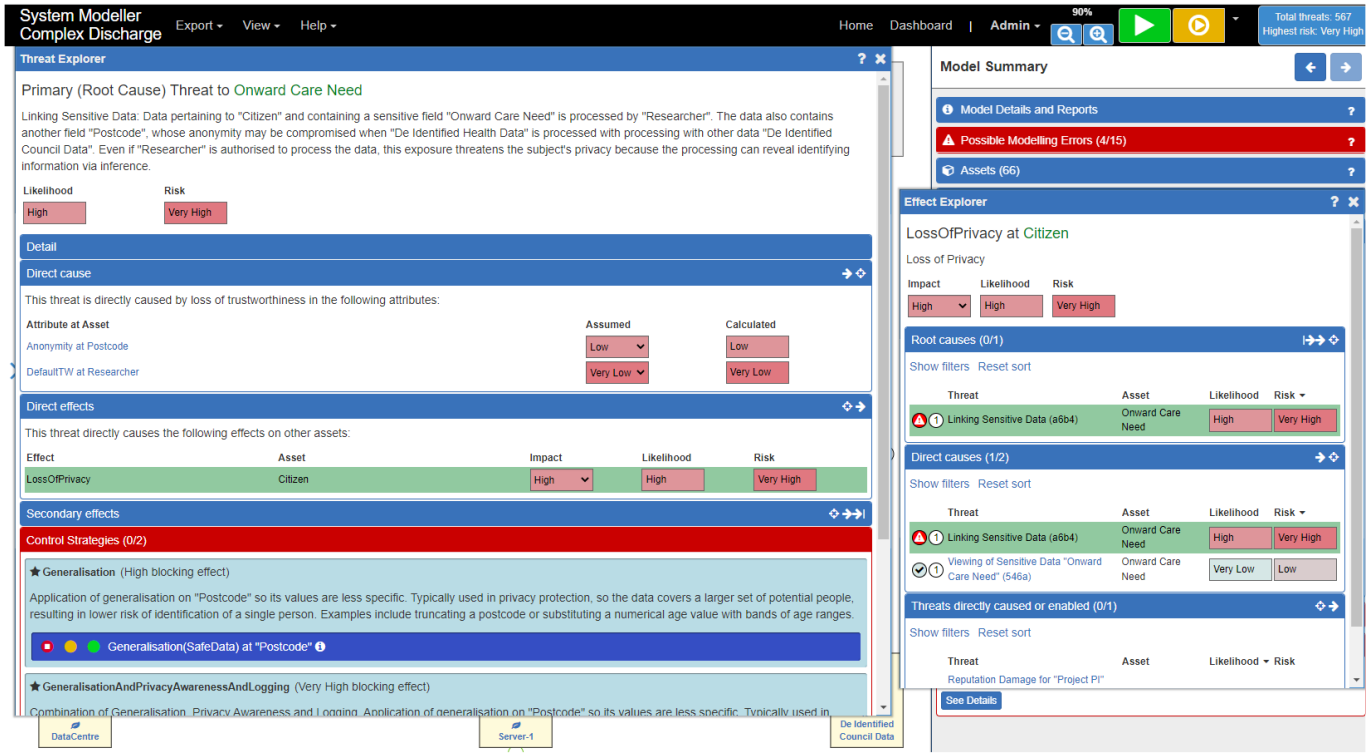


Figure 26: Explorer for "Linking Sensitive Data" Threat (top)

