University of Dundee

SATRE: Standardised Architecture for Trusted Research Environments

Cole, Christian; Sood, Hari; Li, Simon; Oldfield, Katie; Craddock, Matt; Swanepoel, Nel

A Standard Architecture for Trusted Research Environments

Tim Machin, Ed Chalstrey, Christian Cole, Matthew Craddock, James Hetherington, Simon Li, Jim Madge, Martin O'Reilly, James Robinson, Nel Swanepoel

03/10/2023

Contents

1. Approach

In order to provide the greatest utility a multi-level architectural approach has been used to create a standard architecture which Trusted Research Environment (TRE) implementers can use. It is hoped this standardised architecture can improve decision-making, enhance agility and flexibility, optimise costs, aid interoperability, improve compliance and improve communication between implementers.

The broader SATRE specification which contains more detail about the components and controls can be found here. The architecture and associated models have been created in collaboration with the wider SATRE team through community engagement across the TRE community in the UK.

https://satre-specification.readthedocs.io/en/latest/specification.html

Models within this document have been created using the open-source tool ArchiMate. You can find a description of the elements and meta-models in appendix 1.

In architecture as with statistics "all models are wrong, but some are useful" (George E. P. Box). These models should not be considered authoritative or prescriptive. For the architectural models to provide utility we should understand the outputs in two key dimensions: audience and purpose.

1.1. Dimension 1 – Audience

This document and associated models are intended for use by designers and builders of trusted research environments including research infrastructure developers and enterprise and solution architects.

1.2. Dimension 2 – Purpose

This architecture will serve as a tool to:

- Help implementers design and implement environments within a variety of organisations using a mix of existing and novel technologies.
- Provide insight into the underlying components of a TRE and allow improvement planning and road mapping.
- Provide a common understanding of the component elements of a TRE, what functionality they provide and how they fit together to the TRE user community.
- Facilitate the communication of how one or many TREs function alone or together through federation.

1.3. Architecture Views

Within this document there are multiple views documented including capability map view, capability decompositions and 5 safes process view. More views can be created to highlight or demonstrate elements and relationships.

1. Overview

The SATRE architecture will be universally applicable to all TRE implementations and is documented in three parts, principles, capabilities and capability decompositions.
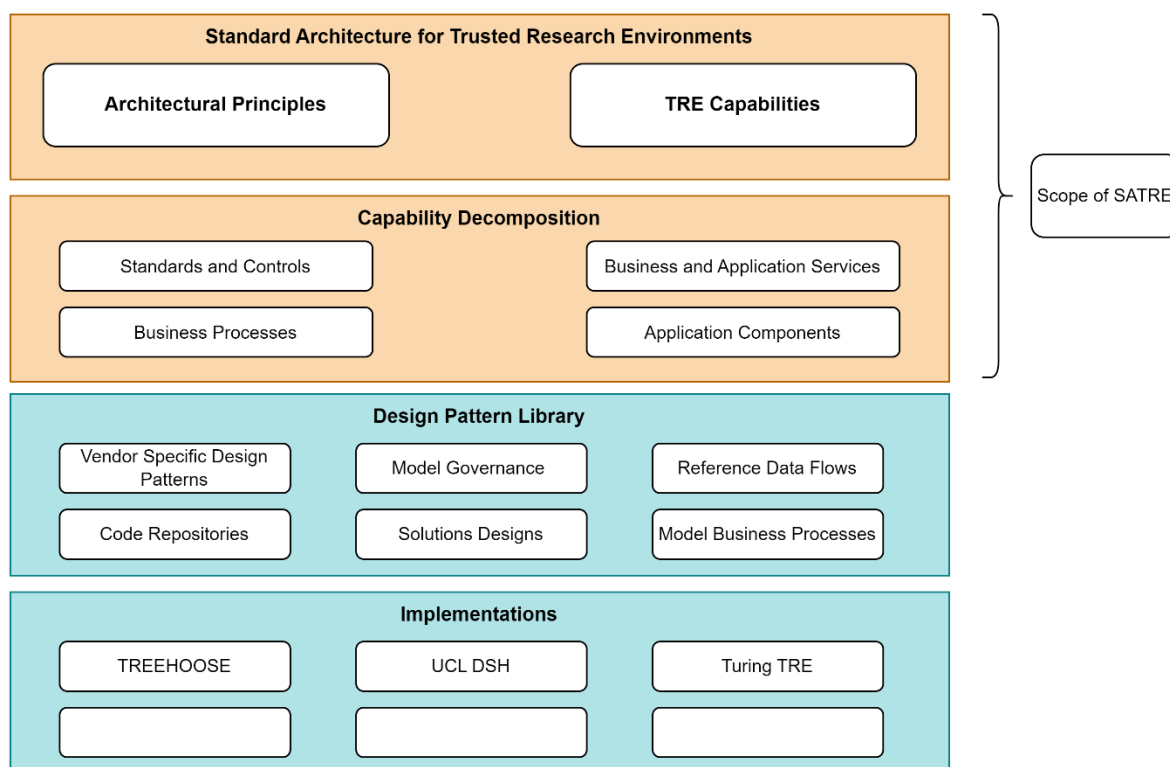
1. **Architecture principles** provide fundamental guidelines that inform the design, decision making and implementation for a TRE. These principles provide a framework to ensure that the design of the underlying components of a TRE are aligned to consistent goals, values and best practices. These principles can be found in the main SATRE Standard (https://satre-specification.readthedocs.io/en/latest/principles.html).

2. **Capabilities** are an ability that an organisation possesses. Capabilities are typically expressed in general and high-level terms and typically require a combination of organ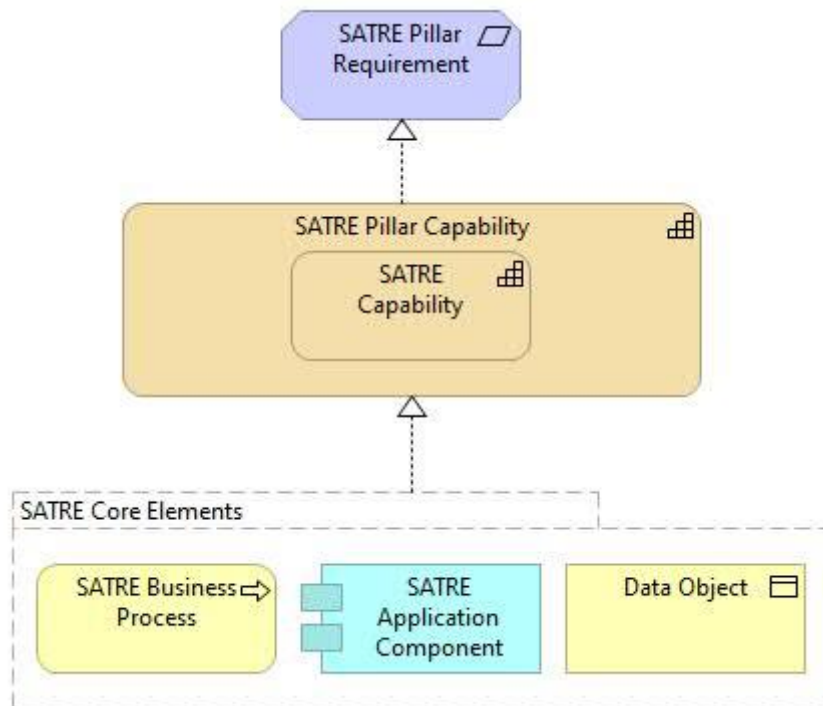isation, people, processes, applications and technology to achieve. The reference map provides a structure for organising and understanding the capabilities required to deliver a TRE. It helps identify, define, and categorise the key elements needed to create and operate a TRE.

3. **Capability decomposition** will map out the components that realise a capability. These components will vary depending on the nature of the capability. Business focused capabilities will be primarily realised by business processes, roles and services with more technology focused capabilities realised by applications, interfaces and technology. In addition to the components realising the capability a catalogue of standards, frameworks and controls linked to the capabilities will provide guidance on how to implement the capabilities safely. These can be found in the SATRE standard (https://satre-specification.readthedocs.io/en/latest/).

1.1. Meta-model

Below is the meta-model to which the architecture conforms.

