eosc | ENTRUST

European Network of Trusted Research Environments

Deliverable D13.1

Draft Roadmap for EOSC-ENTRUST Blueprint

Project Title (Grant agreement number)	EOSC-ENTRUST Grant Agreement 101131056		
Project Acronym (EC call)	EOSC-ENTRUST		
WP No & Title	WP13: Trusted research environment blueprint - Establishing a common terminology and approach		
WP Leaders	Stefan Negru (CSC), Pål Sætrom (NTNU)		
Deliverable Lead Beneficiary	2. CSC & 12. UiB		
Contractual delivery date	31/05/2024	**Actual delivery date**	28/06/2024
Delayed	Yes		
Partner(s) contributing to this deliverable	CSC, UiB		
Authors	Pål Sætrom (NTNU) https://orcid.org/0000-0001-8142-7441 Stefan Negru (CSC) https://orcid.org/0000-0002-6544-5022		
Contributors	Miikka Kallberg (CSC) https://orcid.org/0000-0002-1457-3999 Heikki Lehväslaiho (CSC) https://orcid.org/0000-0002-6263-1356 Stefanie Kirschenmann (CSC) https://orcid.org/0000-0002-2773-4798 Susanna Repo (CSC) https://orcid.org/0000-0003-3488-4767 Korbinian Bösl (UiB) https://orcid.org/0000-0003-0498-4273 Ingeborg Winge (ELIXIR-NO) https://orcid.org/0000-0001-6139-5999 Stian Soiland-Reyes (UNIMAN) https://orcid.org/0000-0001-9842-9718		

Funded by
the European Union

Acknowledgements (not grant participants)	Jonathan Tedds (ELIXIR Hub) - https://orcid.org/0000-0003-2829-4584 Monica Abrudan (ELIXIR Hub) https://orcid.org/0000-0003-0228-4353
Reviewers	Heidi Laine (CSC) https://orcid.org/0000-0002-6658-0664 Jan-Willem Boiten (Lygature) https://orcid.org/0000-0003-0327-638X

Log of changes

Date	Mvm	Who	Description
07/06/2024		Pål Sætrom, Miikka Kallberg Stefanie Kirschenmann	Restructuring of the deliverable according to reviewers' feedback

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Table of contents

1. Executive Summary **3**
2. Abbreviations **5**
3. Introduction **5**
4. Contribution towards project objectives **5**
5. Methods **7**
5.1. Deliverable scope 7
5.2. Architecture 8
6. Description of work accomplished **9**
6.1 Workshop Agenda 9
6.2 Methods for gathering input 11
7. Results **12**
7.1 Blueprint roadmap 12
7.2 Roadmap themes 13
7.2.1 Potential conceptual frameworks for a blueprint 13
7.2.2 Interoperability requirements from Drivers 14
7.2.3 Interoperability capabilities from Providers 16
7.2.4 Input from aligning projects 17
7.2.5 Input from open discussion on interoperability 17
7.2.6 Input from open discussion on governance 18
7.2.7 Input from open discussion on standards and interfaces 19
8. Discussion **20**
9. Conclusions **22**
10. Impact **22**

1. Executive Summary

EOSC-ENTRUST aims to create a European network of Trusted Research Environments (TREs) for sensitive data and drive European interoperability between TREs by development of a common blueprint for federated data access and analysis – the EOSC-ENTRUST Blueprint & Interoperability Framework (Blueprint, for short). This document is the first version of the EOSC-ENTRUST Blueprint Roadmap (Roadmap, for short), presenting the steps needed to arrive at the Blueprint, the architecture modelling framework chosen for the work, and the Blueprint's initial set of main requirements.

The Roadmap is divided into three phases, consisting of (Phase 1, 2024) identifying Blueprint requirements and mapping these to existing TRE capabilities, (Phase 2, 2025) developing and validating key components of the Blueprint, and (Phase 3, 2026) releasing the final Blueprint. Along with the Blueprint, we will develop and release training material for the Blueprint and a TRE Provider Catalogue describing the interoperability capabilities. Updated versions of the Blueprint, training package, and provider catalogue are scheduled for November each year. To support this work and ensure alignment with use cases (Drivers) and TRE providers, we will arrange yearly workshops focused on requirements and capabilities (May) and evaluation and adoption (September). We will revise the Roadmap (May '25 and '26) based on the workshop outcomes.

The Blueprint's initial set of main requirements is primarily based on the outcomes of the First EOSC-ENTRUST Requirements and Capabilities Workshop, held on May 7-8, 2024. The workshop identified the following five main categories of requirements for the Blueprint: 1) Data transfer between environments, 2) User identity and access management, 3) Governance and compliance, 4) Data lifecycle management, and 5) User training and certification. The requirements, along with the workshop presentations and discussions, emphasised the need for a diverse portfolio of TRE solutions, each providing specific technical solutions and security measures that collectively cover the needs of diverse scientific domains and private actors. Consequently, the Blueprint should focus on the services and architecture needed for TRE interoperability instead of on specific technologies and technical details.

2. Abbreviations

ECRIN: European Clinical Research Infrastructure Network
EHDS: European Health Data Space
eID: Electronic identification
FEGA: Federated European Genome-Phenome Archive
GA4GH: Global Alliance for Genomics and Health
HDAB: Health Data Access Body
HPC: High-Performance Computing
SATRE: Standard Architecture for Trusted Research Environments
SPE: Secure Processing Environment
SSHOC: Social Sciences & Humanities Open Cloud
TOM: Technical and Organisational Measure
TRE: Trusted Research Environment
W3C: World Wide Web Consortium

3. Introduction

The EOSC-ENTRUST project aims to create a European network of Trusted Research Environments (TREs) for sensitive data and drive European interoperability between TREs by development of a common blueprint for federated data access and analysis – the EOSC-ENTRUST Blueprint & Interoperability Framework (Blueprint, for short). This document is EOSC-ENTRUST Deliverable D13.1 (D13.1 deliverable, for short) and is the first version of the EOSC-ENTRUST Blueprint Roadmap (Roadmap, for short), presenting the steps needed to arrive at the Blueprint, the architecture modelling framework chosen for the Blueprint work, and the Blueprint's initial set of main requirements.

