October 25, 2024

Trusted Research Environments

Landscape report

+31 (0)20 89 32 007
contact@egi.eu
www.egi.eu

Advanced Computing Services for Research

Landscape report

EGI Working Group on Trusted Research Environments and Sensitive Data Management

Authors	Ville Tenhunen (EGI Foundation), Lucas van der Meer (ODISSEI), Abdulrahman Azab (TSD/UiO), Mario Reale (GÉANT), Andrea Cristofori (EGI Foundation), Michal Růžička (Masaryk University, Brno), Matthijs Moed (SURF), Andrea Manzi (EGI Foundation), Gergely Sipos (EGI Foundation)
Acknowledgements	Matej Antol (Masaryk University, Brno), David Blundell (CyberHive), Robert Jan Bood (SURF), Marian Bubak (SANO/Cyfronet), Mark Dietrich (EGI Foundation), Lars Eklund (NBIS, SND, Uppsala University), Machiel Jansen (SURF), Maciej Malawski (SANO/Cyfronet), Michal Orzechowski (SANO/Cyfronet), Jerome Pansanel (CNRS), João Pina (LIP), Alan Platt (CyberHive), David Rodriguez (CSIC), Miroslav Ruda (Masaryk University, Brno), Roberto Sabatino (HEAnet), Renato Santana (EGI Foundation), Jonas Söderberg (SciLifeLab), Sanjay Srikakulam (University of Freiburg), Viet Tran (IISAS), Paweł Turkowski (Cloudferro), Carmina Vica (SURF), Roksana Wilk (Cyfronet)
Technology	Trusted Research Environments and Sensitive Data Management
Last update	25.10.2024
Status	Final version
License	CC BY 4.0: https://creativecommons.org/licenses/by/4.0/

Document log

Issue	Date	Comment	Author
Disposition	6.2.2024	The first version of the disposition	VT
Disposition	19.2.2024	Added couple of points based on the WG discussion	VT
First draft	19.8.2024	Collected information from meetings and other sources	VT
Second draft	13.9.2024	All chapters in place and texts mostly added	VT, LvdM, AA, MR
Third draft	27.9.2024	Finalised version	VT, LvdM, AC, MR
Final draft	2.10.2024	Finalised introduction	VT
Final version	25.10.2024	Comments from Lecce session included	VT

Terminology

For the purpose of this document, the following terms and definitions apply:
The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", “MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119. For a complete list of term definitions see the EGI Glossary (http://wiki.egi.eu/wiki/Glossary).

Abbreviations

Term	Description
AAI	Authentication and Authorization Infrastructure
AI	Artificial Intelligence
CERT	Computer Emergency Response Team
CMDB	Configuration Management DataBase
EHDS	European Health Data Space
EOSC	European Open Science Cloud
FAIR	Findability, Accessibility, Interoperability, and Reusability
FISMA	The Federal Information Security Modernization Act
GDPR	General Data Protection Regulation
HIPAA	The Health Insurance Portability and Accountability Act
ICT	Information and Communication Technology
ISP	Internet Service Provider
KVM	Kernel-based Virtual Machine
LLM	Large Language Models
NREN	National Research and Education Network
NTNU	The Norwegian University of Science and Technology
RHEV	Red Hat Enterprise Virtualization
RPO	Research Performing Organisation
SANE	Secure ANalysis Environment
SATRE	Standard Architecture for Trusted Research Environments
SP	Service Provider
SRE	Secure Research Environment
SIG-DHD	Special Interest Group - Digital Health Data
TEE	Trusted Execution Environments

TRE	Trusted Research Environment
TSD	Tjenester for Sensitive Data (Service for Sensitive Data)
VRE	Virtual Research Environment
UiB	University of Bergen
UiO	University of Oslo
WG	Working Group

Table of Contents

Executive Summary

The landscape of Trusted Research Environments (TREs) in Europe is extensive, with this report identifying 40 TREs across various themes and disciplines. These TREs are produced by National Research and Education Networks (NRENs), Research Performing Organizations (RPOs), and specific projects. It is evident that not all existing TREs have been included in this list. Members of the EGI Federation are also involved in TRE activities. Organisations such as CNRS, CSIC, INFN, IISAS, and SURF participate in projects or operate advanced services in this field.

The terminology surrounding TREs is still evolving, with different terms such as "Secure Processing Environment" and "Secure Research Environment" used interchangeably.

TREs share some key characteristics: computing and data management within these environments are highly secure and controlled; only approved users, such as researchers, can access, store, and analyse sensitive data remotely; data remains within a secure server environment and does not leave it; and user management is based on trusted Authentication and Authorization Infrastructure (AAI), which may include accreditation processes.

The technological foundations of TREs are well understood, with no major open issues or research questions in this area. However, continuous development requires new technologies. For instance, Trusted Execution Environments (TEEs) represent an emerging technology that enhances security at the processor level.

As the TRE landscape continues to mature, new players bring fresh technological, policy, and business model approaches. While sensitive data necessitates strict security and privacy measures, there are opportunities to improve usability. User experience is crucial for overall security, but traditionally, it has not been a primary focus in TRE development due to the emphasis on security. In the future, usability will play an increasingly important role alongside risk management.

Operating a TRE is both a technical and configurational challenge, heavily influenced by various regulations, standards, and options. Consequently, a purely technology-focused approach is insufficient, and there is no comprehensive "deployable TRE package" available.

The development of TREs also requires methodological and governance standardisation. Organisations providing standardised and federated TRE services with high security and privacy levels must obtain certifications, such as ISO/IEC 27001.

In TREs, AAI solutions are critical, especially when creating federated TREs, which require federated AAI. It is important to note that AAI solutions must meet stringent security requirements, such as Multi-Factor Authentication (MFA). Federations enable cross-border interoperability, provided organisational and legal frameworks allow it. Although semantic interoperability presents a challenge for federated TREs, this issue is common to all TRE approaches.

Artificial Intelligence (AI) will also play a role in future TRE development. For example, generative AI methods applied to sensitive data can produce synthetic data for research purposes, which can be shared without the restrictions associated with sensitive data. Producing synthetic data requires a deep understanding of AI and machine learning algorithms, especially in handling sensitive information.

Initiatives and regulations in the health sector, such as the European Health Data Space (EHDS), are key drivers in the evolution of TREs. Other major drivers include recently launched HORIZON-INFRA-2023-EOSC-01-06 projects like EOSC-ENTRUST, SIESTA, and TITAN. These projects are particularly significant for EGI, complemented by other potential European Open Science Cloud (EOSC) initiatives.

Specifications and guidelines for TRE components are already mature and stable. The SATRE specification, for instance, serves as a comprehensive enterprise architecture model for TREs, while many national requirements support the secondary use of social and health data.

Numerous national or institutional TREs exist across Europe, often integrated or federated. The next challenge involves cross-border interoperability and federated operational models.

To address current challenges, we need to simplify TRE setup, automate processes, enhance user experience, develop training programs, and create guidelines. Increasing the number of TREs will help improve their overall quality.

Interconnectivity among TREs must also advance to the next level to support research, innovation, and adherence to FAIR (Findable, Accessible, Interoperable, Reusable) principles. TREs should be accepted across diverse environments, which requires an understanding of different data sensitivity levels and technological solutions. Certifications are essential for harmonising various solutions; if standard modes of interoperability and interconnectivity can be agreed upon, multiple practical solutions can be more easily adopted in production.

Introduction

The European data strategy defines principles to develop data related services and infrastructures. The strategy aims to make the EU a leader in a data-driven society. Creating a single market for data will allow it to flow freely within the EU and across sectors for the benefit of businesses, researchers and public administrations.[1] Data Act emphasising fair access and user rights, while ensuring the protection of personal data.[2] Data Governance Act seeks to increase trust in data sharing, strengthen mechanisms to increase data availability and overcome technical obstacles to the reuse of data for example in the health data.[3]

Sensitive data management and Trusted Research Environments (TREs) are critical components in the realm of research infrastructures, particularly when dealing with confidential or personal data or information. Effective sensitive data management ensures that data is handled, stored, and shared in ways that protect its integrity, confidentiality, and availability, while complying with legal and ethical standards.

TREs are secure digital platforms designed to facilitate research while safeguarding sensitive data. They allow researchers to access and analyse data without compromising its confidentiality, ensuring that only authorised users can interact with the data and that data remains within the secure environment. This setup is crucial for maintaining trust between data providers, researchers, and the public, and for enabling high-quality research that can inform policy and practice without risking data breaches or misuse.

In research infrastructures, the integration of TREs and robust data management practices is essential. These tools enable large-scale, collaborative research projects while ensuring compliance with data protection regulations and ethical guidelines. By fostering a secure environment for data use, they play a key role in advancing research in fields such as healthcare, social sciences, and genomics, where sensitive data is often central to the research objectives. Even if these are traditionally mentioned domains of the sensitive data, there are multiple other domains and disciplines that use controlled or sensitive data. TREs are often thematic infrastructures but there are a lot of generic implementations as well.