The decomposition elements of people (actors and roles), processes, data and applications combine to provide an organisation with the ability to fulfil or realise a requirement. Business processes are triggered by events such as changes or requests, applications serve those business processes through digitisation and automation with data being read and written programmatically via the application or potentially manually by the operators of the process.

1.2. Design Pattern Library and Implementations

Built on top of this standard architecture will be a variety of implementations. These will have varied process flows, application sets, code and underlying platforms and technology. For example, TREs built in Azure and AWS would conform to the same standard architecture with alternative code and technology component choices.

To make the reference architecture as useful as possible for implementers a design pattern library will provide a set of configurable resources implementers can customise to implement a TRE in line with the SATRE architecture. These designs will provide recognised solutions which conform to the SATRE architecture and standard. Such patterns can be provided as design documents, models or shared as code. Implementers can then utilise these design patterns in their organisation's TRE implementation.

The design pattern library is beyond the scope of this document and could form the basis for future work.

2. Capability Map

The capability map represents all the capabilities an implementing organisation needs to run a TRE safely.
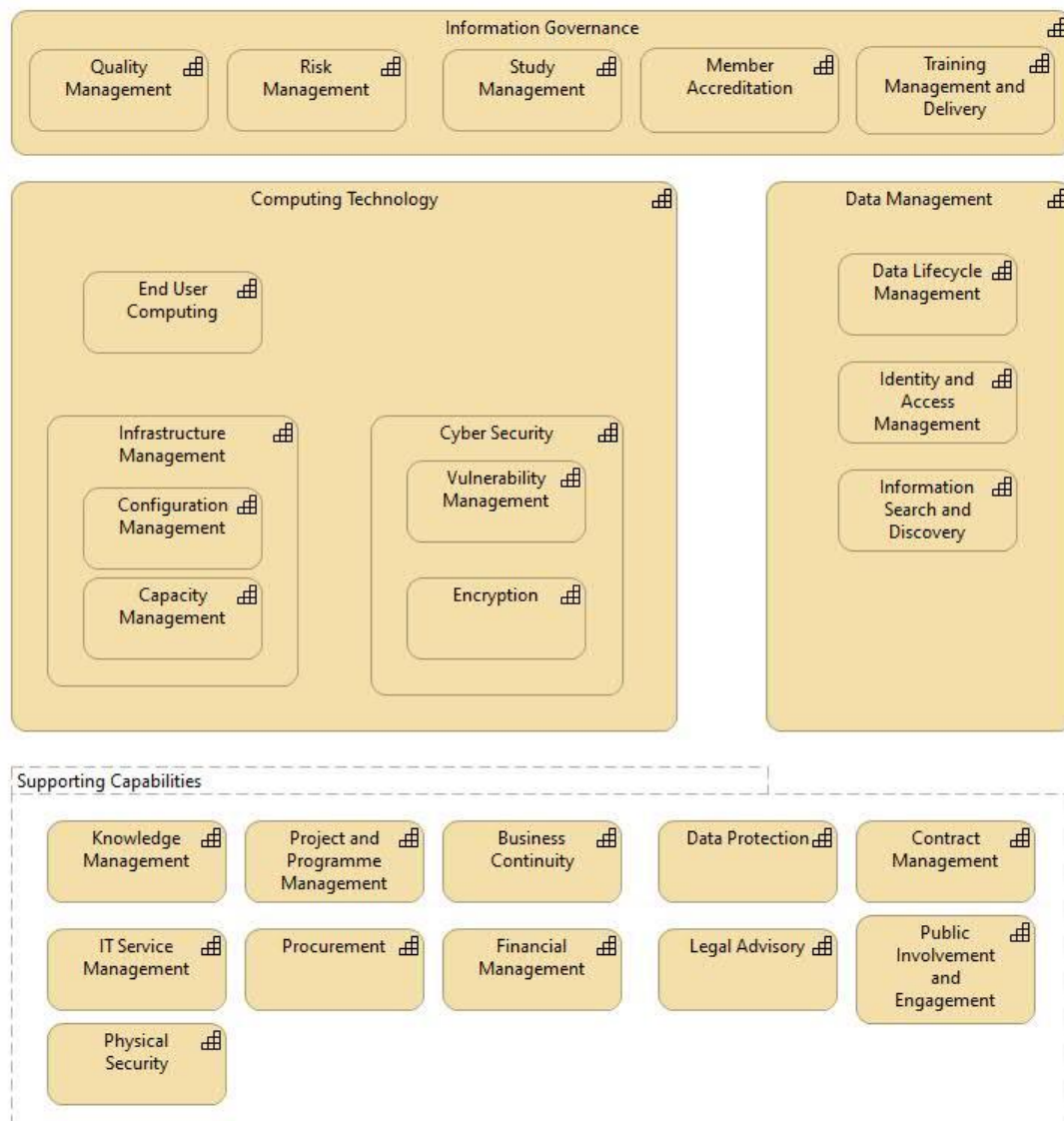
There are 4 core pillar capabilities required to run a TRE safely:

- **Information governance** - What the organisation does to ensure information risk is measured and managed to an acceptable level.
- **Computing technology** - What the organisation does to manage systems for storing, processing, retrieving, and sending information.
- **Data Management** - What the TRE organisation does to manage data assets and ensure information remains secure.
- **Supporting Capabilities** – These organisational abilities are likely to be shared across the wider enterprise and unlikely to relate solely to a TRE. For example, procurement is likely to be part of a finance function and while the TRE organisation may stress the need to secure supply chains as part of procurement it is unlikely there would be a need to create parallel procurement processes for the TRE.

There will considerable variation in how these supporting capabilities work to support the TRE. For example, in a large organisation legal advisory may include access to in-house council, whereas, in smaller organisations this may be contracted advice from external legal firms. The TRE organisation is likely to leverage and augmenting these processes, but they are unlikely to have overall control or responsibility. Due to these variations the supporting capabilities are not modelled as decompositions in this document, however, they are covered in the SATRE standard linked above.

3. Capability Decompositions

3.1. Information Governance

> **What the organisation does to ensure information risk is measured and managed to an acceptable level.**
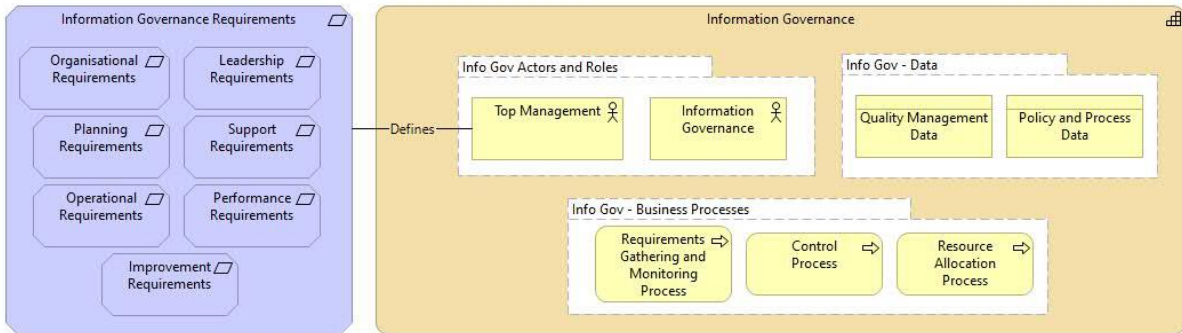
Information governance requirements are defined by top management within the TRE organisation. These will draw from the context of the organisation, the work it performs and the nature of the data it processes. Gathering and monitoring these requirements is a key process for ensuring the TRE aligns with the requirements. These requirements processes will trigger the control process which triggers actions taken to control risk. Control implementation requires prioritisation and resourcing, resources are allocated through the resource allocation process. The TRE top management must have control of adequate resources to control risk and suitable authority to act. Quality data is a key input to measuring and improving the effectiveness of the TRE in relation to governance, risk and compliance.

