VRE4EIC

A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration

Deliverable D5.4

Strategies for the VRE end-users to handle security, privacy and trust issues – second version

Document version: Final

VRE4EIC DELIVERABLE

Name, title and organisation of the scientific representative of the project's coordinator:

Mr Peter Kunz t: +33 4 97 15 53 06 f: +33 4 92 38 78 22 e: peter.kunz@ercim.eu

GEIE ERCIM, 2004, route des Lucioles, Sophia Antipolis, F-06410 Biot, France

Project website address: http://www.vre4eic.eu/

Project	
Grant Agreement number	676247
Project acronym:	VRE4EIC
Project title:	A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration
Funding Scheme:	Research & Innovation Action (RIA)
Date of latest version of DoW against which the assessment will be made:	31 May 2017 Amended Grant Agreement through amendment n°AMD-676247-8
Document	
Period covered:	M1-M33
Deliverable number:	D5.4
Deliverable title	A strategy for the VRE end-users to handle security, privacy and trust issues – second version
Contractual Date of Delivery:	30.06.2018
Actual Date of Delivery:	30.06.2018
Editor (s):	Yi Yin (TU Delft), Anneke Zuiderwijk (TU Delft)
Author (s):	Jacco van Ossenbruggen (CWI), Cesare Concordia (CNR), Maria Theodoridou (FORTH), Laurent Remy (euroCRIS)
Reviewer (s):	Laura Hollink (CWI), Valerie Brasse (euroCRIS)
Participant(s):	All
Work package no.:	5
Work package title:	Information management policy, security, privacy and VRE trustability
Work package leader:	CWI
Distribution:	PU
Version/Revision:	1.0
Draft/Final:	Final
Total number of pages (including cover):	30

What is VRE4EIC?

VRE4EIC develops a reference architecture and software components for VREs (Virtual Research Environments). This eVRE bridges across existing e-ERIs (e-Research Infrastructures) such as EPOS and ENVRIPlus, both represented in the project, themselves supported by e-Is (e-Infrastructures) such as GEANT, EUDAT, PRACE, EGI, OpenAIRE. The eVRE provides a comfortable homogeneous interface for users by virtualising access to the heterogeneous datasets, software services, resources of the e-ERIs and also provides collaboration/communication facilities for users to improve research communication. Finally it provides access to research management /administrative facilities so that the end-user has a complete research environment.

Disclaimer

VRE4EIC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676247.

Table of Contents

1 Introduction

For the eVRE developed by the VRE4EIC project, the security, privacy and trust requirements significantly vary amongst the potential end-users from various research domains and the public. In deliverable D5.1 and D5.3, the VRE4EIC project has defined the strategies handling potential issues and eRIs regarding security, privacy and trust aspects. However, these strategies are mainly produced to help the technical developers better design the eVRE and choose appropriate technologies for the implementation of the eVRE. We assume that most of the potential end-users have limited ICT knowledge so that they might have a different understanding or interpretation regarding these strategies to deal with security, privacy and trust related issues. In order to help our potential end-users understand the logic and consideration of the strategies developed in D5.1, in the deliverable D5.2, we have clarified these strategies in the early stage of the development of eVRE. After implementing an eVRE prototype that have be tested by EPOS and ENVRIPLUS user groups, this deliverable defines strategies and recommendations regarding security, privacy and trust for the end-users to take when they use the eVRE. The user is a generic actor representing any user accessing the eVRE (according to its definition, "VRE users" only concern people that want to access research data). This approach will also help the engagement of potential end-users. Upon acceptance of this strategy document, it will be made publicly available and especially distributed to target users of the existing e-RI initiatives and VRE-related initiatives.

This deliverable is the results of Task 5.2, which identifies issues related to security, privacy, and trust for the use of the eVRE, and which defines a strategy to handle these issues. The strategies provides the end-users of the eVRE insight in the various security, privacy and trust issues that they may face when they use the eVRE. Since those issues may considerably hinder the use of VREs similar to eVRE for excellent research, the strategy is expected to stimulate VREs uptake.