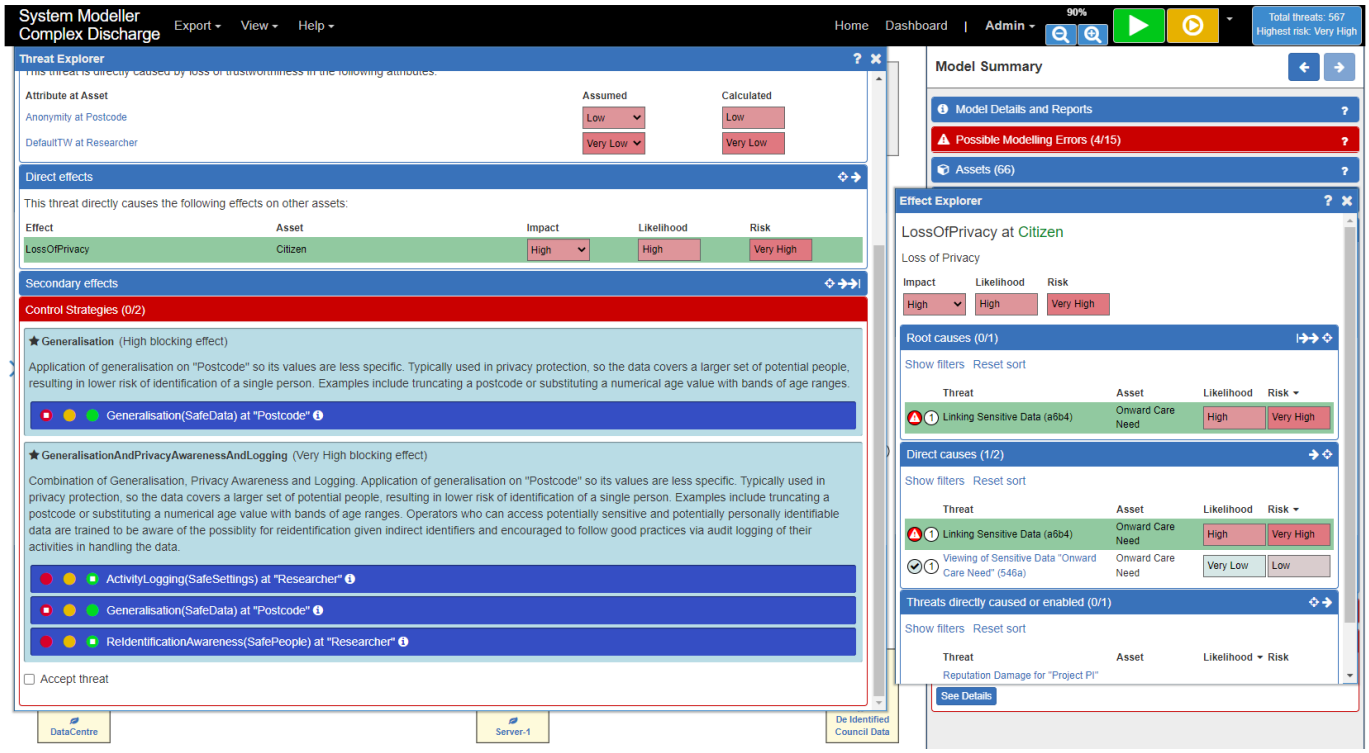


Figure 27: Explorer for "Linking Sensitive Data" Threat (bottom)

A point to note in Figure 27 is that some of the controls in the combined control strategy of “Generalisation and Privacy Awareness and Logging” are already selected. This is because it shares controls with the control strategies needed to address the threat of “Viewing Sensitive Data”, which have already been applied as described above. The only remaining control needed to activate this control strategy is Generalisation, and activating this control results in Figure 28.