3.2. Quality Management

> **What the organisation does to measure and control quality of processes, documentation and outputs.**
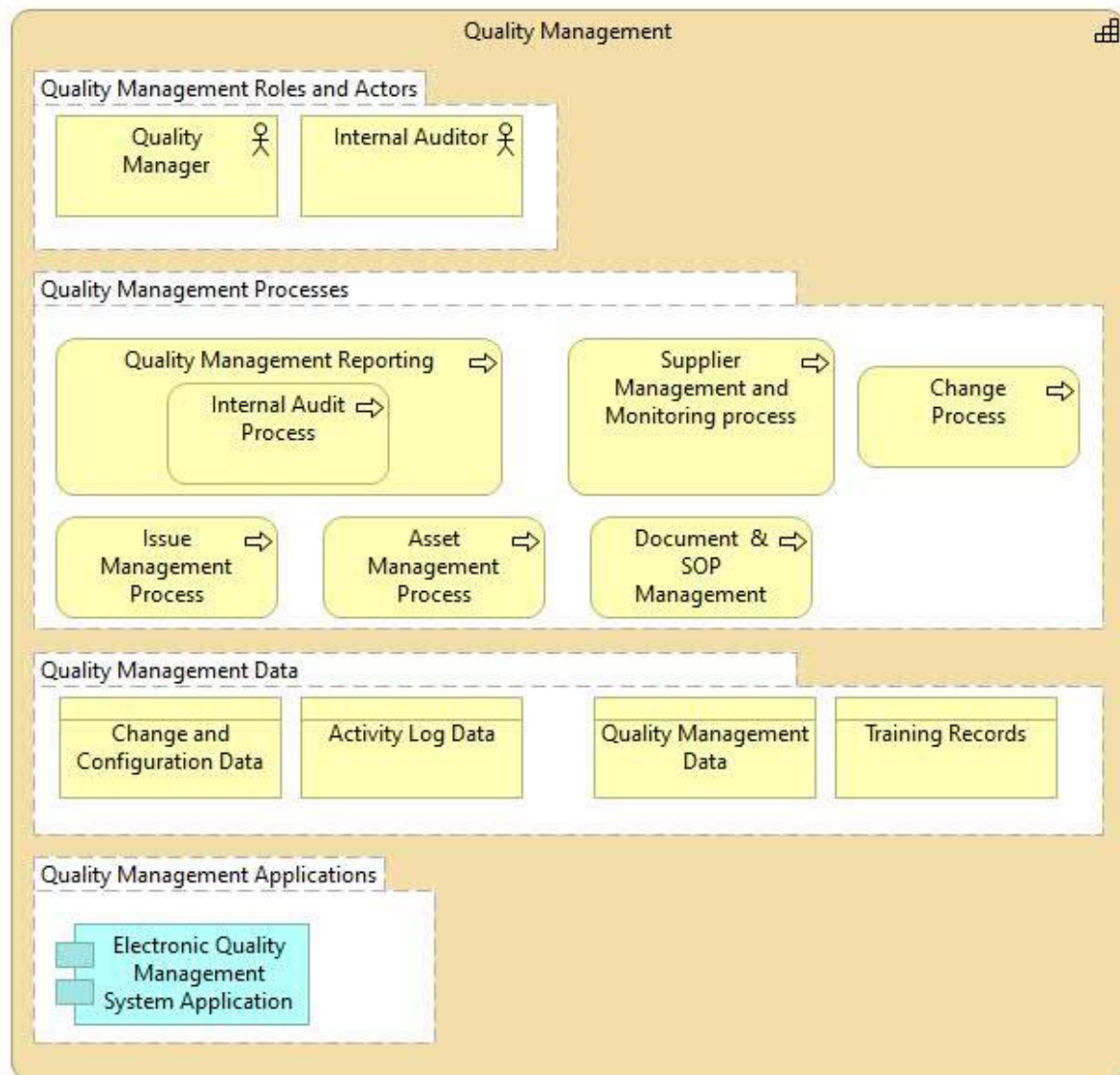
Quality management reporting would is continual but may be triggered by a scheduled event (such as governance meeting). Internal audits will be triggered according to a policy or an issue or change. Supplier management and monitoring will be continual but may also be triggered by a change or addition of a supplier. The issue process will be triggered by an adverse event, incident or audit outcome. Document and Standard Operating Procedure (SOP) and Asset management are continually running.

Quality management data is crucial to understanding whether the organisation is operating as intended. The quality manager role is responsible for keeping the TRE operating within risk tolerance with an internal auditor assessing the effectiveness of the TRE through the internal audit processes.

Measurement is a key element to quality management. All processes across the TRE will generate data and those data must be trusted, retained and be made available to the quality manager and internal auditor during an audit.

Documentation is a key control across the TRE for example, policies, SOPs, work instructions and contracts. Control of these documents is critical to the operation of the TRE and maintaining quality.

Quality management processes and data can be extremely complex and the ability to collate, measure, report and track data across many processes can be daunting. The implementation of an electronic quality management system that tracks all the processes within the capability and draws data from multiple sources for alerting and reporting purposes can reduce the overhead of performing quality management.

3.3. Risk Management

> **What the organisation does to ensure information risk is measured and managed to an acceptable level.**

Risk within the TRE organisation is ultimately owned by the top management with risk ownership for data assets delegated to information asset owners. Risks related to operations are assessed and treated by operational teams.

Risk ownership involves the understanding of risk appetite within the organisation and agreement to proceed with operations given the current or future risk landscape. The risk assessment process scores risk according to an agreed matrix usually based on likelihood and impact. Once a risk is assessed the risk treatment process is triggered that applies technical, policy and process controls to bring them within accepted tolerances. Where risks fall outside of the automatically accepted tolerances, they will be escalated to risk owners.

3.4. Study Management

> **What the organisation does to create and maintain research projects and work packages within the TRE.**
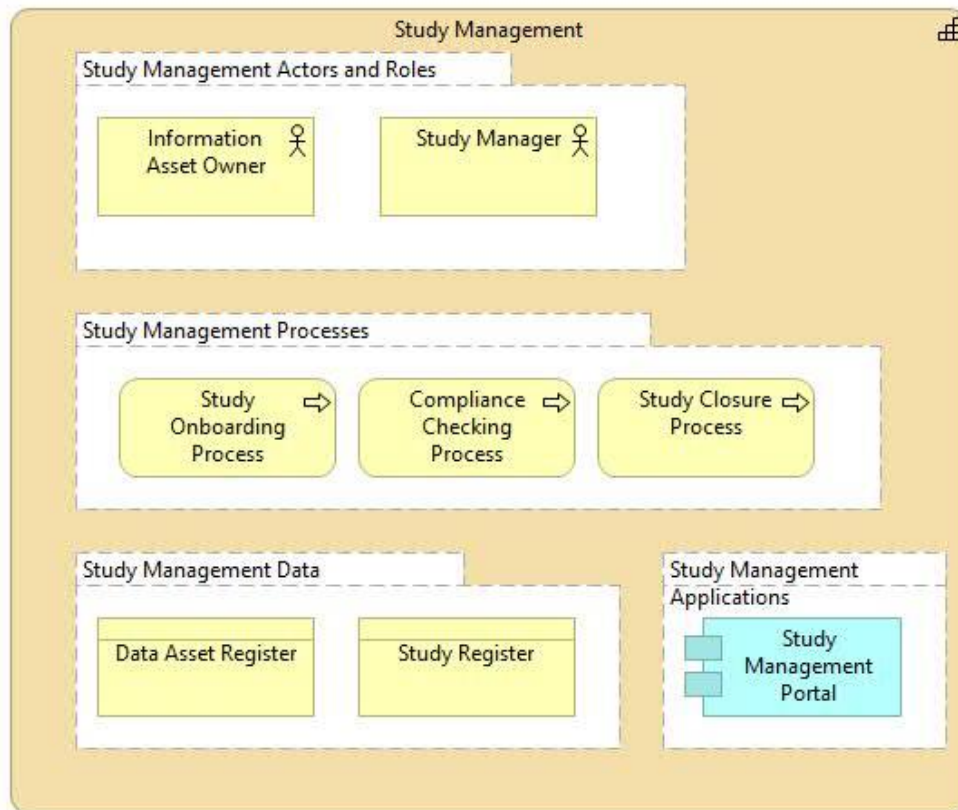
A TRE may have a single study or contain many separate studies. Where there are multiple studies, each study must be separately identified and logically isolated from others.