More specifically, the operations performed as part of Task 5.2 include the following:

- Identify and analyse strategies of existing VRE-related projects for users to handle security, privacy and trust issues;
- Take the requirements regarding security, privacy and trust issues from WP2 and analyse the gaps between those and how existing user strategies handle these issues (e.g. by training users);
- Closely collaborate with the whole project partners and define:
    - how VRE4EIC users can deal with security issues;
    - how VRE4EIC users can deal with privacy issues;
    - how VRE4EIC users can make use of contextual metadata for data citations (to enhance trust of researchers in the data and in the data provider)
    - how measures of certainty of data collection techniques and analyses affect users' trust in system results.
- Translate the strategies to tutorial videos for the VRE end-users and closely collaborate with the WP6 project partners for this;
- Provide the results of these task operations to WP6 for implementation in the training materials;
- Provide the results of these task operations to WP3 for implementation in the learning environment;
- Use the feedback obtained through the evaluations in WP2 and WP6, and update the strategy;
- Make the strategies publicly available so that users of other VRE related initiatives and VREs can also benefit.

This deliverable describes the strategies from both technical and data governance perspectives regarding the security, privacy and trust aspects recommended by the VRE4EIC project.

1.1 Potential security issues for users

Security is a broad concept. In this deliverable we focus on information security. There are many security issues around the development and the use of the eVRE. It is rather difficult to provide a complete detailed list of information security issues, so only key security issues are discussed in this section. Sipoen et al. (2007) propose four high-level abstractions of information security issues including access to information systems, secure communication, security management, and the development of secure information systems (Siponen & Oinas-Kukkonen, 2007). Table 1-1 presents key security issues on the basis of this categorization that need to be considered in the VRE4EIC project. These key issues have been identified from the requirements interviews in the Working Package 2.

Potential Security-related issues		Potential Risks
Access to the eVRE	Uncontrolled subjects' (refers to information entities, e.g., processes, humans) access to objects (e.g. files, directories, services, tools), e.g., subjects are not the real subjects which they claim to be.	Information may be stolen, modified or used nefariously or resold by unauthorized subjects. Individuals are able to deny an action that they have carried out.
	Undesired subjects' access to objects	Information may be stolen, modified or used nefariously or resold by unauthorized subjects.
	Unwanted information flows between objects and subjects, e.g., social engineering attacks, malware attacks, denial-of-service	Information may be stolen or disclosed to unauthorized or albeit authorized subjects. Process of dwindling takes places (Siponen & Oinas-Kukkonen, 2007)
Secure communication	The act of information communication is open and can be touched	Information can be stolen and disclosed unwillingly
Security management	Lack of information management policies, e.g., data backup, recovery, contingency mechanism	Information may be lost No appropriate response actions against ceRIsis events
Secure development of eVRE	Security requirements are not properly collected, are misinterpreted or are missing	Security requirements cannot be properly reflected in the development of eVRE
	Security requirements cannot be fulfilled by certain technological solutions.	The security cannot be guaranteed, which damages service providers' reputations and reduces user trust

Table 1-1 Potential security issues for end-users of the eVRE

1.2 Potential privacy issues for users

Privacy is a fundamental human right, enshrined among various individuals, cultures and legislations. However, there is no universal definition of privacy. The term "privacy" derives from the Latin word "privatus" and "privo" meaning "to deprive" (Leino-Kilpi et al., 2001). In English-language dictionaries, privacy refers to "a state in which one is not observed or disturbed by other people". Although the boundaries and specific contents of privacy vary significantly in different countries, commonly identified elements of privacy are 'the right to be left alone' and the 'control of information about ourselves' (Pearson, 2009). The new General Data Protection Regulation defines 'personal data' as "any information relating to an identified or identifiable natural person" (EU, 2016). Table 1-2 shows the potential privacy-related issues for end-users. These key issues have been identified from the requirements interviews in Working Package 2.

Potential privacy issues	Potential Risks
Beyond the initial data processing purpose, such as research purpose	Violate the purpose limit principle of the GDPR
No consent from data subject or no consents logs	No transparency about the data processing
Researchers are forced or persuaded to be tracked or to provide personal data against their will (Pearson, 2009)	Individual information may be stolen, used nefariously or resold unauthorized
Big data analytics can be used to identify individuals (Montjoye, Radaelli, Singh, & Pentland, 2015)	Individual information can be exposed
Rapid changing data challenges the service providers' ability to maintain consistent security standards	The quality of data service cannot be ensured. There is negative impact or even damage to the data service providers' reputation
Dynamic movement of data in an online environment to share research information. It is not clear which party is responsible for ensuring legal compliance for personal data The sub-contractor involved in processing the data cannot be properly identified, checked or ascertained (Sun, Chang, Sun, & Wang, 2011)	Damage to the service providers' reputation
Privacy breaching event	Damage to information holder's reputation Governmental supervisor might lose credibility due to lack of proper governance
Regulation and legislation on privacy protection are behind the development of technologies	Responsibilities are not clear. Privacy information might be exposed. Governmental supervisor might be blamed and lose credibility due to lack of proper governance