In this report we describe some definitions for TREs, present some examples of real life implementations, list some requirements for the TRE, discuss interoperability schemes and emerging technologies. Finally, the document lists some active initiatives and projects and discusses integration scenarios in the EGI infrastructure.

---

1 https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en#a-single-market-for-data

2 https://digital-strategy.ec.europa.eu/en/policies/data-act

3 https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

This document does not belong to any project and does not try to define all aspects of TREs, main focus is in practical steps to build or integrate TRE.

The report was prepared by the EGI Working Group on Trusted Research Environments and Sensitive Data Management. The WG started 31.10.2023 and ended its work in the EGI2024 conference session 3.10.2024 in Lecce, Italy. The WG had 34 members from 17 organisations.

Definitions

Trusted Research Environments or Secure Research Environments have not one commonly recognized definition but several more or less similar ones. In this chapter we present some central terms and definitions. Additionally, conclusions of terms used in this deliverable have been presented.

General terms and characteristics

Data space
A distributed system defined by a governance framework that enables secure and trustworthy data transactions between participants while supporting trust and data sovereignty. A data space is implemented by one or more infrastructures and enables one or more use cases.[4][5]

Data visiting
Data visiting is an approach where data stays at the owner and allows the consumers (e.g. analysts or machine learning algorithms) to come to the data to work with it.[6]

FAIR data
FAIR data is data which meets the FAIR principles of Findability, Accessibility, Interoperability, and Reusability. Highly sensitive or personally-identifiable data can also be FAIR-data. It means the publication of metadata to facilitate discovery, including clear rules and information regarding the process for accessing the data even if the data itself is not public.[7]

Five Safes
The Five Safes is a framework for helping make decisions about effective use of data which is confidential or sensitive[8]

- Safe projects: is this use of the data appropriate?
- Safe people: can the users be trusted to use it in an appropriate manner?
- Safe settings: does the access facility limit unauthorised use?
- Safe data: is there a disclosure risk in the data itself?
- Safe outputs: are the statistical results non-disclosive?

[4] https://dssc.eu/space/BVE/357073747/2+Core+Concepts
[5] The European Data Strategy:: https://digital-strategy.ec.europa.eu/en/policies/data-spaces
[6] https://datascience.codata.org/articles/10.5334/dsj-2022-004#1-introduction
[7] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
[8] https://en.wikipedia.org/wiki/Five_safes

Health data hub
Minimal inclusion criteria:

- A digital technical infrastructure with the core mission of enabling health data sharing
- It provides health data from different sources
- It allows discovery of health datasets
- It has a metadata discovery service
- It has a data accessibility mechanism in accordance with existing regulation
- It has an authorization functionality, provided by the same Data Hub or by an external institution.[9]

Personal data and sensitive personal data
Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data.[10]

Sensitive personal data include information related to racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership and data concerning the health or sex life of an individual. These data could be identifiable and potentially cause harm through their disclosure.[11]

The following personal data is considered 'sensitive' and is subject to specific processing conditions[12]:

- personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
- trade-union membership;
- genetic data, biometric data processed solely to identify a human being;
- health-related data;
- data concerning a person's sex life or sexual orientation.

Secure Processing Environment
The physical or virtual environment and organisational means to ensure compliance with Union law, such as Regulation (EU) 2016/679, in particular with regard to data subjects' rights, intellectual property rights, and commercial and statistical confidentiality, integrity and accessibility, as well as with applicable national law, and to allow the entity providing the secure processing environment to determine and supervise all data processing actions, including the display, storage, download and export of data and the calculation of derivative data through computational algorithms.[13]

[9] HealthyCloud Glossary, https://doi.org/10.5281/zenodo.6787119
[10] https://commission.europa.eu/law/law-topic/data-protection/reform/what-personal-data_en#:~:text=Personal%20data%20is%20any%20information,person%2C%20also%20constitute%20personal%20data.
[11] HealthyCloud glossary: https://doi.org/10.5281/zenodo.6787119
[12] https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en
[13] European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R0868

Secure Research Environment
The Secure Research Environment (SRE) is a service that allows researchers to store and analyse sensitive research data subject to regulatory standards including for example HIPAA and FISMA in the USA.[14]

Sensitive data
Information that is regulated by law due to possible risk for plants, animals, individuals and/or communities and for public and private organisations. Data may be sensitive because of privacy, commercial or copyright considerations.

Virtual Research Environment
Virtual Research Environment (VRE): “A virtual research environment (VRE) or virtual laboratory is an online system helping researchers collaborate. Features usually include collaboration support (Web forums and wikis), document hosting, and some discipline-specific tools, such as data analysis, visualisation, or simulation management.”[15]. Secure processing environments might be considered a subset of virtual research environments.

Trusted Research Environment

Term “Trusted Research Environment” has a few slightly different definitions. Here are some of them.

- Trusted Research Environments (TREs) are highly secure and controlled computing environments that allow approved researchers from authorised organisations a safe way to access, store, and analyse sensitive data remotely.[16]

- Trusted Research Environments (TREs) (also known as secure data environments) are highly secure computing environments containing de-identified data. Many operate under the principles of the Five Safes framework. Researchers and their projects must go through the TRE's rigorous accreditation process to access and use this data.[17]

- Trusted Research Environments (TREs) take the form of a secure data environment that allows analysts and researchers to undertake in-depth analysis on rich, joined-up datasets

[14] https://case.edu/utech/departments/research-computing-and-infrastructure-services/services/secure-research-environment
[15] https://en.wikipedia.org/wiki/Virtual_research_environment
[16] Hadley Sheppard, https://www.lifebit.ai/blog/what-is-trusted-research-environment
[17] ADR UK, https://www.adruk.org/data-access/trusted-research-environments/

without them seeing any identifiable information. Data is held within a secure server and does not leave that server.[18]

- Trusted Research Environments exist across the UK. In Scotland, they are often referred to as data safe havens, while in the rest of the UK they are sometimes called secure data environments or secure research environments.[19]

Five Safes, Governance and Technical aspects

To connect Five Safes explained above with the TRE concept it is needed to connect those principles with TRE governance and technical aspects as follows.

Safe	Example	Governance	Technical aspects
Safe projects	Is this use of the data appropriate? Is the data minimised?	Data is treated to protect any confidentiality concerns; GDPR, DMP, ISM, SMS	AAI (access), monitoring, reporting
Safe people	Can the users be trusted to use it in an appropriate manner?	Research projects are approved by data owners; ISM, accreditation, agreements	Virtual organisations, AAI (authorization), logging
Safe settings	Does the access facility limit unauthorised use? Are actions logged? Can the user see the data? Is the system encrypted?	Researchers are trained and authorised to use data safely; DMP, metadata, agreements, training. Logging	Data in use (processing environments), data in transit (networking), data at rest (storages, back-up), encryption, catalogues, AAI, physical security of environments
Safe data	Is there a disclosure risk in the data itself?	A processing environment prevents unauthorised use; ISM, data de-identification i.e.	Data in use (processing environments), data at rest (storages), encryption, de-identification, data disposal, AAI

[18] NHS, UK, https://transform.england.nhs.uk/ai-lab/explore-all-resources/develop-ai/working-with-a-trusted-research-environment/

[19] https://www.researchdata.scot/our-work/data-explainers/what-are-trusted-research-environments/

		(pseudo)anonymization	
Safe outputs	Are the results non-sensitive?	Approved outputs that are non-disclosive; reviews, metadata, PIDs, (pseudo)anonymization, agreements, FAIR principles	Processing environments, repositories, catalogues

Terminology conclusions

In this document, the term “Trusted Research Environment” is used in the context of the previous definitions. Main characteristics are

- Computing and data management environments are highly secure and controlled
- Environments allow approved users such as researchers access store, and analyse sensitive data remotely
- Data is held within a secure server environment and does not leave that environment
- User management (including separate accreditation process in some cases) based on trusted AAI

Demand, use cases and fields of applicability

Demand and fields of applicability

Sensitive data is increasingly used for research purposes which means also that Trusted Research Environments are needed where the data is sensitive but used for research purposes of innovation development. There is also a need for data management, computing, analysis, visualisation, production of data products and data archiving and long term retention.

The same operations should be done with the sensitive data than is possible with non-sensitive data.

Fields where we can see this kind of sensitive data could be in every domain and discipline. Naturally, for example disciplines that use health information are more widely represented in TRE than some others. Social sciences and other humanities are another focus area.

Key role in selection of data for TREs is risk analysis of the data. This should be implemented with all research data management cases.

For example, the SIESTA project has stated that data sensitivity definition is not static; it depends on the context and dynamically defined via policy tooling. Different data sensitiveness is also coupled with different access models.

In the SIESTA project concept[20] following tiered model for data sensitiveness has been presented:

0. Fully open data. No need to use a trusted research environment.
1. Very low risk. Pseudonymised data with very low linking risk. Unlikely to cause harm.
2. Low risk. Strongly pseudonymised datasets with some indirect identifiers.
3. Average risk. Pseudonymised personal data and confidential organisations information.
4. High risk. Weak or no de-identification and very sensitive commercial data.
5. Very high risk. Very sensitive personal data or highly confidential government or commercial data.