The following sections describe the D13.1 deliverable's contributions towards the project objectives, the methods used, the accomplished work, and results before discussing, concluding, and outlining the deliverable's impact.

4. Contribution towards project objectives

The D13.1 deliverable has contributed to the following project objectives:

	Key Result No and description	Contributed
Objective 1 Create a European network of	1. A catalogue of suitable national or institutional TREs as part of the EOSC offering (WP4, WP5)	Yes
	2. A 'starter pack' of exemplar projects to demonstrate how networks of TREs can address European research priorities (WP2, WP3)	No

Trusted Research Environments, linked to EOSC and EuroHPC, to enable transnational collaborative research on sensitive or restricted data.	3. European researchers are aware of capabilities through communication and outreach events (WP2) and materials delivered to support national TRE training programmes (WP2, WP5)	No
	4. Enable federated use via standards and technology for trusted researcher identity, data use and data access linked to developing European framework for trusted electronic identification of individuals (WP6)	No
	5. Enable researchers and software developers to deploy across multiple TREs via secure FAIR digital objects and workflows (WP6)	No
	6. EuroHPC capacity that meets the need for secure exascale and GPU (e.g., AI) computing can be identified and connected using the EOSC-ENTRUST framework (WP4, WP5).	No
Objective 2 Trusted Research Environment providers implement, validate, and promote their capabilities through a European framework using common standards and shared legal, operational and technical language.	1. An established European network of national and institutional TRE Providers (WP4)	No
	2. A service blueprint that allows technical interoperability between TRE based on the EOSC Interoperability framework (WP5)	Yes
	3. National and institutional TREs consistently set out their capabilities with common representation for validated legal, operational, semantics and technical aspects (WP4).	No
	4. Define the security baseline and auditing procedures for TREs to support the Five Safes[1] principles and capture requirements in guidelines for FAIR sensitive data in EOSC (WP5)	No
	5. Drive TRE composability via policy and process interoperability and set out an EOSC compliant governance model for a TRE services network (WP4, WP5).	Yes
Objective 3 National funders and	1. A machine-readable catalogue of TRE capabilities allowing detailed, comparative analysis of technical capabilities and identification of gaps (WP4, WP5).	No

[1] Ritchie, F. (2017, September). The "Five Safes": A framework for planning, designing and evaluating data access solutions. Paper presented at Data for Policy 2017, London, UK https://doi.org/10.5281/zenodo.897821

governments understand the network of TRE capabilities serving their needs, and how TREs support their national priorities and their contributions to selected transnational programmes	2. Policy briefs on the capabilities of the European TRE Provider Forum (WP2, WP4) and Use Cases of their application in research domains of high societal impact (WP3).	No
	3. Connection between the EOSC-ENTRUST Provider Forum and the European Data Spaces (WP1, WP2).	No
Objective 4 The European Network of Trusted Research Environments (ENTRUST) is embedded in the European Open Science Cloud and the European Data Spaces and fosters an ecosystem of public, private and joint-venture providers of TRE services.	1. National and organisational providers are incorporated into EOSC via national members and the European network forms part of EOSC long-term strategy (WP1, WP2).	No
	2. The emerging European Data Spaces build their capabilities on the network of existing and developing TRE providers (WP2).	No
	3. Technological developments required by one Data Space activity can be directed to a forum of TRE specialists, reducing the need for duplication and coordinating investment in foundational technologies (WP4).	No
	4. A driver project to demonstrate opportunities for public-private partnerships (WP3)	

5. Methods

5.1. Deliverable scope

The D13.1 deliverable is the initial version of the EOSC-ENTRUST Blueprint Roadmap document presenting the steps needed to develop an EOSC-ENTRUST blueprint for building an interoperable network of TRE services. The roadmap is primarily based on the First EOSC-ENTRUST Requirements and Capabilities Workshop (RC1 Workshop), held on May

7-8, 2024, as a virtual workshop on Zoom, organised by the Architecture work package (WP13).

The workshop started the process of gathering requirements from Drivers (WP7) and mapping these to existing capabilities within the TRE Provider forum (WP10). Specifically, the workshop sought to define a set of minimal requirements for a blueprint, a minimal conceptual description of what a blueprint is, and the terminology to use. Moreover, the workshop sought to identify needed new software development work by providing concrete examples of existing environments, to gather expectations from providers and drivers, focusing especially on TRE interoperability, and to identify common challenges or "pain points" that could hinder interoperability. Whereas the workshop also sought to gather input for the TRE inventory, including its high-level information structure and specific information needs, here we focus on the workshop outcomes contributing to the initial EOSC-ENTRUST Blueprint Roadmap.

5.2. Architecture

We use an enterprise architecture methodology following the ArchiMate[2] 3.2 specification to understand and model the underlying strategic decision-making process. This allows us to use well-defined vocabulary to express the model without having to specify everything from scratch. In addition, it enables being more formal and explicit about the aims, restrictions, and discrepancies underlying various points of view. In this document, we have used enterprise architecture to capture and model capabilities necessary for the Blueprint, as identified during the RC1 Workshop.