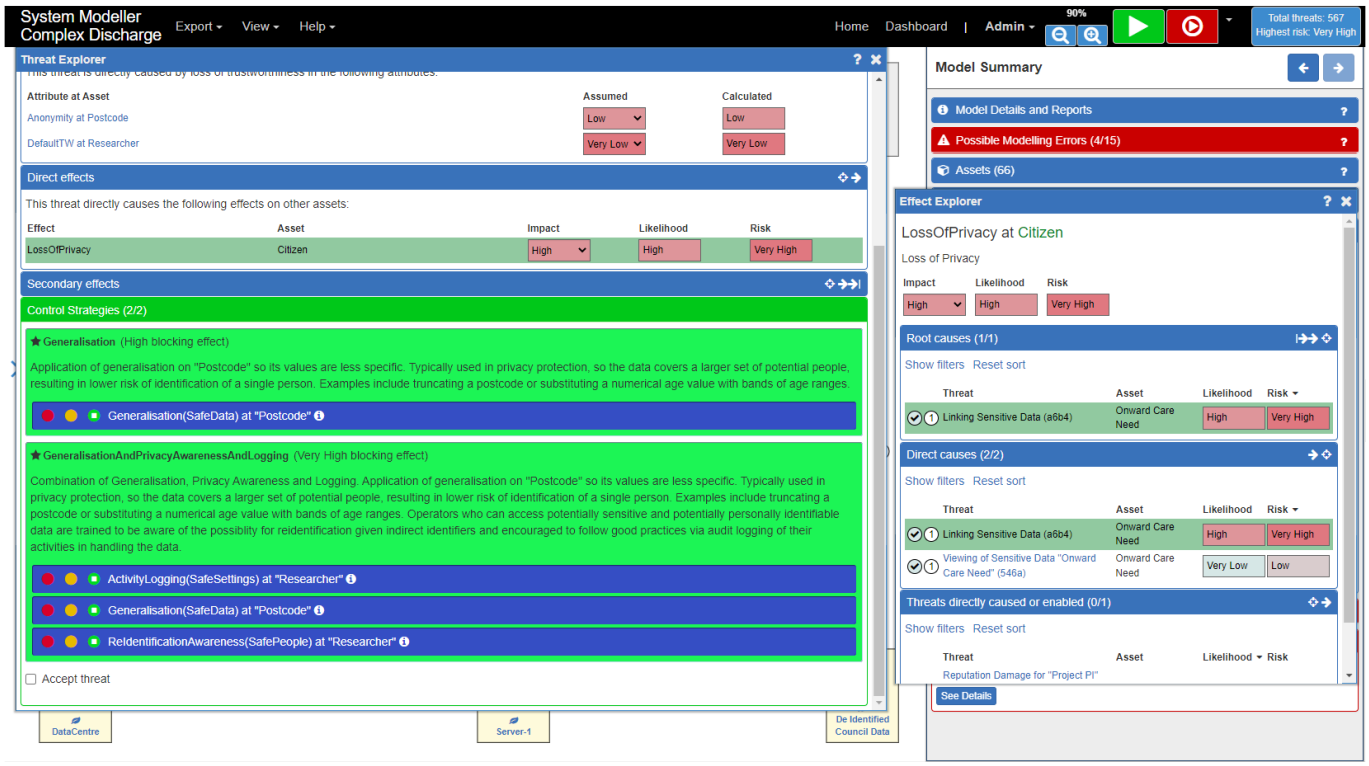


Figure 28: Activating Controls to Address "Linking Sensitive Data" Threat

Activating the Generalisation control satisfies the “Generalisation” and “Generalisation and Privacy Awareness and Logging” control strategies as shown in green. This addresses the “Linking of Sensitive Data” threat, as shown in the Threat Explorer at centre right. The risk recalculation can now be rerun to determine the updated threat likelihood given the new controls applied (Figure 29).