Even if security and privacy measures are very important for TREs there is also increasing demand to implement FAIR principles[21] on sensitive data. This has been notified for example in the health data sector.[22]

Use cases and examples

In this section we present some examples of TREs.

SANE

SANE (Secure ANalysis Environment)[23] is a data provider-agnostic, off-the-shelf Trusted Research Environment hosted by SURF (the Netherlands). It relies on SURF Research Access Management for AAI, SURF Research Drive for data sharing and SURF Research Drive for (private) cloud deployment.

SANE is GDPR-compliant, pentested, ISO 27001-certified and designed to provide a secure and controlled environment for researchers to analyse sensitive data. The secure environment includes pre-approved analysis software, such as RStudio and Jupyter Notebooks, providing a secure gateway for researchers to access sensitive data.

SANE comes in two variants – Tinker SANE and Blind SANE – providing two ways of working with and analysing sensitive data. In Tinker SANE, the researcher is able to get access to the sensitive data as well as his own dataset via a Windows virtual machine and can manipulate the data. In contrast, in Blind SANE the data provider executes the analysis provided by the researcher and the researcher cannot see the data. In both variants, the data provider verifies the output before the results can be exported to the researchers' computer.

Setting up SANE typically takes about 30 minutes and involves collaboration between the data provider and a researcher.

[20] EOSC-SIESTA presentation at EOSC Symposium by Álvaro López García on 21.10.2024
[21] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
[22] https://open-research-europe.ec.europa.eu/articles/3-87/v1
[23] http://www.odissei-data.nl/sane

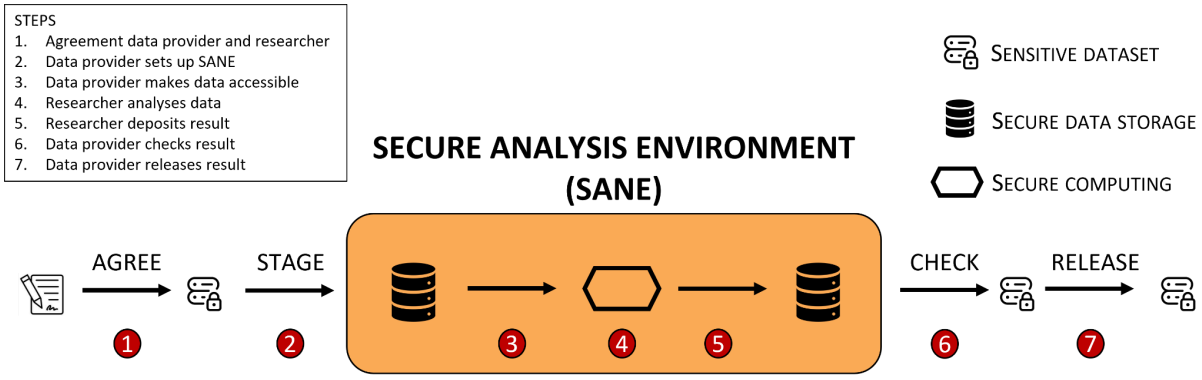

Fig 1. The Secure Analysys Environment (SANE)

TSD@UiO

The TSD – Service for Sensitive Data[24], is a platform for collecting, storing, analysing and sharing sensitive data in compliance with the Norwegian privacy regulation. TSD is primarily used by researchers working at public research institutions (such as universities, hospitals). The TSD is primarily an IT-platform for research even if in some cases it is used for clinical research and commercial research.

TSD is developed and operated by UiO. The main bulk of the HPC resources and some of the storage resources are owned by Sigma2 and are a part of the national e-infrastructure.

TSD is part of the Norwegian Trusted Research Environment Collaboration (Nor-TRE), which involves TRE services at UiO[25], UiB[26]. and NTNU[27]. The suite of Nor-TRE services are potential candidates to join the national Norwegian EOSC Node[28].

Architecture of the TSD based on the idea where all projects/user groups are hence issued with their own dedicated virtual network interconnecting any number of dedicated project servers (Windows and/or Linux).

The solution is run on dedicated computers in a separate location in the organisation's data centre where only operational personnel have access. To achieve complete separation of project environments running on the same hardware, TSD uses Red Hat Enterprise Virtualization (RHEV) Kernel-based Virtual Machine (KVM) as a hypervisor. This means that a physical computer can be divided into several separate virtual computers which for all intents and purposes are working independently.

[24] https://www.uio.no/english/services/it/research/sensitive-data/
[25] University of Oslo https://www.uio.no/
[26] University of Bergen https://www.uib.no/
[27] Norwegian University of Science and Technology https://www.ntnu.edu/
[28] https://open-science-cloud.ec.europa.eu/about/eosc-eu-node

EGI FedCloud as a sensitive data management platform

The LETHE[29] project is about personalised prediction and intervention models for early detection and reduction of risk factors causing dementia, based on AI and distributed Machine Learning (ML).

The project will establish novel digital biomarkers, for early detection of risk factors, based on unobtrusive ICT-based passive and active monitoring. Expansion of digital-enabled health preventive approaches, by reaching out to large populations, can save healthcare systems costs on expensive traditional interventions and confer benefits for the wider society.

In the LETHE project the EGI FedCloud[30] and other EOSC services are used as an infrastructure solution. It is possible to use the EGI FedCloud as a comprehensive sensitive data management solution where on the one hand it is possible to offer data processing and computing services and on the other hand also offer the platform own applications of the organisation or the project.

The project has had two phases. In the first phase the retrospective database has been created. This phase will be used to generate the initial prediction model, based on 4 different data sets provided by the clinical partners of the LETHE consortium. All kinds of connections are encrypted, data stored in the environment is in encrypted volumes and access to the environment is strictly controlled.

Second phase, so called prospective data has been collected via variable apps (such as smartphones, glasses and smartwatches), tools and wearables. The data has been delivered to the models created based on the retrospective data and the data has been processed for clinicians to make conclusions and perform analyses. In this case the data in the EGI FedCloud is anonymised before it arrives to the platform.

Both of these phases have been deployed on the top of the EGI FedCloud with strict security measures and the EGI Check-In[31] has been used as a Authentication and Authorization Infrastructure (AAI) solution within the infrastructure.

[29] https://www.lethe-project.eu/
[30] https://www.egi.eu/egi-infrastructure/
[31] https://www.egi.eu/service/check-in

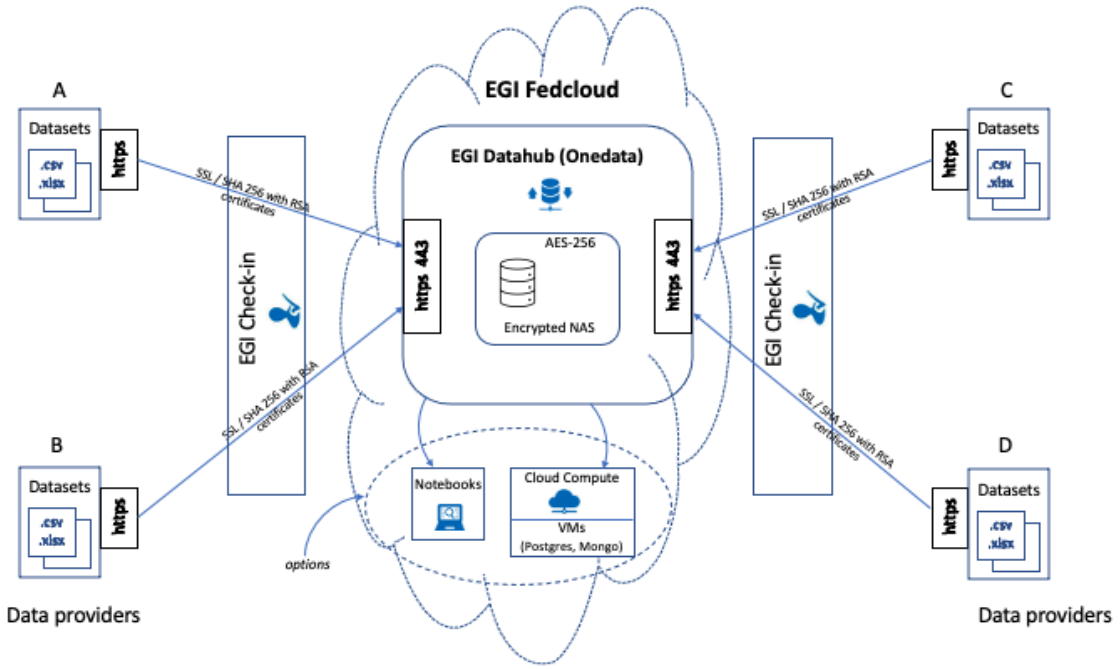

Fig 2. The EGI FedCloud in the LETHE project phase I

Requirements for the TRE

Introduction to requirements

Requirements on the Trusted Research Environment are possible to define based on several sources and approaches. One notable source of the definitions is The SATRE[32] (Standard Architecture for Trusted Research Environments) specification[33].