A study (sometimes called a project or work package) has a lifecycle. Beginning with an initial setup to bring a study onboard. Once within the TRE, compliance with internal and external standards must be checked as the project evolves and regulations and policies change.

The study onboarding process is triggered by the information asset owner. Compliance checks can be triggered by the information asset owner and the onboarding process or through a scheduled event like an audit or change. Study closure is triggered by the information asset owner or a scheduled event such as a contract termination.

Within the data asset and study register critical information related to the study will be maintained. These metadata will include study role information, contracts, data management plans and finance information along with links to quality data related to the study held, for example, user permissions and training records.

In order to allow owners and managers to manage their studies data will be provided. Ideally this would be through a portal. The portal can also provide a self-service interface to ensure the study data is correct within the registers.

3.5. Member Accreditation

Ability to ensure that people with access to data are identified correctly and they are suitably qualified.
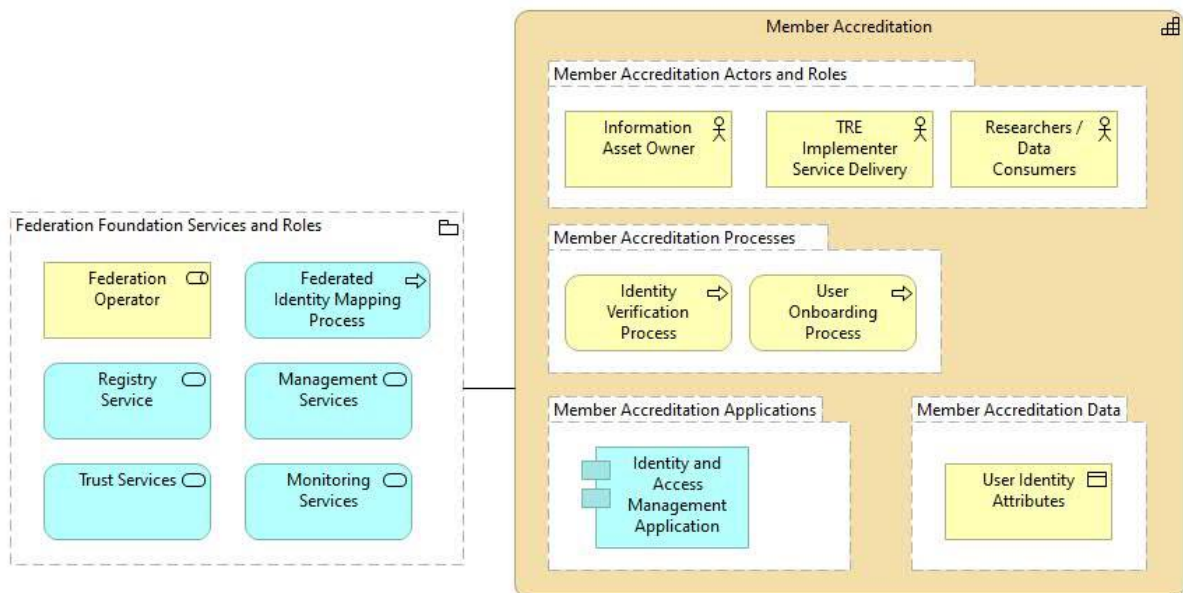

Information asset owners make access changes to the permissions on assets they own to provide access to researchers and data consumers. These changes trigger the identification process to ensure that the right user is given appropriate access to the resource. Once identified, the onboarding process accesses the user identity attribute data.

Member accreditation identifies TRE users and brings them into the environment. All users within the TRE will pass through the identity verification process. Data related to the person will be recorded after which they are issued with an identity, password and other authentication factors. This process will also involve agreement of fair usage statements and contracts etc.

Federation Foundation Services are a key element in the federation of identity across multiple TREs. Details can be found in the Federated Architecture Blueprint, DARE UK Delivery Team.

3.6. Training Delivery and Management

Ability to deliver, track and maintain adequate training levels to ensure competence of all people within the TRE organisation.
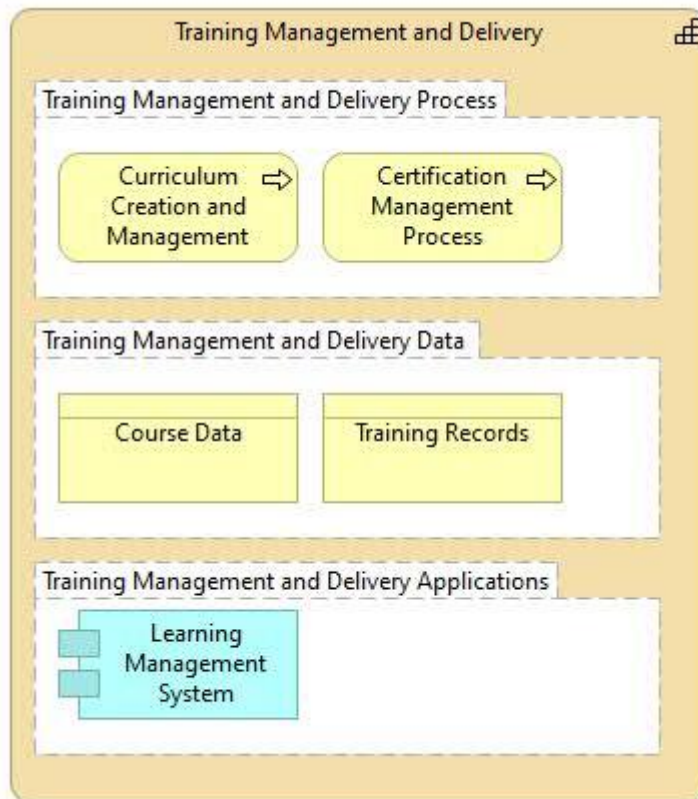
Course data is created and maintained by the curriculum creation and management process. Certification is issued as part of the training record following the completion of courses using the learning management system application.

Competency of people within the TRE is a key control for risk across multiple roles and processes. To ensure competency, a training curriculum should be developed with training needs aligned to roles within the TRE. This training should be delivered through a learning management system and certification and training records maintained for users within the TRE. Competency maybe proven through professional qualifications or external training in which case certifications may be accepted when issued by a trusted third party.

3.7. Computing Technology

What the organisation does to manage systems for storing, processing, retrieving, and sending information.
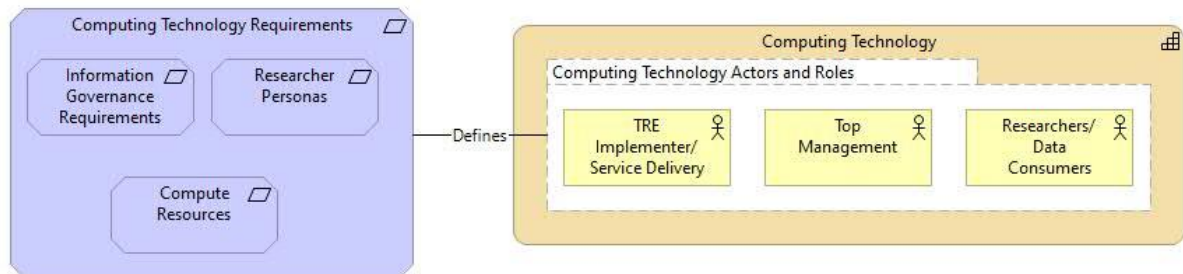
Computing technology requirements are defined by top management within the TRE organisation in collaboration with TRE operators and developers and researchers themselves. Requirements are drawn from information governance requirements which primarily specify the controls needed within the technical and computing infrastructure. Researcher personas influence them according to the technical knowledge and experience of researchers along with the work they need to perform within the system, for example, a data scientist will have very different requirements to a clinician. The required compute resources will vary according to the scale of data and computational techniques employed during research.