We need to clearly define the conditions for the diverse and sometimes conflicting requirements that influence the capabilities of Secure Processing Environments (SPEs), including their Technical Organisational Measures (TOMs) (Figure 5.1). When needed for clarity, we will then be able to specify resources that are components of capabilities. They are best divided into three main categories: (i) competencies and human resources, (ii) processes and models, and (iii) data and information systems. These resource categories are reflected at the business level when implemented services are understood to be explained by (i) roles, (ii) processes, and (iii) objects. Existing business services can then be compared to requirements.

[2] https://www.archimatetool.com/

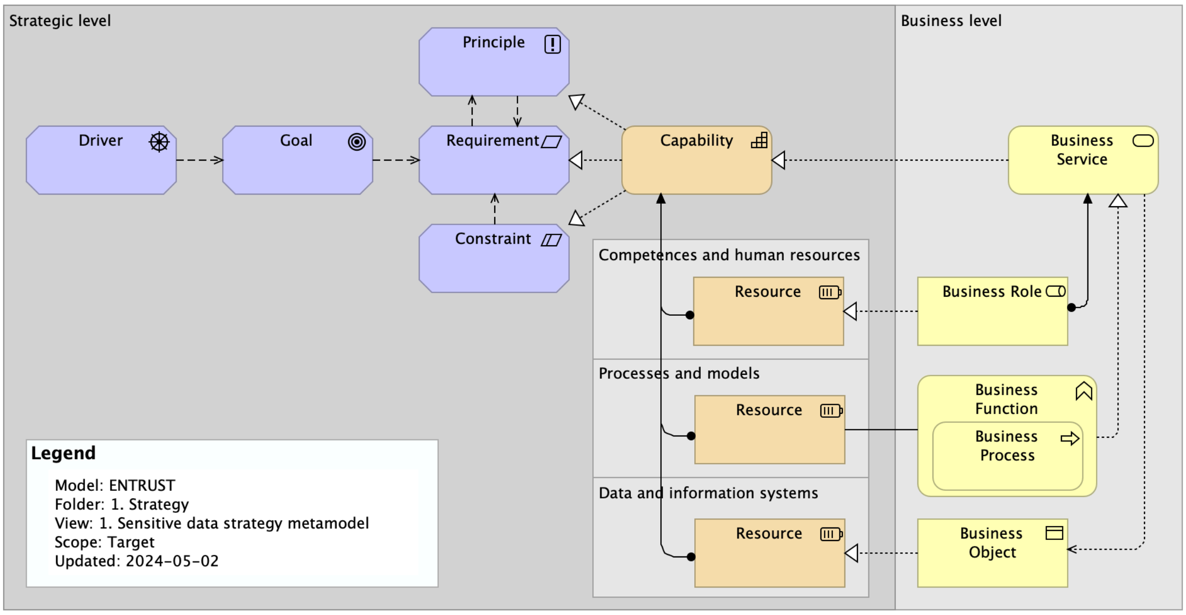

Figure 5.1. Metamodel for capturing the strategic and business levels of SPEs and their TOMs. The strategic level items are not directly related to the business level, but affect the business level decisions.

6. Description of work accomplished

The first milestone of the Architecture work package (WP 13) has been to organise the RC1 Workshop. Beyond bringing together participants from WP13, this event included contributions from other WPs, especially from the Drivers and Providers Forum (WP7 and WP10, respectively), as well as presentations on associated projects. As such, it provided a forum for the discussion of the project's objectives from different perspectives.

The following sections describe the RC1 Workshop agenda and the methods used for gathering and recording input and information from workshop attendees during the event.

The workshop's outcomes, including collaborative minutes and presentations, were saved in the dedicated EOSC-ENTRUST Google Drive folder and presentations were published in Zenodo[3].

6.1 Workshop Agenda

The virtual event covered two half days: The first day focused on the introduction and the Drivers' and Providers' perspectives, the second day on associated projects as well as discussions on the Roadmap.

[3] https://doi.org/10.5281/zenodo.11221124

Agenda May 7 - Requirements and capabilities

Time	Session	Presenter
12:00-12:30	**Opening** • Words of welcome • Housekeeping • Round-table of introductions • Objectives of the workshop	Susanna Repo Pål Sætrom
12:30-13:30	**Introduction** • Five safes • First architecture overview • Discussion	Laura Ward/Chris Cole Heikki Lehväslaiho
13:30-13:45	**Break**	
13:45-14:45	**Session on requirements (Drivers)** • Introduction • Driver 1 - Federated Human Genomics • Driver 2 - (SSHOC) Lessons learned - international remote access connections between TREs • Driver 3 - Clinical Research • Driver 4 - Health & Environmental Science in PPP • Discussion	Jan-Willem Boiten Jordi Rambla de Argila Beate Lichtwardt Sergio Contrino Anne-Marie Tuikka
14:45-15:00	**Break**	
15:00-16:00	**Session on capabilities (Providers)**	10+5 minutes each
	• Presentation on SATRE • Presentation on Tryggve • Presentation on de.NBI	• Simon Li/Chris Cole • Abdulrahman Azab • Nils Hoffmann, Fabian Paz, Christian Buggedei
	Discussion	
16:00-16:15	**Closing and preparing for the next day**	Susanna Repo

Agenda May 8 - Blueprint Roadmap

Time	Session	Presenter
12:00-12:30	**Opening** • Welcome • Discussion on main takeaways from previous day	Pål Sætrom

12:30-13:15	**TRE inventory and survey** ● TRE inventory / survey information model ● Timeline for the survey ● Invitation to contribute - link to survey	Heidi Laine
13:15-13:30	**Break**	
13:30-14:15	**Input from aligning projects** ● Presentation on TEHDAS2 - Alignment with EOSC-ENTRUST ● Presentation on EHDS-NORTRE gap analysis ● Discussion	Helena Lodenius Christine Stansberg
14:15-14:30	**Break**	
14:30-16:00	**Open discussion on the blueprint roadmap** **Breakout rooms:** ● Interoperability ● Governance ● Standards and interfaces	Susanna Repo
16:00-16:15	**Closing remarks**	Pål Sætrom Susanna Repo

6.2 Methods for gathering input

The presenters in the Drivers and Providers sessions at the RC1 Workshop were each given presentation templates and asked to address specific questions related to the workshop's goals.