It outlines a framework for managing secure, compliant, and effective research environments. It emphasises three core pillars: Information Governance, Computing Technology, and Data Management, with supporting capabilities. The specification is designed to ensure that TREs maintain high standards in usability, public trust, observability, and standardisation. Key technical requirements include mandatory components for data protection, user access control, and system observability, with roles clearly defined for governance, data management, and infrastructure. SATRE specification is rather comprehensive enterprise architecture for the TRE.

[32] https://satre-specification.readthedocs.io/en/stable/
[33] https://satre-specification.readthedocs.io/en/stable/specification.html

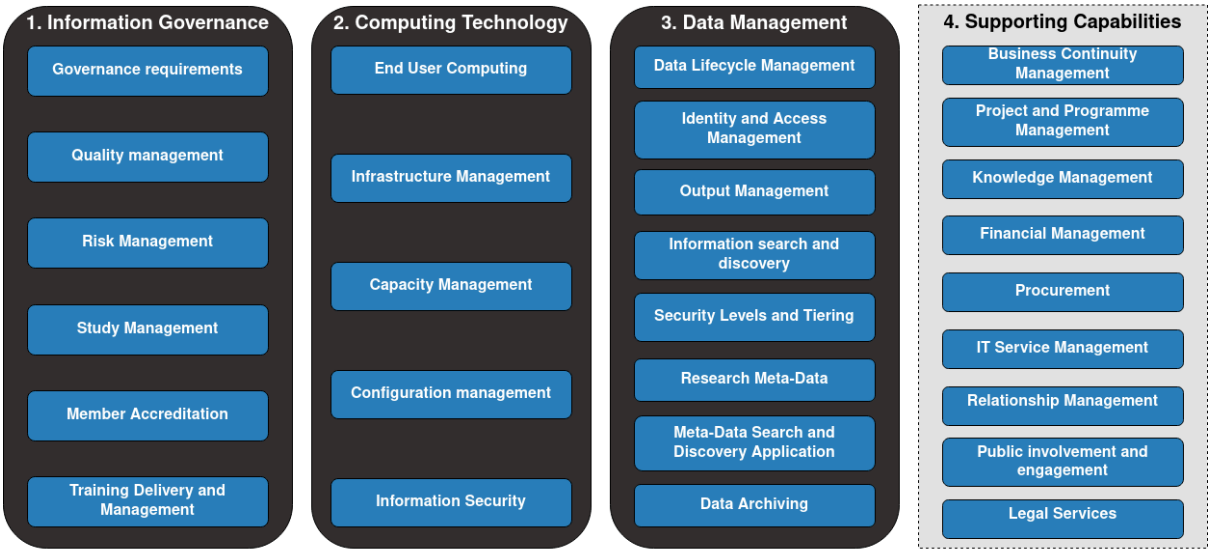

Fig. 3 SATRE Pillars Capability Map

Because the nature of the requirements depend on risk analysis of the data managed within the TRE, there are also other specifications with different granularity of the requirements.

One example of these sets of specifications is Findata regulation on secure operating environments[34]. These regulations are useful when we define TRE technical functionalities and requirements even if interoperability makes compliance with all requirements challenging.

It is important to remember that lowering security requirements may affect the data which it is allowed to handle in such an environment. All these considerations have to be based on risk analysis and following measures.

In this document we present one example set of requirements. In the real life implementations, the final requirements have to be evaluated and modified case by case.

Components of the TRE

In practice, projects and implementations make detailed specifications based on the real life use cases and national and European level regulations and guidance.

The following components are essential when technical requirements are described from the TRE

- Restricted and controlled access to the environment and to all components
- Secure infrastructure solutions (hardened servers, networking, firewalls)
- Permission management

[34] FINDATA - Regulation on secure operating environments;
https://findata.fi/en/kapseli/regulation-on-secure-operating-environments/

- Activity monitoring
- Activity and operation logging
- Data back-ups
- Data management services
- De-identified data, decryption and encryption
- Data, software and AI model sharing
- Data long term preservation
- Metadata services
- Computing services
- Analysing services and SW
- Visualisation services

All these fields are a large part of the virtual research environment and require their own specifications in the real life use cases. In this document, we have to content ourselves with descriptions on a more general level. It should be also noted that in practice local and national conditions and European requirements (such as GDPR) must be taken into account.

In this document, the focus is on technical requirements. Other fields of the TRE security and requirements - such as operational requirements, organisational requirements, legal requirements - are very essential, but not described in this document.

Functional requirements

Basic functional requirements of the TRE are:

- Users log in to the TRE using reliable authentication sources and identity credentials;
- Multi-factor authentication is typically required when the user log in to the TRE;
- Enter or remove the data from the TRE is controlled operation i.e. it has to be logged and if defined, based on predefined permissions by authorities;
- User only has access to materials specified in the data permit in question;
- Data transfer between TRE has been strictly controlled, i.e. possible only based on predefined permits;
- Transfer of user's own data sets to a TRE takes place via security processes and controls;
- Direct internet connections are not permitted in the TRE;
- Log management and backups must take place in an environment with as strict security measures as TRE itself.

The following figure presents one possible generic architecture on the Trusted Research Environment.

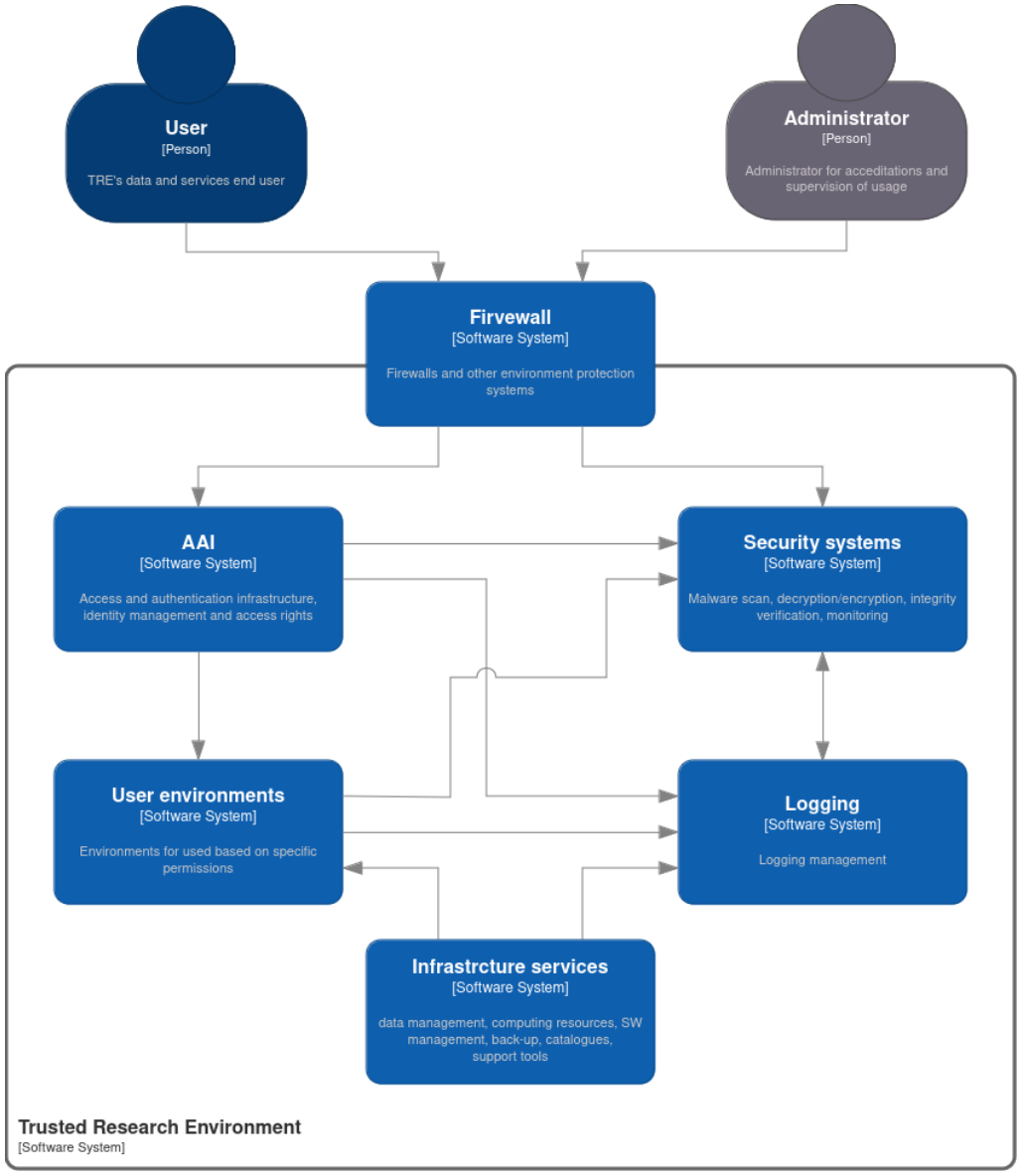

Fig. 4 Generic TRE Architecture

It contains the following components:

- Firewalls: Firewalls and other environment protection systems; all network traffic is controlled;
- AAI: Access and Authentication Infrastructure, identity management and access rights;
- User environments: Work environments for users based on specific permissions. Multiple environments could be accessible by a single user s. Data transferring between these environments are blocked or controlled by permits;
- Security systems: Malware scan, decryption/encryption, integrity verification, monitoring;
- Logging: Logging management;
- Infrastructure services: Data management, computing resources, SW management, back-up, catalogues;

- support tools. Security measures of the TRE components have to be controlled with the same level as user environments etc.