Table 1-2 Potential privacy issues for end-users of the eVRE

1.3 Potential trust issues for users

There are many definitions of trust. According to Rotter, trust refers to "a generalized expectancy held by an individual or group that the word, promise, verbal or written statement of another individual or group can be relied upon"(Rotter, 1967, p. 444). Trust is defined as a vulnerable willingness or expectation that the commitments will be fulfilled by other people or organizations which entails Risks for the trusting party (Mayer, Davis, & Schoorman, 1995; Panteli & Sockalingam, 2005; Rotter, 1971). In Table 1-3, the potential trust issues for users are presented, derived from the requirements interviews in Working Package 2.

Potential trust issues		Potential Risks
Technical trust	Data quality	Low quality data will damage the trust on the data
	Various data standards	Various data standards hinder the interoperability of data
	Secure storage of data	Data might be lost, stolen or leaked
	Data Confidentiality and auditability	The source changes of data may not always be traceable
	Usability of services/Service continuity	Data service might be interrupted or out of service capability
	The stability and security of the underlying IT infrastructure	The underlying infrastructure might be unable to deliver reliable computing or storage services
Social trust	Legal compliance	Not all parties involved undertake the legal compliance
	Human behaviour	If the privacy information cannot be ensured, the level of trust will be reduced

Table 1-3 Potential trust issues for end-users of the eVRE

1.4 Structure of the document

First of all, this document provides an overview of current issues with regards to security, privacy and trust issues. Section 2 recalls the requirements clarified in D5.1 and D5.3 on these three aspects. Section 3-5 describes the security, privacy and trust strategies developed by this project for the end users of eVRE. Finally, some recommendations are provided in the conclusion section.

2 Requirements related to security, privacy and trust

D2.1 describes the steps that have been followed to elicit requirements for the eVRE. The requirements regarding security, privacy and trust have been further described in D5.2, "A strategy for the VRE4EIC project to handle security, privacy and trust issues". We expanded the list of requirements related to these three aspects. These requirements are used to design the AAAI component of the eVRE. In this

section we review the security, privacy and trust requirements. Chapters 4, 5 and 6 discuss the implications of these requirements for the end-user of the eVRE.

2.1 Requirements related to security

The requirements elicitation process in WP2 has provided the following list of requirements (expectations from the users, Table 2-1). Work done in WP3 has identified functionalities (functions) and technical solutions (down to microservices) to fulfil those requirements including ensuring several security features thanks to the AAAI microservices.

Requirement ID (from D2.1)	Requirement	Requirement description
CTRQ1	Login	Login with a user account and password
CTRQ2	Continuous access	Access to the software, services and datasets anywhere with internet connection
CTRQ3	Single login	Ability to gain multiple accesses to the system with one login
CRQ6	Data Storage & Preservation	Ability to deposit (over long-term) the data and metadata or other supplementary data and methods according to specified policies, and make them accessible on request.
SRQ12	Secure storage	Secure storage of data, especially sensitive data
SRQ15	Physical access control	Identity control of the access to the physical infrastructure
CLRQ1	Instrument Integration	Ability to create, edit and delete an instrument or sensor which will generate data
CLRQ4	Instrument Access	Ability to read and/or update the state of an instrument
CLRQ15	Data Transmission	Ability to transfer data over communication channel using specified network protocols.
PRQ35	Data backup	Ability to backup datasets according to specified policies
SRQ6	Use log	Logs of the system usage for auditing and legal compliance
PRQ31	Accounting	Accounting services for data and services provider

Table 2-1: Identified requirements related to security, derived from WP2.

At the VRE level, this means that the VRE should (1) be compatible with several external access mechanisms, (2) be able to include new ones when new e-ERIs connect to the VRE and (3) allow unrestricted access to open data. In the latter case, the e-ERIs users should be warned about potential additional privacy Risks when their data is combined with other datasets (differential privacy).

Physical access control (in addition to the standard digital access control) is not used by the eight currently characterized e-ERIs in D2.1, even though it was identified as a requirement. Individual e-ERIs should determine how much priority this requirement has for their user groups.