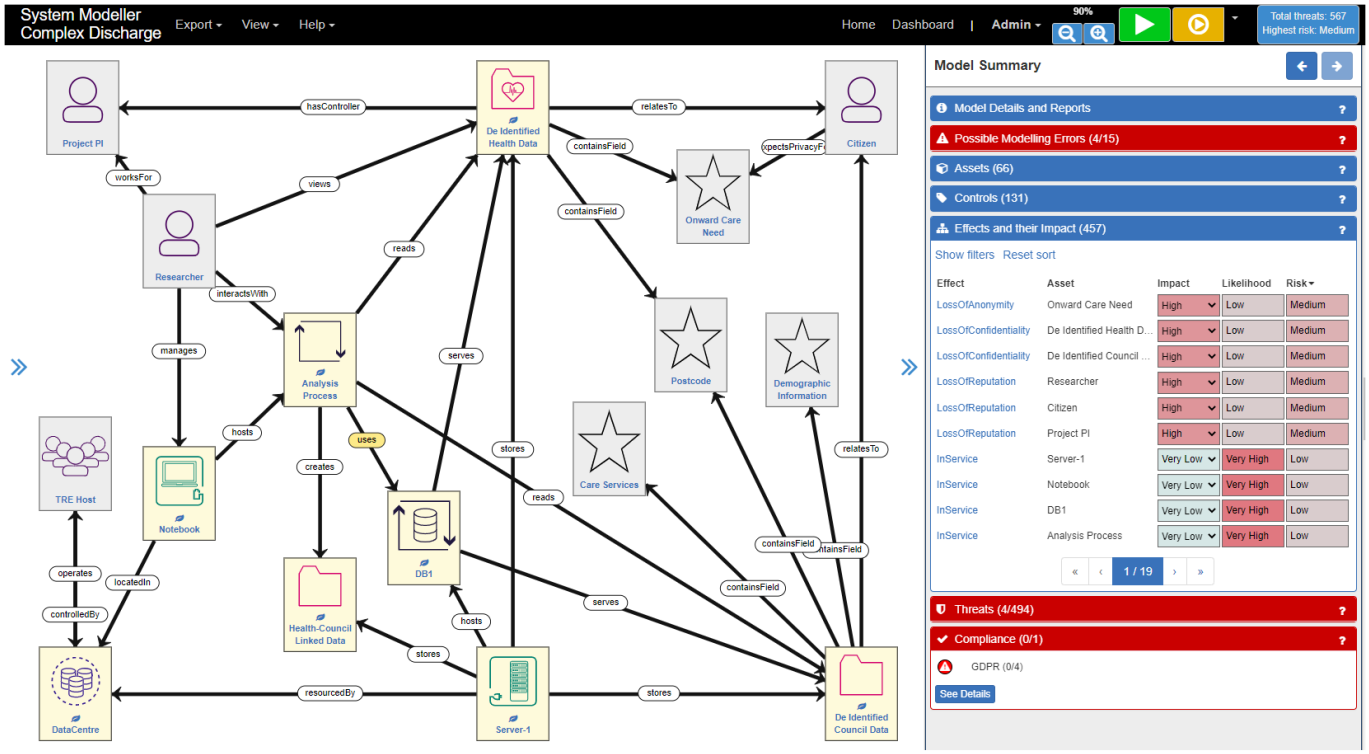


Figure 29: Highest Risks After addressing Viewing Sensitive Data and Linking Sensitive Data Threats

Following recalculation, the worst case risk (Loss of Anonymity on Onward Care Need) has a risk level of “Medium”, which is the overall risk level for the system. Loss of Privacy cannot be seen on this page because the risk level has been reduced, and the risks are displayed by default worst first. We can search for “Loss of Privacy” in the Effect Explorer (Figure 30) see that the likelihood of “Loss of Privacy on the Patient data asset has reduced to Very Low resulting in a Low level of risk.