Technical requirements

The technical requirements in this document are based on the presentations and materials in the EGI's WG on Trusted Research Environments and Sensitive Data Management.

User management and access rights

- Users are initially identified primarily through strong electronic identification.
- Multi-factor authentication should be used for user authentication.
- Access rights to the environment are restricted to ensure users can only access the materials and resources for which they have been granted permission.
- Access rights are assigned based on data permissions. If a user holds multiple data permissions within the operating environment, they may access data sets from several research permits. However, the transfer of data sets between user environments is strictly controlled by these permissions (if applicable).
- Access rights are granted based on the principle of least privilege.
- Only tokens from trusted identity providers should be used within the environment.
- Materials within the Trusted Research Environment (TRE) will be deleted after a predefined period once the access rights have expired, unless national law or other regulations specify otherwise.

Securing the infrastructure

- Data has to be encrypted in transit and at rest, in use the data could be decrypted if environment security is high enough.
- The TRE environment has been separated from other respective environments
- The connection of the TRE to the one(s) of another classification level requires the use of a firewall in minimum.
- Data traffic exceeding the perimeter of a controlled physical security area have to be encrypted using an approved encryption solution
- The TRE environment has been separated from other respective environments.
- If a user's workstation or other terminal is not located within the same physically and technically protected area as the user environment, multi-factor authentication must be used for log-in.
- Additionally, the user's working environments must be protected, and the user's ability to import or export data from the TRE environment must be controlled.
- Users are not granted administrative rights to computers in the TRE operating environment.
- User rights shall be managed based on the principle of least privilege, meaning users should only have the rights necessary to perform their duties.

- The management of encryption keys used to encrypt traffic must be organised and controlled. Key management offered by a cloud service provider can be used if the confidentiality of secret keys is ensured at an adequate level.
- In protecting the system, the principles of minimality, least privilege, and defence-in-depth should be applied where applicable.
- Regular updates of malware identifiers can be arranged while strictly limiting the necessary traffic, for example, by using firewall rules.
- To manage software vulnerabilities, information updates from the Computer Emergency Response Team (CERT) community and suppliers are recommended.
- The network, its services, and the servers and workstations connected to it should undergo regular inspections (e.g., vulnerability scans).
- An inventory of software and hardware must be maintained (e.g., CMDB).
- Handling detected vulnerabilities should follow a predefined procedure.
- Regular security scans must be performed on the operating environment.
- Archived data within the TRE should be read only.
- Long-term archives must be held in simple, standard formats to ensure accessibility.

Logging, monitoring and admin rights

- The usage of the TRE has to be carefully logged.
- Logs must be monitored and analysed systematically and regularly.
- Data in logs must be processed with the same level of security as special category personal data.
- Technical log data must be collected comprehensively to ensure that any errors or data breaches can be thoroughly investigated.
- The operating environment is documented.
- The operating environment must be automatically monitored, with clear instructions and assigned responsibilities for responding to incidents.
- When monitoring the user environment, special attention must be given to monitoring information security.
- Maintenance of the operating environment must be conducted from appropriate premises.
- The servers of the operating environment must be located in secure premises that meet the regulatory requirements for such facilities.
- Admin user rights for the operating environment must be personal and assigned specifically based on duties.

Implementation and interoperability architectures

Trusted research environments have had and will have interoperability architectures. This is a trivial requirement because data, even sensitive data, is valuable when it is used. There are few basic concepts for TREs within research communities and in the federations. The idea is to describe different operational modes of TREs: institutional or self hosted, private cloud and federation. These models are generalisations of use cases where there are plenty of local variations in the technical details based on local policies and project goals.

Institutional TRE

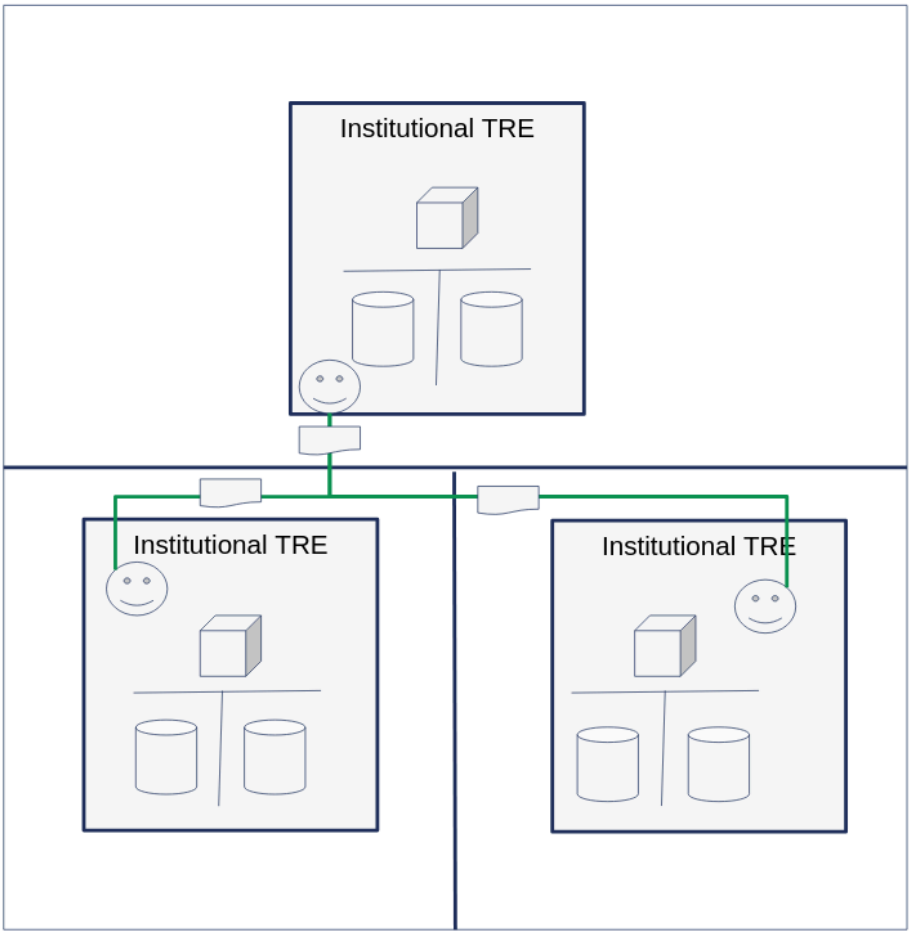

Fig. 5 Institutional operation mode

Traditional architecture where researchers process and manage data in their own institutional environments. Only results or snapshots of the data are shared manually via users.

Features:

- Quite typical institutional solution
- Researchers share only results or preprocessed and (pseudo)anonymised data
- Different policies and principles in institutions
- Different technical solutions
- No common integrations
- FAIR data based on manual work
- Local security and privacy measurements

Private cloud TRE

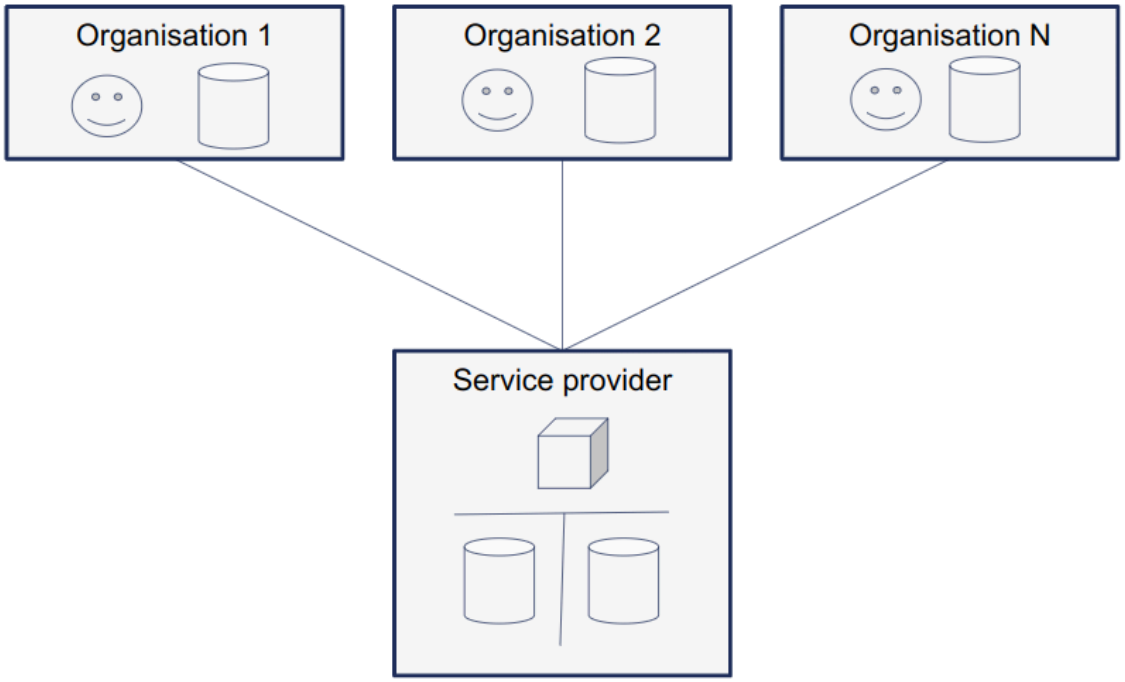

Fig. 6 Private cloud TRE

The second concept of the TRE is the ISP model or cloud model where services and resources are delivered by Service Provider (SP) for all users. SP have their own policies and providers have their own.