The logging of user actions and accounting may be implemented at both the e-RI level and the VRE level. The VRE logs allow for a complete picture of user actions across the various e-ERIs. Note that while accounting relies on an identification of users, logging of actions of non-registered users is useful as well to provide overall usage statistics.

Secure data storage, backup and secure transmission of data are handled at the e-RI level. Here, the task of the VRE is to provide (CERIF) metadata about the provided level of security, e.g. whether encryption is used.

2.2 Requirements related to privacy

In the implementation and operation phase, the eVRE system will process data containing personal data. In general, there are three sources of data containing personal data on the eVRE, including: 1) The metadata retrieved from ERIs or connected VREs, describing the information of registered users, such as names, contact information, affiliation, access roles, etc.; 2) data retrieved from the research datasets shared by ERIs or connected VREs, can be directly or be combined with other data to identify an unique person, such as GPS location information, IP address, device MAC address, or biometric information; 3) data generated by the eVRE itself, such as user registration information, user access logs. The first two conditions are the main sources of personal data processing in the eVRE. However, eVRE needs to design comprehensive privacy preserving mechanisms to safeguard all personal data from these three sources.

The VRE should guarantee the privacy of both users of the eVRE and of sensitive research data that is stored through the VRE. Access Control (CRQ6), secure storage (PSRQ2) and transmission (DRQ14) of research data were already mentioned as security-related requirements. We mention them under privacy again since they are fundamental in protecting privacy-sensitive research data. In addition, the identities, access credentials as well as transaction logs of users of the VRE should be stored securely (PSRQ3). This includes the metadata stored in the VRE Metadata Manager (in CERIF format).

Differential privacy: The fact that the VRE bridges across several e-ERIs poses additional challenges with regard to privacy. In D2.1 on requirements elicitation, it was noted that "Datasets often require removing privacy sensitive variables from it before publication. [...] Moreover, the combination of data with other sources might still make it possible to track the identity of an individual person, especially when open data are combined with social media data. " This means that the privacy levels of data in an e-RI are not always strict enough for a VRE. This results in additional requirements related to resetting access control settings (e.g. to disallow combination of data when an e-RI becomes part of the VRE), creating awareness with data providers (that their previous privacy policy might no longer be enough). These requirements are described below in Table 2-2.

Requirement ID	Requirement	Description
CRQ6	Data Storage & Preservation	Ability to deposit (over long-term) the data and metadata or other supplementary data and methods according to specified policies, and make them accessible on request.
SRQ12	Secure storage	Secure storage of data, especially sensitive data
CLRQ15	Data Transmission	Ability to transfer data over communication channel using specified network protocols.

SRQ13	Credentials protection	Ability to protect the user's' digital identities and credentials
Additional requirement 1	Consent	Consent by data subjects for data collection, data processing and data storage
Additional requirement 3	Transparency	Adequate insight into the purposes and means of a data processing, informing data subject about what personal data is being processed, when, and by whom for which reason.

Table 2-2: Identified requirements related to privacy, derived from WP2

2.3 Requirements related to trust

The elicited requirements (see Table 2-3) show a clear need of users for methods to cite data (IRQ4): they need to be able to uniquely identify datasets (IRQ1), including parts of datasets (IRQ1) or specific versions of datasets (CRQ4); they need a guarantee that identified data will not change and will remain accessible (CRQ6). This enhances the reproducibility of studies done on the basis of these data. In addition, these identification mechanisms provide a means to keep track of changes made to datasets, in other words, to record the provenance. Finally, the opportunity to verify the quality of the data (CRQ6 and CRQ3), improves the transparency of the research process.

We observe that in some cases there may be a tension between the need to record provenance of datasets, including information on who did what, and the need to protect the privacy of users, including their identities and access logs (SRQ6 in Table 2-1 above). A VRE needs to have a clear policy regarding this issue.

Requirement ID	Requirement	Description
IRQ1	Data Identification	Ability to assign (global) unique identifiers to data contents.
CRQ4	Data Versioning	Ability to assign a new version to each state change of data, allow to add and update some metadata descriptions for each version, and allow to select, access or delete a version of data.
CRQ6	Data Storage & Preservation	Ability to deposit (over long-term) the data and metadata or other supplementary data and methods according to specified policies, and make them accessible on request.
CRQ6	Data Quality Checking	Ability to detect and correct (or remove) corrupt, inconsistent or inaccurate records from data sets.
CRQ3	Data Quality Verification	Ability to support manual quality checking.
CRQ7	Data Replication	Ability to create, delete and maintain the consistency of copies of a data set on multiple storage devices.
CLRQ18	Data Publication	Ability to provide clean, well-annotated, anonymity-preserving datasets in a suitable format, and by following specified data-publication and sharing policies to make the datasets publicly accessible or to those who agree to certain conditions of use, and to individuals who meet certain professional criteria.