3.8. End User Computing

The ability of the TRE operator to provide and manage devices, workspaces, interfaces and applications used by researchers to interact with underlying systems and data.
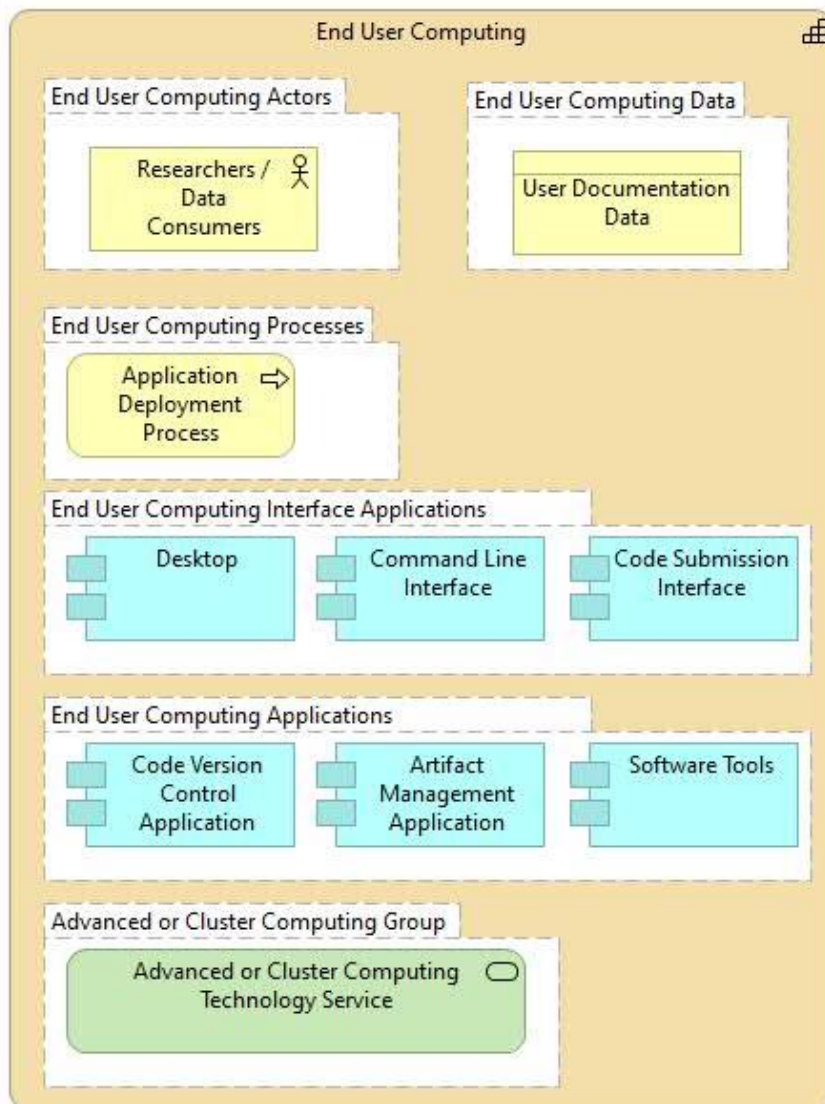
End user computing is realised by interfaces and applications providing researchers and data consumers access to data. Changing researcher requirements trigger the application deployment process to update applications and interfaces. The advanced or cluster computing technology service provides resources for advanced analytics capabilities Machine Learning.

The TRE may contain a single or multiple end user computing environments. Multiple computing interfaces provide familiar and flexible ways for users to interact with data. Interfaces may also provide an opportunity to obfuscate data by allowing the submission of code without providing direct access to the data itself. Tools to version control code support reproducible science. Artifact management applications allow users limited access to external repositories outside the environment.

3.9. Infrastructure Management

> **The ability of the TRE operator to instantiate, deploy, change or remove deployed infrastructure.**
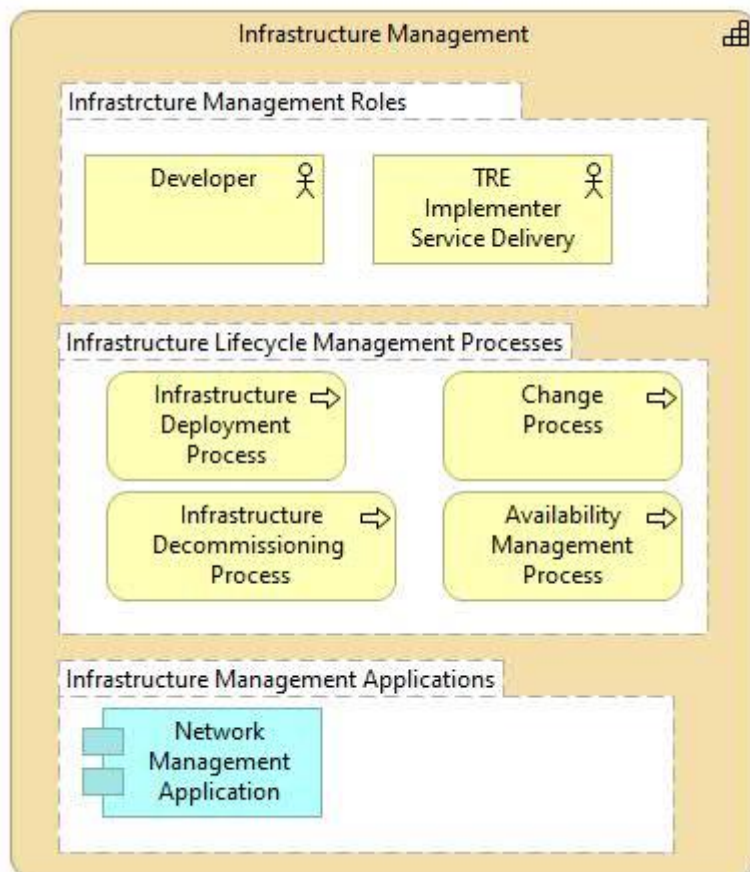
Infrastructure developers and TRE operators deploy and manage infrastructure within the TRE. Infrastructure is likely to be in part or fully virtualised allowing the automation of deployment processes. The change process ensures infrastructure changes are risk assessed, carefully managed and recorded. Infrastructure is decommissioned ensuring data and infrastructure services are managed off the platform. Availability management ensures infrastructure maintains uptime with data and systems available to the level agreed with service users. There will be a set of management applications including, network management to ensure the technical environment remains suitably isolated.

3.10.   Capacity Management

What the organisation does to ensure the right amount of resources are available at the right time to provide a service.
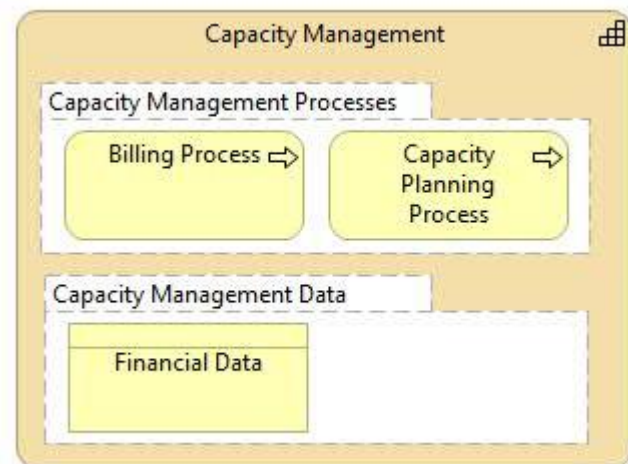
Capacity planning looks at patterns of activity, compute and storage requirements in order to determine suitable resources for deployment into the TRE infrastructure. The main control for capacity is finance and billing processes to ensure operational and capital costs can be recovered to ensure the TRE is sustainable and remains available to users. Capacity increases identified in the planning process trigger the billing process and update financial data.