Features:

- Typical for projects (such as LETHE).
- Secure measurements vary by organisation.
- Centralised technical solutions and procedures are needed.
- FAIR data based on common agreements of data owners.

Federated TRE

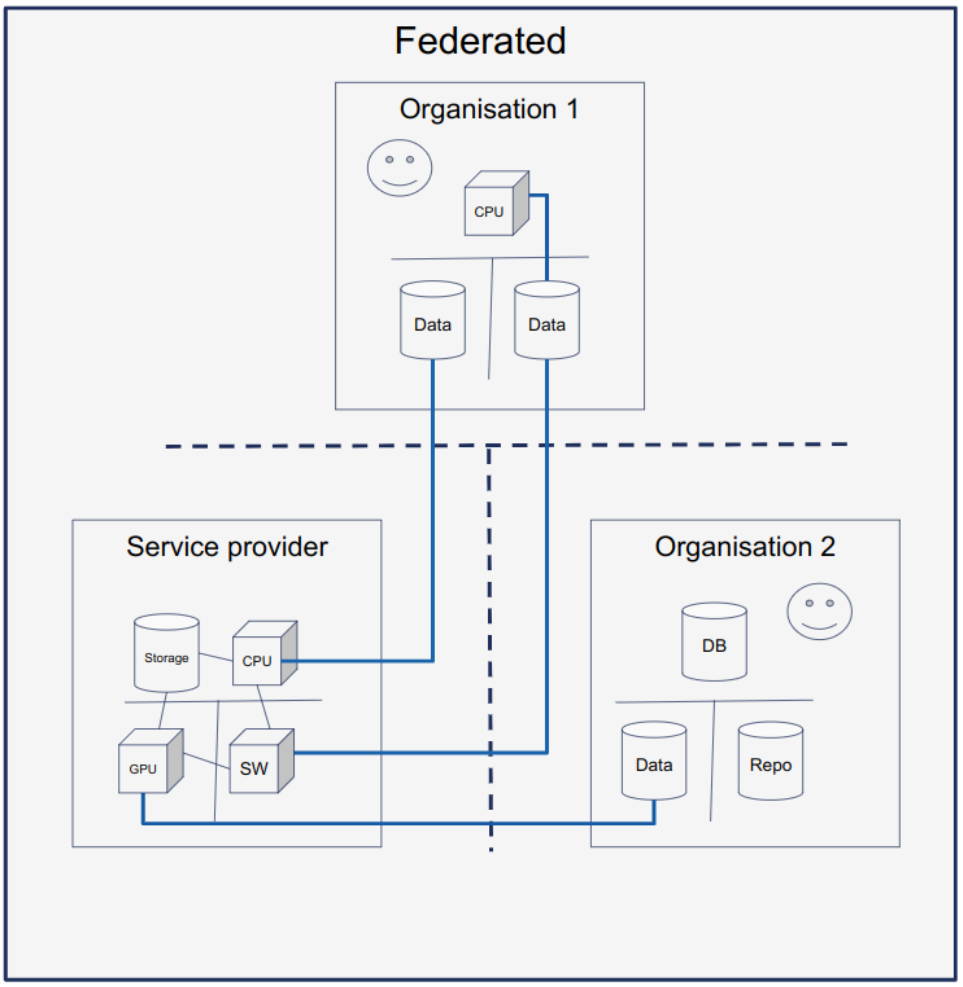

Fig. 7. Federated TRE

The federated model of the TRE contains several possible service providers and users. Access and authentication management based on the federated AAI solution. This conceptual model makes it possible to build services across national or organisational borders. In this model it is usual that data located in institutional or national (NRENs) repositories and computing resources are delivered by other service providers.

Features:

- Common principles, rules and standards
- Shared services, resources and usage across borders
- Common security and privacy arrangements and implementation
- Common API definitions and protocols to move data
- FAIR data based on common agreements and measures

Emerging technologies

In this chapter we describe some interesting technologies which are not yet in wide use in communities. These technologies or concepts have some potential to be a remarkable part of the TRE development in the future. Naturally, it is obvious that there are or will be plenty of other emerging technologies.

Trusted Execution Environments (TEE)

Trusted Execution Environment (Enclaves) is a secure area within a main processor that runs an isolated environment parallel to the main operating system. Through this hardware-level isolation, the TEE guarantees that data and code uploaded into it cannot be tampered with by malicious agents.[35]

Constellation

Constellation[36] is a Kubernetes environment which enables users to migrate sensitive data workloads to the cloud. Constellation is designed to keep all data always encrypted and to prevent any access from the underlying (cloud) infrastructure. This includes access from datacenter employees, privileged cloud admins, and attackers coming through the infrastructure. Such attackers could be malicious co-tenants escalating their privileges or hackers who managed to compromise a cloud server.

Homomorphic computing

Homomorphic encryption is a form of encryption that allows computations to be performed on encrypted data without first having to decrypt it. The resulting computations are left in an encrypted form which, when decrypted, result in an output that is identical to that produced had the operations been performed on the unencrypted data. Homomorphic encryption can be used for privacy-preserving outsourced storage and computation. This allows data to be encrypted and outsourced to commercial cloud environments for processing, all while encrypted.[37]

Synthetic data usage

Synthetic data is artificially generated data using algorithms. Idea is to present similar than real life phenomenons using data produced by AI/ Machine Learning (ML) models or other mathematical models. If there is personal data, synthetic data hides all the connections between data and real

[35] https://confidentialcomputing.io/wp-content/uploads/sites/10/2023/03/Everest_Group_-_Confidential_Computing_-_The_Next_Frontier_in_Data_Security_-_2021-10-19.pdf

[36] https://www.edgeless.systems/products/constellation

[37] https://en.wikipedia.org/wiki/Homomorphic_encryption

world people. If there is fully synthetic data it is possible to share it as public data. Synthetic data methods are also used for fraud detection and confidentiality systems, ML and AI development and other scientific purposes[38].

Other initiatives and projects

National projects

There have been numerous national and international TRE initiatives and projects. Some of them are thematic, some institutional and some national level solutions.

Van der Meer et al.[39] gathered a list of Trusted Research Environments that are used often in the social sciences:

Existing infrastructures services	Geographic	Infrastructure services
anDREa	Netherlands	TRE
Austrian Micro Data Center	Austria	TRE
Bianca	Sweden	TRE
BioMedIT network	Switzerland	TRE
CASD	France	TRE
DataSHIELD	International	TRE component
de.NBI Cloud	Germany	TRE
EHDEN	Europe	Community, TRE
EJP RD Virtual Platform	Europe	TRE
EMBL-EBI Embassy Cloud	United Kingdom	TRE
EPIC Cloud	Italy	TRE
ePouta	Finland	TRE
Federated EGA	Europe	TRE
FINDATA	Finland	Permits
GAIA-X DataLoft	Europe	TRE
Galaxy / UseGalaxy.eu	Germany	TRE, User portal
HONEUR	Europe	Community, TRE
HUNT Cloud	Norway	TRE
INFN EPIC Cloud	Italy	TRE
L3S	Germany	TRE

[38] https://en.wikipedia.org/wiki/Synthetic_data
[39] https://github.com/odissei-data/awesome-tres-social-sciences

LETHE project infrastructure	Europe	TRE infrastructure
Medical Informatics Initiative (MII)	Germany	TRE
nCloud	Spain	TRE
ODISSEI Secure Supercomputer	Netherlands	TRE
OSSDIP	Austria	TRE Reference architecture
PANCAIM project infra	Europe	TRE infrastructure
Personal Health Train	International	Federated Compute
SANE	Netherlands	TRE, User portal
SATRE	United Kingdom	TRE Reference architecture
SensitiveCloud	Czech Republic	TRE infrastructure
SD Services	Finland	TRE
SeERP	United Kingdom	TRE
Statistics Denmark Remote Desktop	Denmark	TRE
Statistics Finland – FIONA	Finland	TRE
Statistics Netherlands Remote Access	Netherlands	TRE
Statistics Sweden – MONA	Sweden	TRE
TSD	Norway	TRE
UK Longitudinal Linkage Collaboration	United Kingdom	TRE
Vantage6	International	Federated Compute
Virtual Research Workspace	Netherlands	TRE
Wellfort	Austria	TRE

European initiatives

HORIZON-INFRA-2023-EOSC-01-06 projects

Three projects are funded by the EC in HORIZON-INFRA-2023-EOSC-01-06 (Trusted environments for sensitive data management in EOSC) call[40].