Drivers were given the following leading questions.

(i) What are your expectations from a TRE/SPE. Specifically, what are the musts, shoulds, and won'ts in terms of interoperability standards (e.g. GA4GH, W3C), governance, account management (e.g. Federated login), and capabilities.

(ii) Name 3-4 main challenges when using a TRE/SPE.

Providers were given the following leading questions.

(i) In your TRE, what are the musts, shoulds, won'ts as a provider. Specifically, what do you consider the essential priorities, what are the areas where you are willing to negotiate to achieve consensus, and what are the issues that you are indifferent to or have minimal concern about?

(ii) Comparing your TRE to SATRE (Standard Architecture for TREs), is SATRE a good reference framework for assessing a TRE (what is missing; what is considered adequate), and have you identified alternative frameworks for evaluating the capabilities of a TRE/SPE?

Collaborative notes were taken throughout the meeting, including discussions and questions posed orally or in the chat. A summary on how these inputs relate to the Roadmap is given in Section 7.

7. Results

The following two sections present the draft Roadmap for the EOSC-ENTRUST Blueprint & Interoperability Framework and relevant themes for the Roadmap, identified in the RC1 Workshop.

7.1 Blueprint roadmap

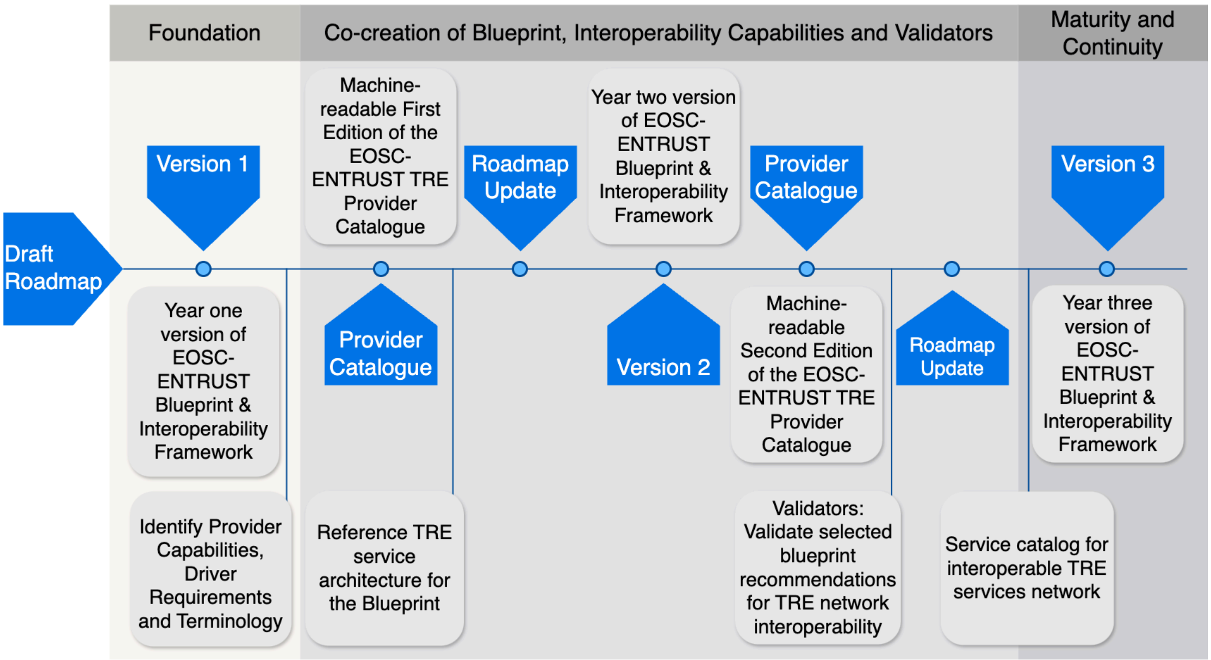
Foundation
Co-creation of Blueprint, Interoperability Capabilities and Validators
Maturity and Continuity
Draft Roadmap
Version 1
Year one version of EOSC-ENTRUST Blueprint & Interoperability Framework
Identify Provider Capabilities, Driver Requirements and Terminology
Machine-readable First Edition of the EOSC-ENTRUST TRE Provider Catalogue
Provider Catalogue
Reference TRE service architecture for the Blueprint
Roadmap Update
Year two version of EOSC-ENTRUST Blueprint & Interoperability Framework
Version 2
Provider Catalogue
Machine-readable Second Edition of the EOSC-ENTRUST TRE Provider Catalogue
Validators: Validate selected blueprint recommendations for TRE network interoperability
Roadmap Update
Service catalog for interoperable TRE services network
Version 3
Year three version of EOSC-ENTRUST Blueprint & Interoperability Framework

Figure 7.1: Development steps and expected updates.