IRQ4	Data Citation	Ability to assign an accurate, consistent and standardized reference to a data object, which can be cited in scientific publications.

Table 2-3: Identified requirements related to trust, derived from WP2

At the VRE level, the main requirement is to correctly convey the information that is already present at the e-RI level (including data ownership, licensing and liability) of each dataset as metadata, preferably in CERIF format.

3 Security Strategies

In order to help VRE4EIC users deal with security issues when using the eVRE system, the VRE4EIC consortium should help making users aware of information security issues and understand the reasons for the technological solutions offered in the eVRE. Since the eVRE system is built on many e-ERIs and connected with many other VREs and e-ERIs, users need to understand the complexity of eVRE and adapt their using behavior to the characteRIstics of the eVRE and develop their own security strategies when using the eVRE. Although information security can be enhanced via technological solutions such as access control, data encryption, firewalls, these technical measures alone are not sufficient in mitigating the Risks to information security. Human interactions with technical measures could lead to serious threats to information security like fraud, social engineering or privacy information re-identification via big data analytics. Therefore, information security needs to be enhanced by considering technical solutions and human behaviors.

Veiga and Eloff (2007) proposed a comprehensive information security governance framework by consolidating four approaches towards information security governance including ISO/IEC 177995 and ISO/IEC 27001, PROTECT model, Capability Maturity Model and Information Security Architecture (A. D. Veiga & J. H. P. Eloff, 2007). This model lists a set of components to consider for information security governance.

In this section we will use this model to discuss the implications of technical solutions regarding security as well as the governance solutions to help users handle potential information security issues.

3.1 Technical strategies

The eVRE reference architecture supports federated authentication and authorization. There are two main scenarios. In the first, the eVRE acts as a Service Provider (SP) that delegates all authentication and authorization to a federated service such as EduGAIN. This allows users that are already part of an organization within the federation to log onto the eVRE using the credentials of their own organization by using their organization's Identity Provider (IdP). In the second scenario, another VRE or RI (such as EPOS) acts as the Service Provider (SP) and uses the VRE4EIC eVRE as the Identity Provider (IdP) that can either provide additional accounts for users of organizations not represented in the federation or further proxy authentication to the same federations as in scenario 1. Both scenarios allow single sign on (SSO), and in both scenarios the user only provides credentials to the IdP service, not the SP. Single sign-on is an advantage for end-users as it allows them access to multiple associated e-ERIs without the overhead of creating multiple accounts.

In the federated authorization scenario, the amount of information about a user that is disclosed to the Service Provider by the IdP is always under control of the user. The user needs to disclose the information needed by the SP to decide if the user is indeed authorized to access a certain resource. In some cases, it is desirable that the SP provider can enrich the user profile with attributes from its secure CERIF metadata catalog, also after prior permission from the user. This, can, for example, be used to support Role Based Access Control (RBAC), where access to an authenticated user is granted

based on the roles and the affiliations the user has according to the metadata recorded in the catalog. This requires that the user needs to be able to rely on the information in the catalog being correct. Research Institutes will typically be the providers of the metadata used for authentication. As discussed in D5.3, the potential security implications of incorrect CERIF metadata are large, and responsibility for the maintenance of this metadata should be clearly allocated within an organization.

Multi-factor authentication is recommended for the eVRE when a user accesses sensitive data. This is a method of confirming a user's claimed identity by utilizing a combination of multiple (typically two) different components (factors). Using two-factor authentication (2FA), users can authenticate using something that only the individual user knows (for instance login and password) plus a one-time-valid, dynamic passcode that is received or generated by the user at the time of authentication. When designing a 2FA method, the IdP must take into account security (e.g. whether messages can be encrypted), privacy (e.g. whether additional user information needs to be stored) and ease-of-use (e.g. the number of (additional) devices that a user must carry). 2FA by definition presents the user with an additional burden. However, several 2FA solutions have been shown to be minimally intrusive. In the scenario where the eVRE plays the role of the SP, the decision to use 2FA or not is also fully delegated to the IdP of the user. In the scenario where the VRE4EIC eVRE serves as the IdP, the current prototype uses Telegram as a 2FA method. Telegram is a free messaging app that encrypts communication (low security risk), does not store personal user information such as telephone numbers (low privacy risk), and can be installed on various devices (relatively low burden of carrying an additional device).