SIESTA

SIESTA (Secure Interactive Environments for SensiTive data Analytics) project[41] aims to design and build secure Interactive environments for Sensitive data analytics in the EOSC ecosystem.

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

41 https://eosc-siesta.eu/

The project describes a set of tools, services, and methodologies for the effective sharing of sensitive data in the EOSC, following a cloud-based model and approach to foster the uptake of sensitive data sharing and processing in the EOSC.

The project will deliver trusted cloud-based environments for the management and sharing of sensitive data that are built in a reproducible way, together with a set of services and tools to ease the secure sharing of sensitive data in the EOSC through state-of-the-art anonymization techniques.

The overall objective is to enhance the EOSC Exchange[42] services by delivering a set of cloud-based trusted environments for the analysis of sensitive data in the EOSC demonstrating the feasibility of the FAIR principles over them.

Objectives of the project are:

- Enhance the EOSC Exchange services by delivering a set of cloud-based trusted environments for the analysis of sensitive data in the EOSC demonstrating the feasibility of the FAIR principles over them.
- Study, explore and demonstrate the feasibility of FAIR management and processing of sensitive data, showcasing the benefits for society, science and research.
- Deliver tools for the secure anonymization or pseudonymisation of datasets, allowing rightholders to safely release sensitive data through the EOSC Exchange.
- Provide right holders and other relevant stakeholders with best practices and methodologies for the release of sensitive data following FAIR principles, including design principles for compute infrastructures allowing access to them, exploring the feasibility of FAIR data workflows over sensitive data.
- Extend the service offer and the capabilities being offered through the EOSC portal, coordinating with the operational and management activities carried out by the EOSC partnership and related projects.

The SIESTA project presents 5 use cases to validate TRE solution used in the project:

- Epidemiology
- Medical imaging
- Energy
- Text anonymization on sensitive data
- Demography

The project is coordinated by Spanish National Research Council CSIC[43].

[42] https://eoscfuture.eu/ker/eosc-exchange/
[43] https://www.csic.es/en/csic

TITAN

TITAN (Trusted envIronments for confidenTiAl computiNg and secure data sharing)[44]

is a 36-month project that proposes to develop secure and trustworthy confidential data processing and sharing capabilities, and demonstrate them in the EOSC ecosystem.

The sharing of sensitive data will follow FAIR data and open science principles. The project puts significant emphasis on privacy preservation and AI technological solutions in line with existing ethical, regulatory and legal EU boundaries.

The developed open-source software platform will focus mostly on the two use cases present in the project: government data and healthcare.

By being under the umbrella of the EOSC Ecosystem, TITAN will take advantage of a commonly created brand, already established networks of contacts and working groups, and close collaboration with several other projects, some of them starting also early this year.

To promote community adoption of TITAN's open-source software artefacts, the solution will be practically demonstrated in several vertical cross-border scenarios – notably in the public administration and healthcare sector.

The project is coordinated by Universidad de Murcia.

EOSC ENTRUST

The mission of EOSC-ENTRUST (A European Network of TRUSTed research environments)[45] is to create a European network of trusted research environments for sensitive data and to drive European interoperability by joint development of a common blueprint for federated data access and analysis.

EOSC-ENTRUST brings together providers of operational TREs from 15 European countries with a shared goal to implement, validate and promote their capabilities through a common European framework using shared standards and common legal, operational and technical language.

This blueprint for interoperability is anchored in the EOSC Interoperability Framework spanning the four dimensions of legal, organisational, technical and semantic interoperability.

EOSC-ENTRUST has identified four driver projects covering genomics, clinical trials, social science and public-private partnerships to benchmark capabilities, inform blueprint design and demonstrate secure data analysis using federated workflows.

[44] https://titan-eosc.eu/
[45] https://eosc-entrust.eu/

Targeted outreach activities will expand this open network with further providers and develop policy papers and guidelines for the full range of stakeholders to create a long-term operational TRE framework within EOSC.

EOSC-ENTRUST project has defined its drives as follows:

- Demonstrate scalability and interoperability of the blueprint in a high data volume (genomics) network of local TRE nodes distributed across multiple countries.
- Demonstrate the applicability of the blueprint across very heterogeneous scientific domains, e.g. social and life sciences.
- Demonstrate the potential ability of the blueprint to bridge traditionally very separated data domains of clinical trials and real-world health data in one solution architecture.
- Demonstrate the applicability of the blueprint beyond the academic context in a public-private network with SME providers.

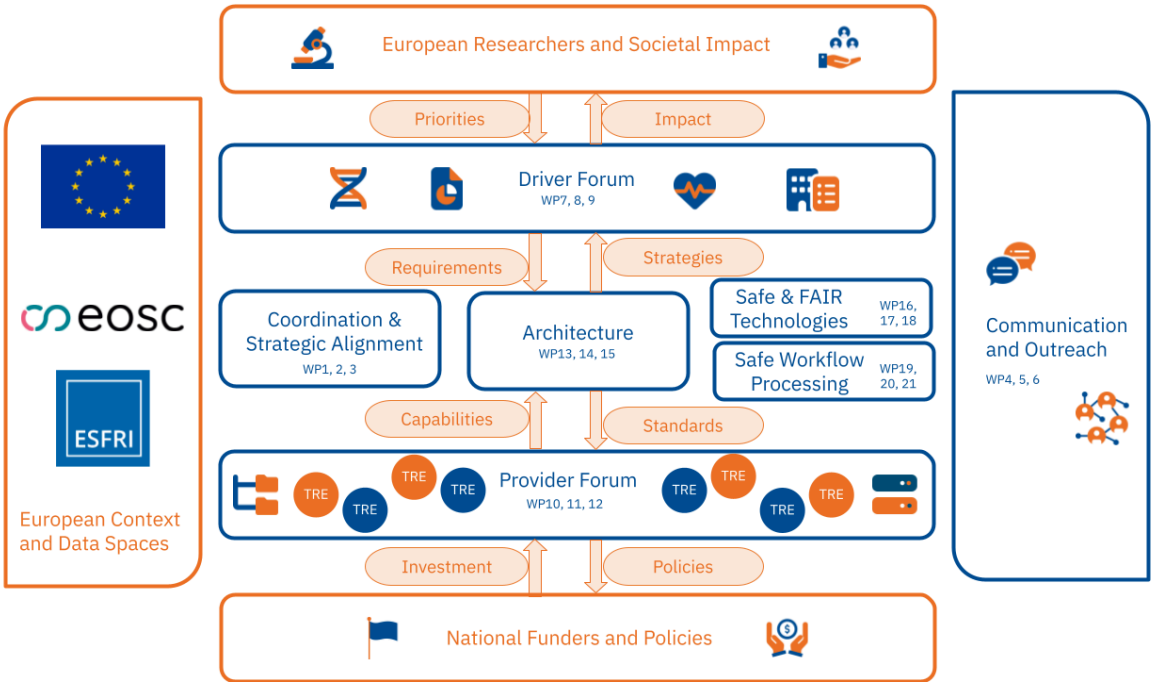

Fig. 8 Structure of the EOSC-ENTRUST

EOSC-ENTRUST is coordinated by ELIXIR[46].

European Health Data Space (EHDS)

Based on the European Commission The European Health Data Space (EHDS) will be a key pillar of the strong European Health Union and is the first common EU data space in a specific area to emerge from the European strategy for data.

[46] https://elixir-europe.org/

The EHDS will[47][48]:

- empower individuals to take control of their health data and facilitate the exchange of data for the delivery of healthcare across the EU;
- foster a genuine single market for electronic health record systems (primary use of data);
- provide a consistent, trustworthy, and efficient system for reusing health data for research, innovation, policy-making, and regulatory activities (secondary use of data).

To build EHDS the Commission supports efforts such as:

- the HealthData@EU pilot project[49]. The project builds a pilot version of the European Health Data Space (EHDS) infrastructure for the secondary use of health data;
- the Xt-EHR Joint Action[50]. The project is dedicated to establishing guidelines for the development of a comprehensive, interoperable, and secure Electronic Health Records (EHR) system that promotes smooth connections among healthcare providers within European Union Member States;
- building on existing infrastructures such as ePrescription and eDispensation, and Patient Summaries which provide information on important health related aspects such as allergies, current medication, previous illness, surgeries, etc.