The RC1 Workshop was the first step in identifying requirements for the EOSC-ENTRUST Blueprint & Interoperability Framework and gathering current interoperability capabilities of TRE Providers. The following Section will elaborate on requirements, capabilities, and other themes relevant for the blueprint roadmap (Figure 7.1). Along with ongoing work in the Driver and Provider forums, these requirements and capabilities will be taken forward in the project's next steps, which will lead to the Year one baseline version of the EOSC-ENTRUST Blueprint & Interoperability Framework (D5A.3/D13.4). Specifically, we will:

1. **Aug. '24**: Update the requirement list based on further input from the Driver forum (Milestone M3A.1/M7);

2. **Sept. ‘24**: Run the first EOSC-ENTRUST Evaluation & Adoption Workshop (Milestone M5A.2/M14);
3. **Nov. ‘24**: Create a machine-readable First Edition of the EOSC-ENTRUST TRE Provider Catalogue (D5A.2/D13.3), based on the TRE inventory survey (Milestone M4A.2/M10);
4. **Nov. ‘24:** Initial report on selected Validators in NORTRE (M5A.3/M15);
5. **Nov. ‘24:** Create a training package for the year one Blueprint & Interoperability Framework (D5A.4/D13.2);
6. **Nov. ‘24**: Document the first version of the EOSC-ENTRUST Blueprint & Interoperability Framework (D5A.3/D13.4).

We plan to do the main development work and update the Provider Catalogue, Blueprint & Interoperability Framework, and Training Package in 2025 (Aug., Nov., and Nov., respectively) and release final versions of these artefacts in 2026 (Aug., Nov., and Nov., respectively). To ensure alignment with the Drivers and Providers, we will arrange yearly Requirements and Capabilities and Evaluation and Adoption workshops (May and Sept., respectively). The Roadmap itself will be revised in May 2025 and 2026 to reflect project developments and the outcomes from the Requirements and Capabilities workshops.

Throughout the project, we want to continue our interaction with EOSC-ENTRUST’s sister projects SIESTA[4] (Secure Interactive Environments for SensiTive data Analytics) and TITAN[5] (Trusted envIronments for confidenTiAl computiNg and secure data sharing), which are funded under the same HORIZON 1.3 call on Trusted environments for sensitive data management in EOSC[6]. ENTRUST will benefit from collaborating with SIESTA and TITAN and will seek alignment of objectives with these two other projects. Initial discussions began at the EOSC-ENTRUST Kick-Off meeting, where representatives from both SIESTA and TITAN introduced their initiatives and discussions will continue at events such as the annual EOSC Symposium.

7.2 Roadmap themes

The following sections describe the RC1 Workshop outcomes by summarising the main points and discussions in each session (see Section 6.1).

7.2.1 Potential conceptual frameworks for a blueprint

Laura Ward and Chris Cole (UNIVDUN) presented the Five Safes framework for confidential or sensitive data and the corresponding five dimensions of people, projects, settings, data, and outputs, where the first three dimensions cover managerial controls and the last two cover statistical controls. Importantly, the Five Safes dimensions are not absolute

[4] https://eosc-siesta.eu/

[5] https://titan-eosc.eu/

[6] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-infra-2023-eosc-01-06

requirements but rather represent adjustable levels ("safe sliders"). By setting different increasing thresholds on these levels, one can create tiered safety levels matching existing classifications of data into sensitivity tiers (Tiered Data), such as the Alan Turing Institute 5 tiers[7]. The Five Safes dimensions were further highlighted by showing how these map to the practices of the Health Informatics Centre at the University of Dundee. The presentation and meeting chat also discussed extensions to the Five Safes framework to facilitate Safe interoperability between TREs. These dimensions included Safe compute and Safe return, reflecting, respectively, challenges related to cloud computing and challenges related to data import, including linking datasets from different sources and return of results for future use, into a secure environment. It was also noted that the term "Safe return" has specifically been used for referring to safely returning research results into an individual clinical care setting; e.g, safely reversing de-identified data for individual care or targeted follow-up research[8]. Finally, Legal interoperability was mentioned as an additional challenge for building interoperable networks of TRE services across national or regional jurisdictions.

Heikki Lehväslaiho (CSC) presented enterprise architecture (see Section 5.2) as a potential modelling tool in the ENTRUST architecture work. One goal of using enterprise architecture modelling is to illustrate concepts at appropriate levels of abstraction to avoid confusion. This was demonstrated by showing enterprise architecture models for the Finnish data security framework, the CSC sensitive data services principles, an extended Five Safes framework, and the SATRE specification[9].

Whereas both presentations had a view towards health data, it was pointed out during the discussions that trusted research environments are needed beyond the health data space; indeed, ENTRUST's Driver forum covers more use cases than health data research and the Drivers themselves represent a subset of the possible use cases for trusted research environments in the context of EOSC. Consequently, the ENTRUST Blueprint should cover different safety requirements – for example, by matching Tiered Data with appropriately tiered TREs. Nevertheless, it was also pointed out that alignment with the ongoing work of the European Health Data Space (EHDS) is still important, as many TREs – including members of the ENTRUST Provider form – aim to serve as providers for EHDS and will therefore face a strict need to implement EHDS requirements.

7.2.2 Interoperability requirements from Drivers

The following drivers presented interoperability requirements based on their current experiences:

1. Federated Human Genomics (Jordi Rambla de Argila, CRG)

[7] https://doi.org/10.6084/m9.figshare.11815224.v6
[8] https://ukhealthdata.org/wp-content/uploads/2020/04/200430-TRE-Green-Paper-v1.pdf
[9] https://satre-specification.readthedocs.io/en/latest/specification.html

2. SSHOC – Social Sciences & Humanities Open Cloud (Beate Lichtwardt, UKDS)
3. Clinical Research (Sergio Contrino, ECRIN)
4. Health & Environmental Science in Public-Private Partnerships (Anne-Marie Tuikka, Turku UAS)

Across the drivers a wide range of requirements were identified. These requirements could broadly be classified into the following five categories:

1. **Data transfer between environments:**
    - Combining different data sources residing in distinct environments (e.g. different repositories for individual participant data from clinical trials, or linking archived genomic data with phenotype data from registries on the level of individuals).
    - Mandatory data encryption during transit.
    - Efficient data transfer protocols for handling large datasets, such as Aspera[10].
2. **User identity and access management:**
    - Verification of the user's identity and role before giving data access.
    - Implementing categorised data access levels based on predefined standards (e.g. standardised roles and access levels).
3. **Governance and compliance:**
    - Development of governance models tailored to varying organisational and research needs (e.g. legal or policy constraints of private companies vs. public research organisations).
    - Compliance with domestic and international standards (such as ISO27001) of data and information security management practices, and certification thereof.
    - Managing compliance with legal frameworks for data sharing.
    - Allowing diverse TREs with different technical or resource constraints (e.g. standardised environments with pre-installed software vs. custom environments allowing users to install software themselves, including supporting software with dial-home licence models).
4. **Data lifecycle management:**
    - Establish mechanisms for the timely deletion of data post-expiration or when not required.
    - Secure integration and management of different sensitive data types, such as genomic, health record, registry, questionnaire, image, sound, and video data.
5. **User training and certification:**
    - Ensure safe researcher training compliant with data access requirements.

[10] https://www.ibm.com/products/aspera

- Creation and dissemination of general training modules focusing on curating and handling sensitive data across varied domains.
- Standardisation of statistical disclosure control, such as output checker training and certification.

7.2.3 Interoperability capabilities from Providers

The following providers presented their experiences with and capabilities for interoperability:

1. SATRE – Standard architecture for trusted research environment (Simon Li, UNIVDUN)
2. Tryggve (Abdulrahman Azab, UiO)
3. de.NBI (Nils Hoffmann, de.NBI; Fabian Paz, EKUT; Christian Buggedei, BIH)

SATRE is an open, UK-wide, community-led specification on how to build and run a TRE. The specification is organised as four pillars consisting of 29 capabilities with 160 concrete statements. Each statement is scored (0, 1, 2, N/A) and statements are either mandatory, recommended, or optional; TREs should score ≥1 on mandatory statements. This classification of statements and statement score thresholds is a potential basis for mapping TREs to tiered safety levels (see Section 7.2.1). The SATRE framework is a potential starting point to evaluate ENTRUST TREs, as it is a robust reference point for comparison. Moreover, whereas the full evaluation of all 160 statements is extensive, the specification can be tailored by focusing on the higher level capabilities or using a selected subset of statements. SATRE is also well aligned with IS027001. The current specification may not be ideal for all European environments, however, but applying SATRE in a broad European setting is a good test that can identify its applicability and areas for improvement.

Tryggve[11] is a collaborative Nordic (Denmark, Finland, Norway, Sweden; recently, Estonia also joined) project, aiming to develop and facilitate access to secure e-infrastructure supporting large-scale cross-border biomedical research studies based on sensitive data. The project has both supported specific research use cases, such as the Nordic twin study on cancer, and developed secure tools for analysing sensitive data across borders and sensitive data archiving technology. The secure tools cover both joint and federated processing scenarios; the archiving technology is the basis for the Federated European Genome-Phenome Archive (FEGA).

de.NBI[12] is the German network for bioinformatics infrastructure, which focuses on providing tools and services for bioinformatics and cloud computing resources for German academia. A subset of these resources are TREs and two of these, Cloud Tübingen and HEALTH-X, were presented in more detail. Cloud Tübingen is primarily a project-centric

[11] https://neic.no/heilsa/
[12] https://www.denbi.de/

data processor, allowing researchers to upload and analyse sensitive data according to standard operating procedures approved by the data controller. In contrast, HEALTH-X is a data-centric federated infrastructure for health data built on Gaia-X[13]. It aims to be an open data ecosystem for health data, allowing individual citizens to gather and combine their data from primary (general practice) and secondary (hospital) health care providers with other sources of health data, such as personal smart watches. These data can then be consented for secondary (research) use.

7.2.4 Input from aligning projects

The discussions and presentations emphasised the need for coherent strategies between various European projects and initiatives like TEHDAS2, EOSC-ENTRUST, and EHDS, focusing on interoperability, security, and effective data use in research environments.
Helena Lodenius presented TEHDAS2, which, starting in May 2024, focuses on the secondary use of health data, particularly through the development of guidelines for Health Data Access Bodies (HDABs), data holders, and users. It aims to establish technical specifications and security requirements for SPEs and de-identification of data. TEHDAS2 will inform the implementation of the EHDS SPE requirements and provide technical inputs to EOSC-ENTRUST. It is of critical importance to align between TEHDAS2 (legislation-driven) and EOSC-ENTRUST (community-operated) to avoid overlap and ensure complementary development.

Christine Stansberg presented Norway's ongoing alignment with EHDS, which is led by the Norwegian Health Directorate through projects like MyHealth@EU for primary use of health data and HealthData@EU for the secondary use of health data. The collaboration involves Norway’s major universities and focuses on reusing existing solutions and establishing minimum SPE requirements.

Discussions on the role of High-Performance Computing (HPC) in TREs highlighted the need for environments that not only secure data but also provide necessary computational resources.

7.2.5 Input from open discussion on interoperability

The discussion in this session underscores the complexity of creating cohesive TRE/SPEs which can balance technical, legal, and governance aspects while ensuring secure, interoperable, and efficient data handling. These aspects are relevant to the blueprint input both for validating TRE/SPE interoperability and for establishing trust between them.

The following key aspects were identified for achieving TRE/SPE interoperability:

1. Composition of TREs and Data Handling:

[13] https://gaia-x.eu/

- From the discussion it results that TREs are best understood as entities that combine SPEs and TOMs.
- Focus on ensuring secure data transfer, including protocols, encryption, and directional data movement.
- Discussion on user privileges and data access, ensuring appropriate access levels for different users.