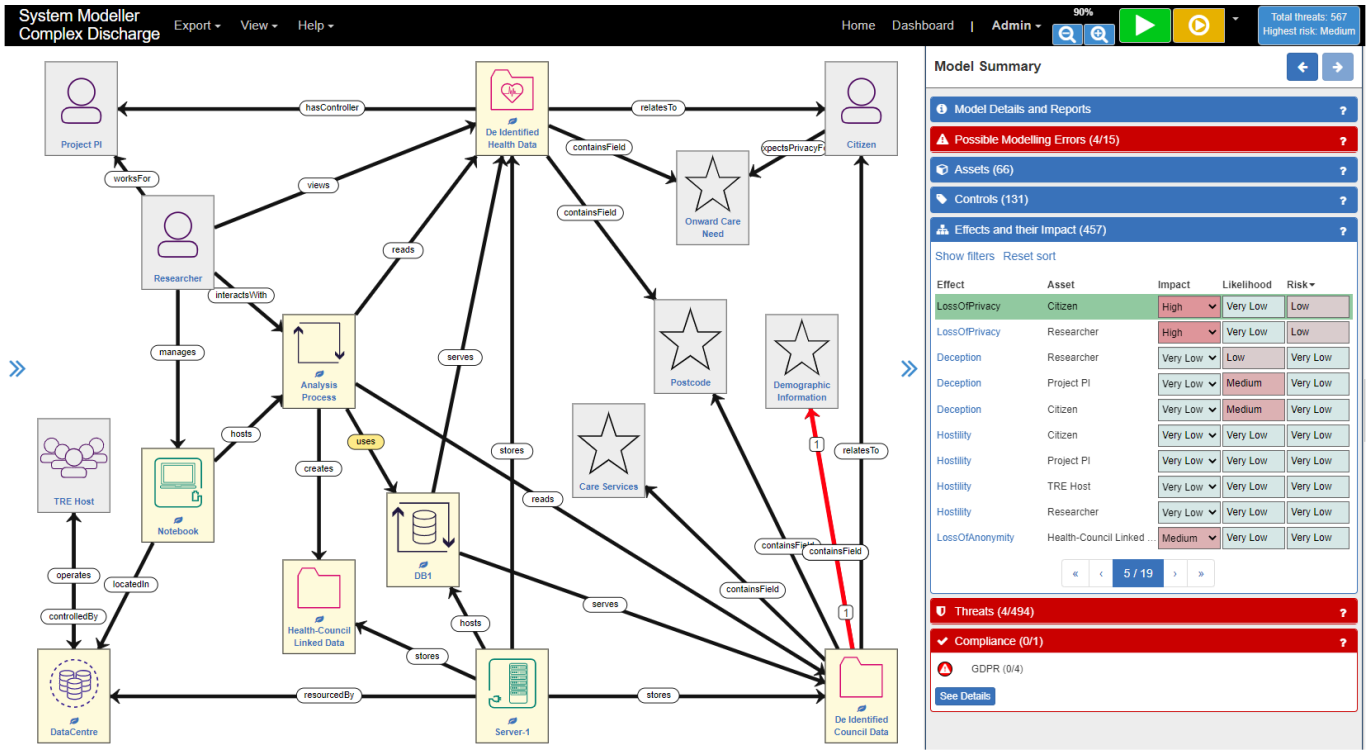


Figure 30: Loss of Privacy Risk after Controls Applied

This section has described the modelling of privacy knowledge, enhancement of the SSM Risk Knowledge Base with privacy risk knowledge and shown the KB in action with an example. The next section draws brief conclusions.

5. Discussion & Conclusion

This report has described a privacy risk modelling approach based on ISO27005 and shown how privacy risk factors can be encoded into the elements needed for such a risk modelling. This includes Assets, their Vulnerabilities, Threats to those Assets, the Consequences that arise from Threat attacks on Assets, and Controls that can lower the likelihood of Threat attacks. New elements of each of these types have been determined specific to privacy protection, specifically addressing risks that lead to privacy harms and these have been modelled as exemplars.

The modelling process has been illustrated by example following a use case scenario for a research project studying complex discharge of a patient from hospital. The research project requires the linking of data from a hospital and a council, who are responsible for delivering onward social care services. The data contains sensitive information for which there is an expectation of privacy on the part of the patient, so there is a risk of privacy violations even though the scenario has assumed that the data from both the hospital and council is de-identified and minimised as a starting condition. This illustrates that federated data analysis and data linking can cause additional privacy risks that must be mitigated.

We have demonstrated how to model privacy risks emerging from Risk Tier analysis of the Five Safes within the knowledgebase of the System Security Modeller and then how the tool supports interaction and automated risk assessment in accordance with ISO27005. We focused on how to apply control strategies to reduce a Loss of Privacy to a Low level.

Some key lessons have been learned from this work and are summarised below:

- The SSM shows the systemic risks given a system configuration and choice of controls, so it can help people to determine an effective risk treatment plan. The SSM had its origins in cybersecurity, but DARE UK PRiAM has shown how different (but related) threats and risks can be encoded into the same knowledge base, providing useful complement to, and integrated with, the cybersecurity knowledge already encoded. This means that existing cybersecurity controls such as access control, encryption and malware detection already encoded in the knowledge base can be utilised to address privacy threats as well as cybersecurity threats, thus increasing their utility.
- Risk Factors, Threats, Consequences and Controls can be classified in terms of the Five Safes. A control strategy for the same threat can combine controls from different aspects of the Five Safes and it is their combination that determines the effectiveness of the strategy in addressing threats.