3.11.   Configuration Management

What the organisation does to identify, maintain, and verify information on IT assets and configurations in the TRE organisation.
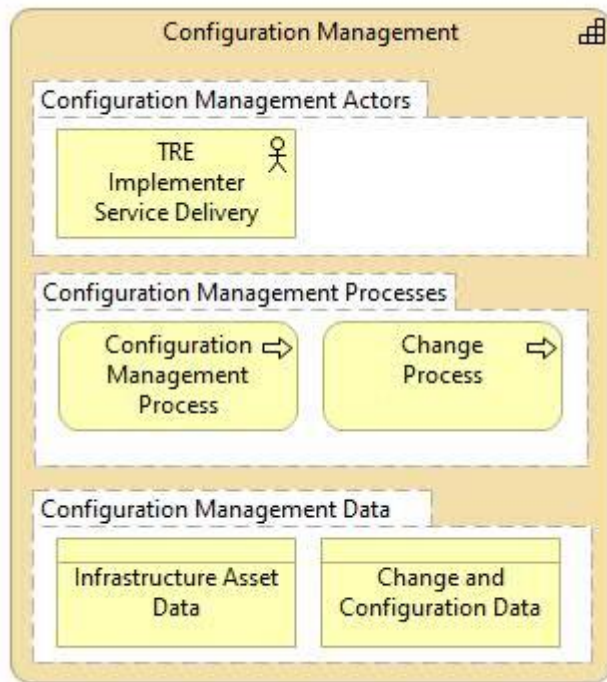
In order to control risk within the TRE infrastructure, configuration should be carefully designed and planned through the configuration management processes. Changes to the infrastructure trigger updates to infrastructure asset data and configuration data.
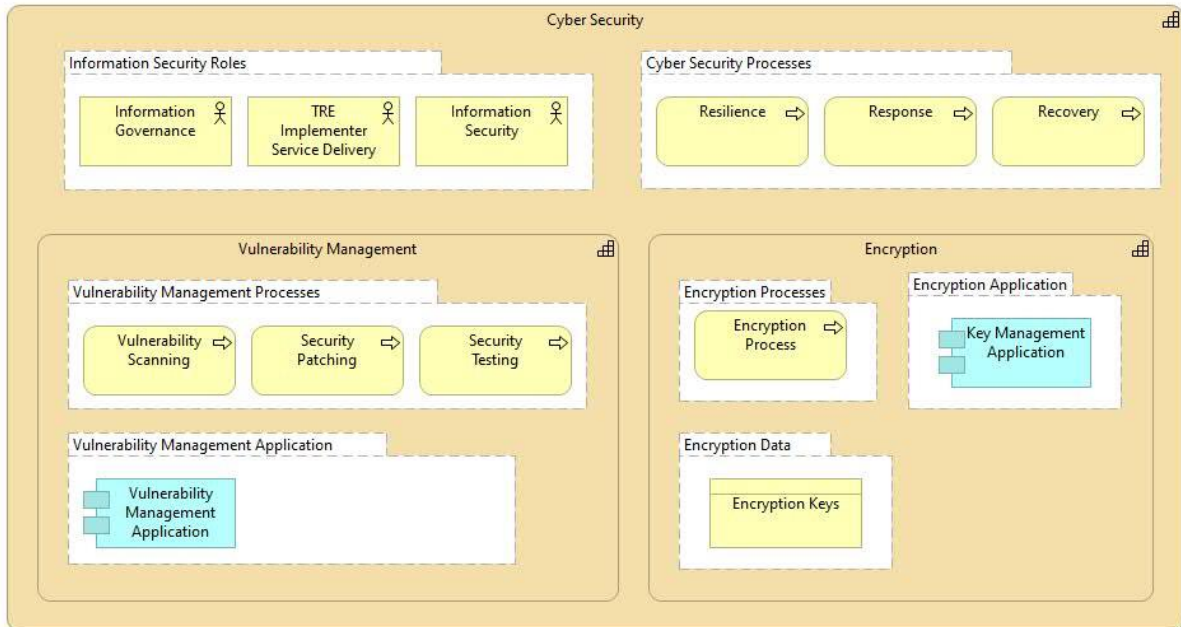
3.12.   Cyber Security

> **What the organisation does to safeguard research to ensure the confidentiality, integrity and availability of research resources and data.**

Cyber security is realised by three processes and two additional capabilities. Resilience ensures the TRE organisation and infrastructure continues to function when threats are realised, or an incident occurs. Following an incident, the response process is triggered and the organisation reacts to reduce risk. Once the immediate response is complete, recovery process is triggered to restore the organisation's services to make data available.

3.13.   Vulnerability Management

> **The ability of the TRE operator to identify, assess, report on, manage and remediate technical vulnerabilities across endpoints, workloads, and systems.**
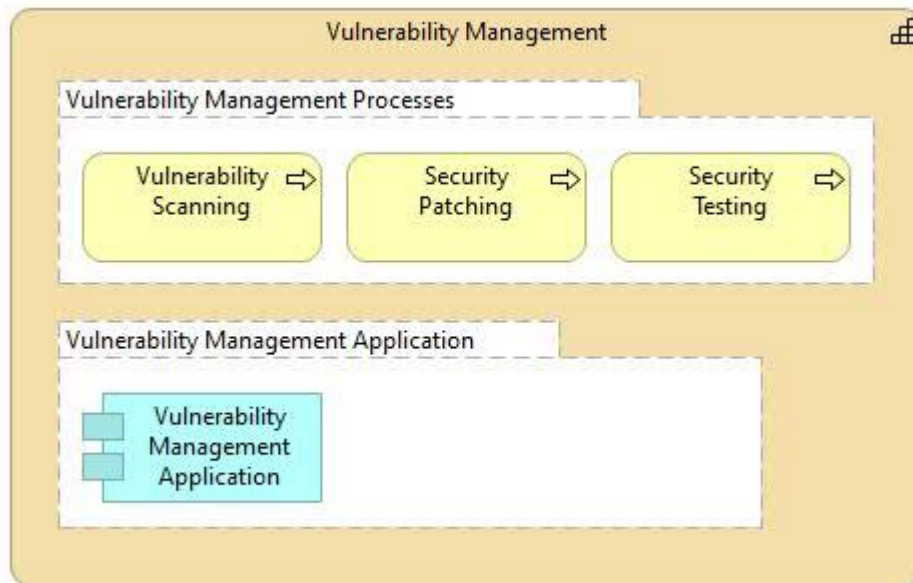
Security patching pre-emptively fixes vulnerabilities within the system. This is usually via automatically utilising software vendor issued updates. Patching may also be triggered by the vulnerability scanning or security testing process. One or more vulnerability management applications serve the processes.

3.14.   Encryption

The ability of the TRE organisation to deploy and manage encryption to protect information assets, including data for TRE research data.
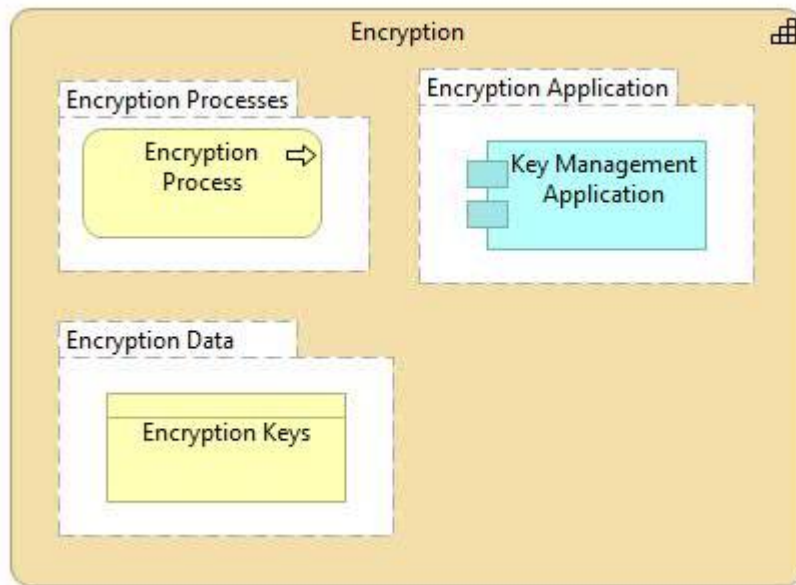
The encryption process is triggered when data is ingressed, egressed or moved within the TRE to ensure it is securely transmitted or stored. The key management application serves the encryption process and updates the encryption key data.