2. Data Management and Query Handling:
   - Consideration of adding query management capabilities that allow users to discover and query data securely.
   - Federated computing and data discovery should be integrated while ensuring metadata security.
   - These capabilities are not a direct requirement of a TRE and can be considered as optional.
3. Trust and Certification:
   - Emphasis on establishing trust through clear governance structures, auditing, and monitoring.
   - Checklist for capabilities to be used by legal representatives when evaluating TREs and a framework for TRE federation.
   - Consideration of eID for user identification.
4. Administrative capabilities:
   - There are clear needs for administrative capabilities such as financial clearance, service requests, and evaluation.
   - At the same time the capability to archive processing environments for future use or peer review should be integrated as part of the vision.

It is important to highlight that the interoperability can be established on different levels to promote diversity in the capabilities of a TRE/SPE. The aim should be on shared processes rather than identical systems.

7.2.6 Input from open discussion on governance

In parallel with the interoperability discussion the Governance session sought input on processes that enable sharing both data and knowledge across TREs, harmonising the differences of legal environments and governing the provider forum.

The discussion focused on four main points:

1. Processes and Data Types:
   - Establish processes for accessing, adding, and linking data, as well as for incorporating external software and computation nodes.
   - Differentiate between data stored within TREs and data to be transferred between TREs.
2. Data Discoverability and Access:

  - There is some debate over whether data discoverability should be within the scope of TREs. On the one hand, there are strong arguments for having TREs focused on secure, project-specific data handling. On the other hand, making TRE-stored data discoverable would provide a real benefit.
  - Data discoverability was also identified in the interoperability breakout session, with the mention that it can be considered an optional capability rather than a required one.
3. Country-Specific Differences and Roles:
   - This should include highlighting variations in data handling and discoverability processes across countries as well as emphasise the importance of defining roles within the TRE ecosystem, including data controllers and providers.
4. Legal and Governance Challenges:
   - Address the need for alignment with EHDS and accreditation criteria for TREs and discuss the role of governmental agencies in harmonising data access processes, particularly outside health data.
   - Consider the involvement of universities and national authorities in establishing legal frameworks.
   - Highlight the role of EOSC Association and EHDS in influencing and aligning TRE initiatives.

In terms of future steps and development it was emphasised that there is a need for clear definition of TREs, including subject-specific vs. general TREs, proposing the development of a framework to inform decision-makers and funders on TRE requirements.

7.2.7 Input from open discussion on standards and interfaces

The standards and interfaces breakout session aimed to identify existing standards and interfaces that are working well, or if there is an opportunity for further development, taking into consideration the interoperability of TRE/SPEs. At the same time a connection was identified with Research Object (RO)-Crate[14] as a standard for packaging research data and their meta-data and workflow processing of the data.

One of the key opportunity areas identified is given by the lack of formal vocabulary with respect to TRE/SPEs and cross-border data transfer, where the different aspects regarding access levels come into play. While different ontologies to describe the level of privacy or data use exist, for example DUO[15], ICO-ontology[16] and DPV[17], still the main challenge is defining what is most relevant for a TRE.

[14] https://www.researchobject.org/ro-crate/
[15] https://github.com/EBISPOT/DUO
[16] https://github.com/ICO-ontology/ICO
[17] https://w3c.github.io/dpv/dpv/

In terms of RO-Crates there is a need to establish legal boundaries for the datasets, whilst the workflows face a compatibility issue due to the use of different workflows for different TREs, although standards such Common Workflow Language[18], GA4GH WES[19] and TES[20] can be utilised alongside Apptainer[21] and Podman[22] containerization.

Also worth noting is regulatory work on secure identification and communication, such as Electronic Identification and Trust Services (eIDAS)[23], eDelivery[24], and the NIS2 Directive[25].

8. Discussion

We have identified the main components required for structured secure data processing to be further developed in the Blueprint. These are depicted as extensions and specialisations of the capabilities identified in the European '1+ Million Genomes' Initiative in Figure 8.1. Focus areas will be:

- The concept of SPE as the paradigm for secure **data processing** will need to be expanded to accommodate automated and semi-automated secure and distributed processing.
- **Knowledge dissemination** will capture the export of anonymous results from SPE, return of enriching information to source datasets, and creation of new datasets to enrich scientific knowledge.
- **Transient dataset management** will capture the creation of merged and minimised datasets by data holders or their representatives to create tailor-made datasets for data users.
- All communication within and between secure environments will be governed by **interoperability** rules for encrypted and directional transfer of information.
- The concept of **project** in data access management and data processing will need to be defined consistently.

[18] https://www.commonwl.org/
[19] https://ga4gh.github.io/workflow-execution-service-schemas/
[20] https://github.com/ga4gh/task-execution-schemas
[21] https://apptainer.org/
[22] https://podman.io/
[23] https://eur-lex.europa.eu/eli/reg/2014/910/oj
[24] https://ec.europa.eu/digital-building-blocks/sites/display/DIGITAL/eDelivery
[25] https://eur-lex.europa.eu/eli/dir/2022/2555