The modelling approach described here is a first step towards the open curation of knowledge for privacy risks. Given the time constraints of the DARE UK PRiAM project, we have demonstrated a proof of principle, methodology and tooling for open curation of security and privacy knowledge and risk assessment. In future we would expect open communities of domain experts and the public to contribute to identification, curation and reuse of an open knowledgebase for modelling and assessing risks in systems.

The Threat Specifications of this report are encoded into the SSM's knowledge base for the purposes of demonstration and evaluation of the mapping between risk factors and ISO27005 concepts, but the concept of Threat Specification is useful in its own right because it ties together the real-world elements (e.g. data subjects, data, processing, operators, etc), their relationships with the cause (vulnerabilities and threats) and effect (consequences) of privacy risks and the controls to address threats.

Further work will be required to provide decision support to help users identify risk factors and vulnerabilities that lead to privacy violating consequences. For example, indirect-identifiers need experience to be identified in data, and exemplary lists of indirect-identifiers (e.g., the Postcode and demographic information) will be useful guidance and a pointer towards classification of data fields by type to help users identify the presence of relevant factors.

Sources of knowledge about privacy risk factors will continue to be identified from literature, expert communities and public consultation, whilst the DARE UK community and wider stakeholders continue to be a further valuable source of privacy risk factors.

A key recommendation from DARE UK PRiAM is to continue to build community expertise in analysis of risk factors and curation of privacy knowledge in both human and machine-readable formats to increase awareness of practitioners and to allow for development of transparent, repeatable and automated privacy risk assessment processes.

6. References

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.

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.

Commission nationale de l'informatique et des libertés (CNIL). (2018a, February). Privacy Impact Assessment (PIA): Methodology. February 2018 edition. Retrieved from: | 55 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.

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.

Dictionary.com. (n.d.). Definition: trustworthy. Retrieved from: https://www.dictionary.com/browse/trustworthy.

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/.

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.

Griffor, E., Greer, C., Wollman, D., & Burns, M. (2017). Framework for Cyber-Physical Systems: Volume 1, Overview, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD. https://doi.org/10.6028/NIST.SP.1500-201.

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

ISO/IEC 27001:2013. Information technology — Security techniques — Information security management systems — Requirements. https://www.iso.org/standard/54534.html

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

Kurdi, D. (2021, February 2). Postcode Facts: Interesting Facts and Information about Postcodes. IdealPostcodes. Retrieved from: https://ideal-postcodes.co.uk/guides/postcode-facts.

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.

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 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/2020/01/1 6/NIST%20Privacy%20Framework_V1.0.pdf.

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.

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.

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.

Surridge, M., Correndo, G., Meacham, K., Papay, J., Phillips, S., Wiegand, S., & Wilkinson, T. (2018), "Trust Modelling in 5G mobile networks," in SecSoN '18: Proceedings of the 2018 Workshop on Security in Softwarized Networks: Prospects and Challenges, 2018. https://doi.org/10.1145/3229616.3229621.

Taylor, S., Surridge, M., & Pickering, B. (2021), "Regulatory Compliance Modelling Using Risk Management Techniques," *2021 IEEE World AI IoT Congress (AIIoT)*, 2021, pp. 0474-0481, doi: 10.1109/AIIoT52608.2021.9454188. y.

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/.

7. Glossary

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

Key term	Definition
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 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.
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).
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 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.
Re-identification	"De-identified data, or data otherwise disassociated from specific individuals, becomes identifiable or associated with specific individuals again. It can lead to problems such as discrimination, loss of trust, or dignity losses" — as defined by NIST (2019).
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).
System Security Modeller (SSM)	An asset-based cybersecurity risk modelling tool created by the University of Southampton, designed to follow the ISO/IEC 27005 methodology for information security risk management, focusing on threats arising in socio-technical systems related to cyber-security and non-compliance (e.g., with the GDPR).
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').