3.15. Data Management

The ability of the TRE operator to manage data assets and ensure information remains secure.
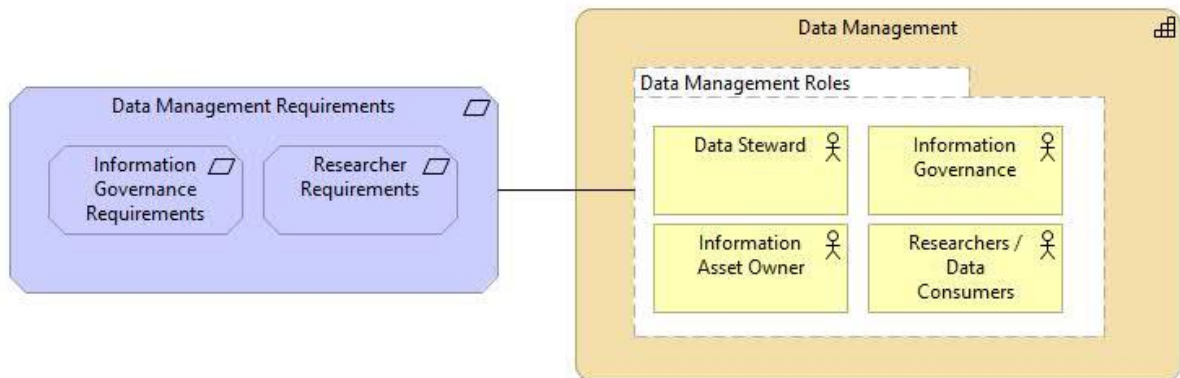
Data management requirements include information governance requirements and researcher requirements. They are defined by the information governance roles and by the owners, users and managers or data.

3.16.   Data Lifecycle Management

> **What the TRE organisation does to manage data assets and ensure information remains secure.**

Data lifecycle management processes manage research data throughout the lifecycle of data and studies. Data lifecycle management processes are triggered as research data are ingressed, egressed or moved within the TRE. Research metadata is updated by these processes with the data asset regi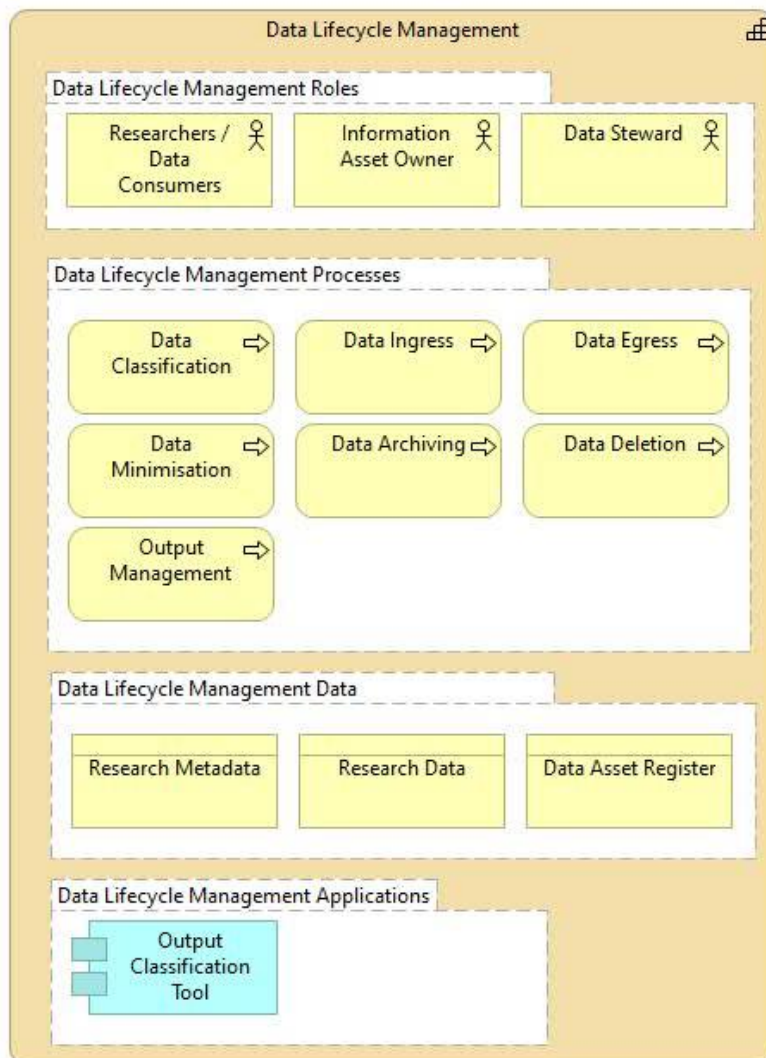ster being updated when data ingresses or egresses. The Data classification application serves the classification process data is classified according to information risk this process can occur prior to ingressed, moved or egressed to ensure the environment is suitably controlled. Guidance can be found in Design choices for productive, secure, data-intensive research at scale in the cloud . One or more data transfer applications serve the ingress and egress application.

Researcher requirements to create or transfer safe outputs from the TRE trigger the data minimisation and output management processes including disclosure control.

3.17.   Identity and Access Management

> **The ability of the TRE operator to ensure the right people (identities) can only access the tools and data they need.**
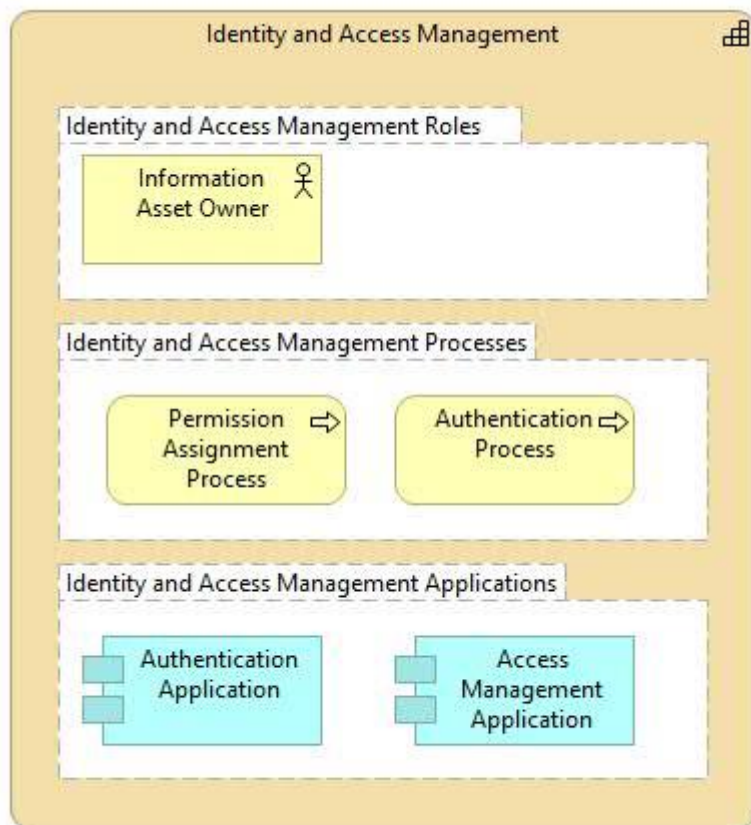
Identity and access management processes manage appropriate access to data during the data lifecycle. The permission assignment process is triggered by an information asset owner to change permissions for a researcher on research data. This process is served by the access management application. The authentication process identifies users and issues the correct permissions at logon.

3.18.   Information Search and Discovery

The ability to query and browse the data within an environment at various levels of abstraction.
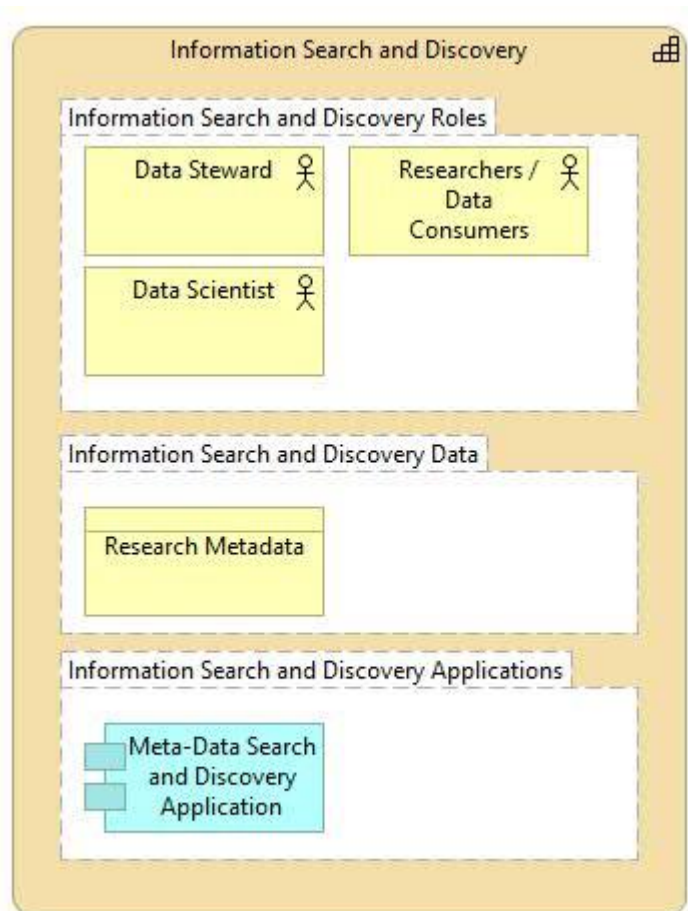
Roles that utilise research data within the TRE utilise the metadata search and discovery application to query and understand the contents of the data prior to analysis. This can be used to identify data, refine research questions and plan analysis without access to the underlying data.