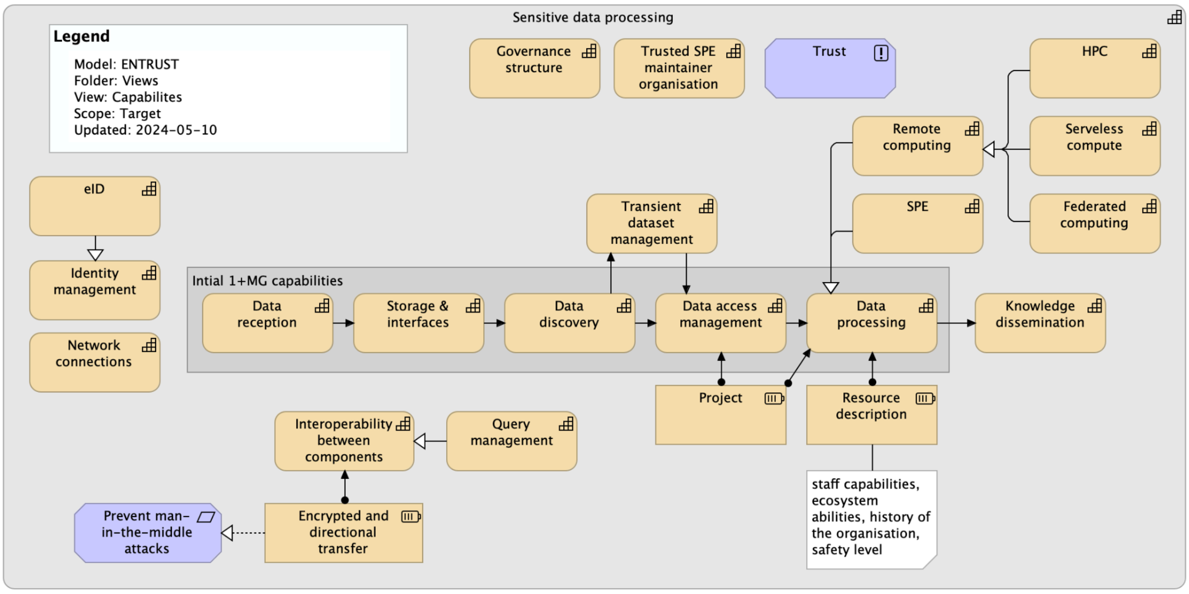

Figure 8.1. Capabilities outline the areas of work.

Based on the requirements uncovered in the Drivers session, the general discussions, and the specific discussion in the interoperability session, the following three areas were identified where new development work is needed: (i) Sensitive data transfer between TREs, (ii) Common user identity and access management, and (iii) Creation of reusable information from secure environments, including reusable datasets, to close the gap in the research data life cycle. These areas represent relevant functionality for interoperability requirement categories 1, 2, and 4 (“Data transfer between environments”, “User identity and access management”, and “Data lifecycle management”, respectively; see Section 7.2.2) and will be taken forward as Validators and sought to be implemented within the ENTRUST consortium of TREs.

During the workshop discussions, mechanisms for establishing trust in users and between TREs were mentioned as important aspects of TRE interoperability. Identity management is therefore crucial; however, defining clear and modular structures for identity management and other secure data processing components that can serve many different governance structures is a major challenge. To illustrate, the upcoming EHDS legislation defines specific structures for health data within Europe, but the ENTRUST project and Blueprint will have to put special emphasis on use cases that will not be covered by upcoming EHDS legislation. This will need to identify data types that are not shared with health data as well as domains of science that have specific security requirements.

The RC1 Workshop has started gathering requirements for TRE interoperability and mapping these to existing TRE capabilities, thereby forming a basis for the first version of the EOSC-ENTRUST Blueprint & Interoperability Framework. While working on this first Blueprint version, we will coordinate with both the Driver and Provider forums, which will continue gathering and refining requirements and capabilities through separate activities in their respective work packages. The first EOSC-ENTRUST Evaluation & Adoption Workshop will be the next important milestone in this coordination work.

9. Conclusions

The RC1 Workshop identified five main categories of requirements for the Blueprint & Interoperability Framework: 1) Data transfer between environments, 2) User identity and access management, 3) Governance and compliance, 4) Data lifecycle management, and 5) User training and certification. The requirements, along with the workshop presentations and discussions, also emphasised the need for diverse TRE solutions, including different technical solutions and safety trade-offs, that can collectively cover the needs of diverse scientific domains and private actors. The blueprint work should therefore primarily consider frameworks that focus on the services and architecture needed for TRE interoperability instead of technologies and technical details. The Five Safes and SATRE are two such general and specific frameworks, but both would require extensions for interoperability (e.g. remote compute, import of data, and return of results.)

The EHDS is currently the driving process for establishing legislation and technical requirements for health data interoperability in Europe. In functional terms, the SPE defined in EHDS together with its TOMs is equivalent to the TRE concept. The ENTRUST Blueprint should align with the EHDS, as health data is an important part of sensitive data, but the Blueprint should also encompass use cases beyond health. Defining categories of TREs supporting e.g. specific tiers of data safety or specific legislations could be one approach for both users and decision-makers to choose between alternative TREs for addressing specific needs.

10. Impact

This deliverable provides the “initial version of a roadmap document presenting the steps needed for the blueprint” (project proposal). The outcomes of the RC1 Workshop were crucial to achieve this. Especially, the deliverable contributed to key results 1 and 5 of the project’s objectives 1 and 2, respectively (see Section 4 “Contribution towards project objectives”.)

In the EOSC-ENTRUST consortium we have several different TRE and SPE environments in different development stages and it will be important to work on the catalogue mentioned in 1) in order to get a comprehensive view on the diverse landscape, but not to “reinvent the wheel”. The discussions in the workshop on the different environments, and the results of the TRE inventory survey conducted by the Provider WP (and presented at the workshop) will feed directly into this. Furthermore, the different breakout sessions in the workshop on the topics of Interoperability, Governance and Standards, and Interfaces provided a valuable forum for 2).

To conclude, the RC1 Workshop was an initial, though very important, step on the longer journey to achieve the project's objectives and drive interoperability between the different secure environments. As the different work packages in EOSC-ENTRUST are closely intertwined, these discussions amongst the participants of the different work packages are very much needed to get aligned. The workshop will be followed by a deeper analysis of the results obtained and the roadmap document will contribute to the more detailed planning of the deliverables and milestones.