The Joint Action (JA) Towards the European Health Data Space (TEHDAS), helps EU Member States, and the European Commission (EC) to develop a common framework for the cross-border secondary use of health data to benefit public health and health research and innovation in Europe. The project have published deliverables where are described for example options for the services and services architecture and infrastructure for secondary use of data in the EHDS (D7.2)[51].

This deliverable has analysed for example the users' journey to make data available for secondary uses through the HealthData@EU. Deliverables also describe some architecture scenarios and the TEHDAS' data lifecycle definition. As deliverable stated, "it is clear that the TEHDAS architecture has a direct mapping in the HealthData@EU one, being mostly a "renaming" of the actors participating on it." One component in these architectures are also Secure Processing Environments ("SPEs") .

[47] https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en
[48] https://www.european-health-data-space.com/
[49] https://ehds2pilot.eu/
[50] https://www.xt-ehr.eu/
[51] TEHDAS, Deliverable 7.2, Options for the services and services architecture and infrastructure for secondary use of data in the EHDS, https://tehdas.eu/tehdas1/app/uploads/2023/07/tehdas-options-for-the-services-and-services-architecture-and-infrastructure.pdf

EOSC Association Health Data TF

The EOSC Association Health Data Task Force themes are also related to the sensitive data management and European wide trusted research environment development. The task force was set by the EOSC-A[52] Board.

Key Focus Areas

- Identify business process models (i.e., needs and requirements) of health data research projects, understanding their limitations and capabilities in the context of EOSC.
- Map and align the health research projects business process models to EOSC specific services and tools (such as Findability, AAI, Technical and Semantic interoperability), expecting collaboration with the TFs in the Technical Challenges Advisory Group.
- Propose solutions to enable the alignment with EOSC actors relevant to the Health Data domain(the research health data community, research infrastructures (ERICs) and the e-Infrastructures).
- Foster the development of partnerships among EOSC and EHDS communities.
- Map and align EOSC solutions with the needs of the European Health Data Space for secondary use infrastructure (HealthDAta@EU).

Global initiatives

RDA TRESD WG initiative

“Trusted Research Environments for Sensitive or Confidential Data: FAIRness for Controlled Data and Processes” is the working group initiative proposed within RDA communities.

The working group should discuss about three main problems:

- Blueprints. There are no community-agreed technical blueprints that provide building blocks and adaptable infrastructure components as well as community best-practice processes.
- Interoperability. A user federation (not in the sense of identity federation) is needed for users to work across several SREs/TREs, i.e. have code move from one to another, work with the data locally such as in federated learning in machine learning settings.
- Risks. Balancing risks of data disclosure and utility of the environment for researchers is an ongoing problem for SREs/TREs.

At the time of this report the WG proposal is under the community review in the RDA.

GÉANT Digital Health Data SIG

The SIG-DHD (Digital Health Data)[53] of the GÉANT[54] aims primarily at enhancing coordination, exchange of best practices and sharing knowledge among the National research and education networks (NRENs) and their implied eHealth institutions & projects in the activities

[52] https://eosc.eu/eosc-association/
[53] https://community.geant.org/sig-ehealth/
[54] https://geant.org

they are individually carrying out to support their eHealth User Communities and the management of Health Data (processing, storage, share, access policies).

The SIG organises workshops, webinars and events aimed at gathering and polling the community around the main issues and challenges that managing, sensitive, health data implies. Several aspects of managing Health Data are of interest for the NRENs:

- the required network, AAI, Cloud, Security services in the eHealth domain- all aimed at boosting the potential exploitation of eHealth services and Health Data by the GÉANT and NRENs community.
- liaising with key initiatives like RUTE-AL[55] in South America, the EGI TRE WG, the EOSC-A Task Force on Health Data is one of the objectives, also in relation to the contribution to a shared, understood, agreed vision on the implementation and implications of the EHDS regulation in the EU.
- SIG-DHD has organised a reference survey[56] about eHealth for the NRENs, in 2024, aimed at understanding what the main challenges, issues, possible benefits of supporting the management of health data is for the NRENs and their user communities.

TREs integration scenarios in the EGI infrastructure

There are at least three different integration scenarios for TREs in the EGI infrastructure. Real integration scenarios will be studied and described in the upcoming work of the EGI's TRE WG. In this section some general level approaches are discussed as a possible seed for the next phase of the TRE WG.

EGI Federation focuses on Private Cloud TRE or Federated TRE models, but institutional TREs are relevant as well, because if an institute of the EGI Fedederation operates like that, then it can be open for helping other institutes in the federation to become such an service provider. EGI Federation would play a knowledge and technology transfer facilitator role.

In the following chapters some preliminary ideas for EGI infrastructure integrations are discussed.

On demand deployment (Private Cloud TRE)

This scenario focused on the delivery "on demand" of a TRE service based on the needs of customers/projects. This scenario has been developed for instance in the LETHE[57] project by EGI Foundation.

This scenario implements among other things:

- Controlled data operations

[55] https://www.redclara.net/en/colaboracion/conozca/red-universitaria-de-telemedicina-de-america-latina-rute-al
[56] https://wiki.geant.org/display/EHE/eHealth+-+Home
[57] https://www.lethe-project.eu/

- Automatic provisioning of services based on the development of TOSCA/Ansible recipes to be then executed via the EGI Infrastructure Manager service[58]
- Integration of EGI Check-in service as AAI
- Possibility to tailor solutions for the single project
- Tailored integrations between applications and services
- Organisational and legal interoperability defined in the project documentation (DPIA etc.)
- Semantic interoperability defined internally
- External FAIR data solution requires collaboration with external communities and should be part of the project (if HE project)

EGI FedCloud based TRE (Federated TRE)

This is an extension of the EGI's FedCloud[59] platform. It uses basic services of the FedCloud and could be managed with the same tools. This is also a further developed possibility from the project based approach.

- Predefined solution between some EGI Federation members to make integrations between local/national/RI level TREs possible
- Technical interoperability requires clear architecture definition by EGI Federation and participating members
- Organisational and legal interoperability have to be agreed the way which takes into account also national differences
- Semantic interoperability is at least partly possible to achieve by general or global disciplinary specifications
- Infrastructure solutions and functionalities for the FAIR data operations are possible to create for all projects involved in the TRE

Integration with EOSC Nodes (Federated TRE)

The third approach extends further the EGI FedCloud based idea. This solution integrates TRE also to the EOSC Node.

- Follow the EOSC EU Node rules.
- Technically similar to the EGI FedCloud version (based on the idea that the EGI FedCloud could be a part of the Node).
- This TRE have to solve organisational, legal and semantic interoperability issues as described above or the way where Node have common rules and legal frameworks
- This offers obvious integration with the EOSC communities.
- FAIR data principles are a clear part of the solution.
- Collaboration with data spaces.

[58] https://www.egi.eu/service/infrastructure-manager/
[59] https://www.egi.eu/service/cloud-compute/

Roles of EGI Foundation and EGI sites

When integrating TREs in the EGI infrastructure, participation of EGI Federation members are needed. Therefore it is important to define roles for actors. The following table has presented some possible roles for EGI Foundations and EGI sites in three scenarios presented in the chapter *Implementation and interoperability architectures.*

In all scenarios EGI Federation members have a service provider role. EGIU Foundation role varies case by case.

	Institutional	**Private Cloud**	**Federation**
EGI Foundation	Promotion, visibility, knowledge sharing	Facilitation of deployment	Federation with several service providers (protocols, SLA/OLA etc.)
EGI sites	Granular solutions	Cloud backend	SP, cloud provisioning

With institutional TREs EGI Foundation roles focus on non-technological activities such as knowledge sharing and promotion. Service providers provide granular solutions for specific needs.

In the private cloud scenario EGI Foundation is able to facilitate deployment and coordinate deployment from the service provider's cloud backend. For example some projects use this model of solutions. In this case there is one service provider for several users served by cloud backend.

Federation models offer services and resources from several service providers and then governance protocols and for example SLA or OLA processes are required. Service providers are responsible for cloud provisioning and other cloud operations.

Conclusions

Based on the 1 year work of the TRE working group, we identify the following challenges on the TRE landscape that is relevant for EGI:

1. While the landscape of operational TREs is broadening, it cannot keep up with the growing demand in Europe and in the various disciplines where researchers need to store and process sensitive data according to institutional, national, domain-specific and European/global regulations. We **need to speed up TRE deployment and adoption**.

2. Current TREs typically serve a single organisation or a single country. The mid-future challenge will be the interconnection of those TREs, and keeping them interoperable over extended periods of time. We need to find suitable **operational, business and governance models for interconnected TRE ecosystems**.
3. The use of TREs is broadening and reaching beyond research institutes, to large initiatives (EOSC), commercial environments, and public institutions. The terminologies and approaches within these settings can be radically different, limiting our abilities to reuse solutions. We need to continue identifying and researching **vocabularies and standards that can facilitate the reuse of TREs in different sectors**.

While the current WG ends, the WG proposes a second phase to it, where we tackle these three challenges. The importance of collaborating with other key players is recognized, and therefore, it is proposed that this second phase be conducted jointly with the GÉANT Special Interest Group on Health Data.