4. Additional Views

4.1. Five Safes Process View

The view below shows the alignment of processes within the SATRE architecture to the five safes model. Each process does not exclusively deliver to that safe within the framework. Processes such as change, issue management and audit in particular deliver to all 5 safes.

Safe Settings

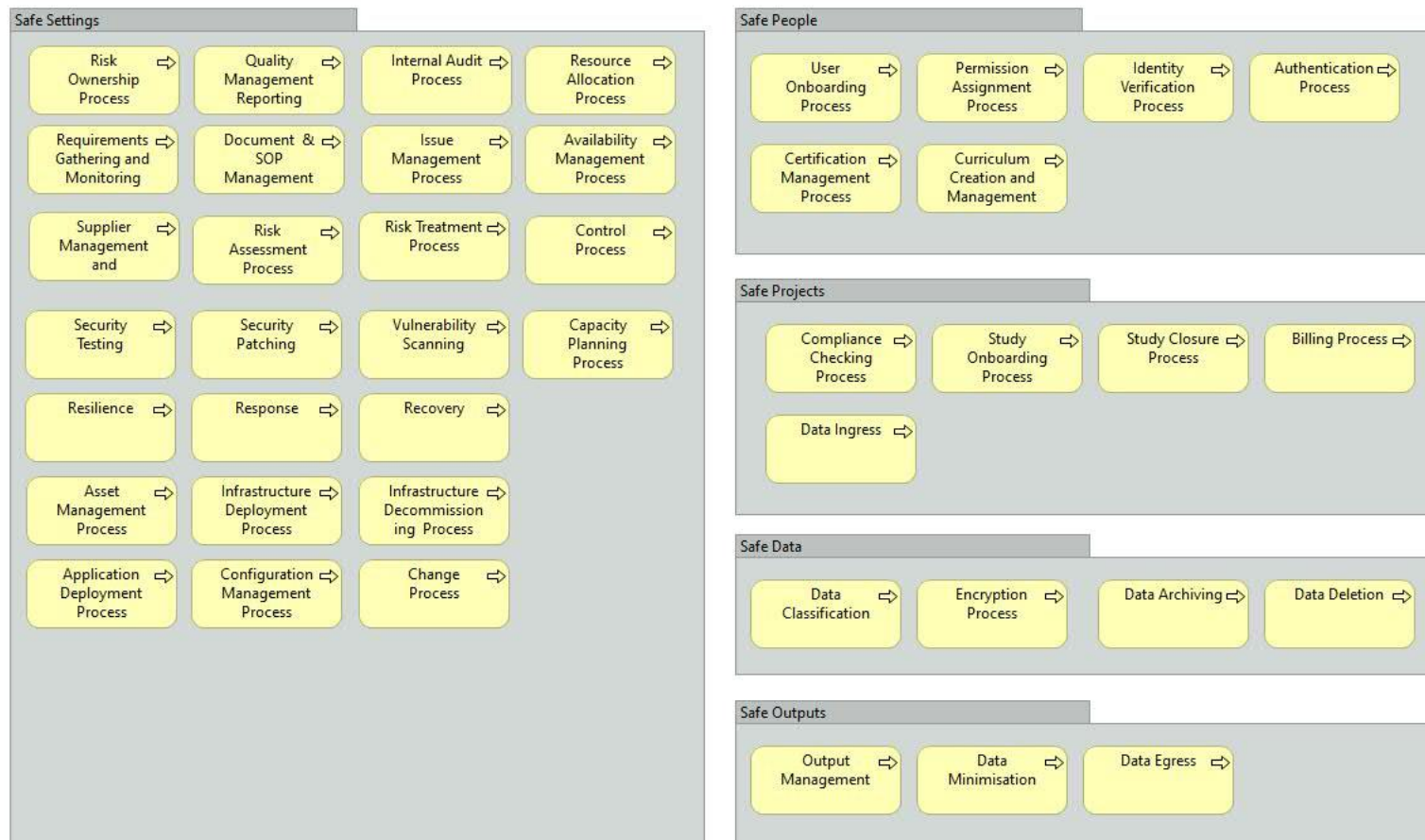|			
Risk Ownership Process	Quality Management Reporting	Internal Audit Process	Resource Allocation Process
Requirements Gathering and Monitoring	Document & SOP Management	Issue Management Process	Availability Management Process
Supplier Management and	Risk Assessment Process	Risk Treatment Process	Control Process
Security Testing	Security Patching	Vulnerability Scanning	Capacity Planning Process
Resilience	Response	Recovery	
Asset Management Process	Infrastructure Deployment Process	Infrastructure Decommissioning Process	
Application Deployment Process	Configuration Management Process	Change Process	

Safe People

User Onboarding Process	Permission Assignment Process	Identity Verification Process	Authentication Process
Certification Management Process	Curriculum Creation and Management		

Safe Projects

Compliance Checking Process	Study Onboarding Process	Study Closure Process	Billing Process
Data Ingress			

Safe Data

Data Classification	Encryption Process	Data Archiving	Data Deletion

Safe Outputs

Output Management	Data Minimisation	Data Egress

5. Bibliography

Arenas, D., Atkins, J., Claire, A., David, B., Cabrejas Egea, A., Carlysle-Davies, S., . . . Whitaker, K. (2019). Design choices for productive, secure, data-intensive research at scale in the cloud. *arXiv*. Retrieved from https://arxiv.org/abs/1908.08737

Cole, D. C., Li, D. S., Gao, D. C., Sutherland, D. J., Beggs, J., & Chuter, A. (2022). TREEHOOSE: Trusted Research Environment and Enclave for Hosting Open Original Science Exploration. Retrieved from https://doi.org/10.5281/zenodo.7085505

DARE UK. (n.d.). Retrieved from dareuk.org.uk: https://dareuk.org.uk/wp-content/uploads/2022/08/DARE_UK-Paving_the_way_coordinated_national_infrastructure_sensitive_data_research-Aug2022.pdf

DARE UK. (2023). DARE UK Federated Architecture. https://dareuk.org.uk/our-work/federated-architecture-blueprint/.

HDRUK Trusted Research Envionrments. (n.d.). Retrieved from HDRUK: https://www.hdruk.ac.uk/access-to-health-data/trusted-research-environments/

Hosiaisluoma, E. (2022). *Archimate Cookbook.* Retrieved from https://www.hosiaisluoma.fi/ArchiMate-Cookbook.pdf

International Organization for Standardization. (2022). ISO/IEC 27001 Information security management systems. Retrieved from https://www.iso.org/standard/27001

Lourinho, R. &. (2017). Mapping and Integration of Enterprise Governance of IT Practices. Retrieved from https://www.researchgate.net/publication/319115313_Mapping_of_Enterprise_Governance_of_IT_Practices_Metamodels

OpenGroup. (n.d.). ArchiMate® 3.2 Specification. Retrieved from https://pubs.opengroup.org/architecture/archimate3-doc/

SATRE. (2023). *SATRE Standard*. Retrieved from https://satre-specification.readthedocs.io/

Varma, S., Hubbard, T., Seymour, D., Brassington, N., & Madden, S. (2021). Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems. Retrieved from Building Trusted Research Environments; Principles and Best Practices; towards TRE ecosystems: https://zenodo.org/record/5767586