3.2 Governance strategies

In the D5.2, we have used the information security governance model proposed by Veiga and Eloff (A. D. Veiga & J. H. Eloff, 2007) to help define our project security strategies. The project consortium defines the VRE4EIC development strategy which includes the project objectives. The consortium collectively defined the risk management strategies including information security strategies and identifies the security requirements. The work package leader defines the security control plans, policies standards which need to be used when developing practical security solutions. The evaluation criteria are also defined by the work package leader and are approved by the whole consortium.

From the view point of end users, they need to adapt their daily behaviors to incorporate with the compliance with information security. They should define their own information access policies, such as what information can be accessible by whom. They need to regularly attend training and education seminars organized by VRE4EIC team to update their knowledge regarding information security.

4 Privacy Strategies

The General Data Protection Regulation (GDPR) came into force on May 25, 2018, in the vision of better safeguarding personal data processing in the new digital era. The Art. 4 of the GDPR defines personal data processing as "means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction"(EU, 2016). Seven principles are defined to enable legitimate personal data processing, including:

(a) Lawfulness, Fairness and Transparency

Sufficient transparency would give adequate insight into the purposes and means of a data processing, informing data subject about what personal data is being processed, when, and by whom for which reason. Transparency needs to cover the whole lifecycle of the data, from the moment of collection onwards to the stages of processing, storage and deletion. It is rather explicit and also inherent in many different articles of the GDPR, for example in: "• Art. 12 (Transparent information, communication and modalities for the exercise of the rights of the data subject) • Art. 13 (Information to be provided where personal data are collected from the data subject) • Art. 14 (Information to be provided where personal data have not been obtained from the data subject) • Art. 15 (Right of access by the data subject) • Art. 30 (Records of processing activities) "(Schlehahn & Zwingelberg, 2017) .

Recommendation 1 : The end users can check whether they can easily get access to the privacy policy of the eVRE and connected datasets, software and services.

(b) Purpose limitation

Purpose limitation in the GDPR expresses a limitation of personal data processing for specific lawful purposes in the first place. Personal data must be collected only for specified, explicit and legitimate purposes (purpose specification). Such specification of a purpose also guarantees the transparency principle of data processing and use of personal data from the data subject. Data processing operation going beyond the original purpose without data subject's consent or other legal ground and as such is not permitted. As a consequence, the purpose of the processing must be determined already prior to the collection of the information.

For the further processing of personal data beyond the initial purpose, Article 6 para. 4 of the GDPR gives a concluding list of permissible cases, which are:

- the consent of the data subject,
- the processing is based on union or Member State law constituting a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1) GDPR,
- the processing is not incompatible with the initial purpose.

In the practice of the eVRE operation, the purpose of personal data processing is not limited to scientific research. We also would like to attract more commercial organization to use shared datasets. Therefore, before sharing datasets containing personal data, eVRE management body need to get the consent of data subject if the initial data process purpose need to be updated.

Recommendation 2 : The end users may check whether the use their personal data is only limited to the delivery of the services provided by eVRE and other connected eRIs and VREs.

(c) Data minimisation

Any personal data collection must be adequate, relevant and limited to what is necessary with regards to the specific purposes. The necessity of data minimisation is closely related to the idea of the unlinkability of data (Shadish & Galindo, 2010). Unlinkability means that personal data should not be linked across different domains for other use beyond the initially specified purposes. Therefore, unlinkability is an enforcement of purpose limitation and necessity (including data minimisation).

Recommendation 3 : The end users may check how eVRE collects and stores the personal data concerning him or her. They can ask for the full records of the data stored by eVRE and check whether limited data is collected.

(d) Accuracy

According to Art. 5 para 1 (d) of the GDPR, data must be *'accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay'*. This reflects

that the end users of eVRE system have control over their own data requiring always effective intervenability on data processing activities. In the literal sense, intervenability is the operational access to processes and data either by effective technical or organisational means, for example being able to modify or delete inaccurate personal data.

Recommendation 4 : The end users may frequently check the accuracy of the data concerning themselves on the eVRE.

(e) Storage limitation

Article 5 para. 1 (e) of the GDPR requires that data must be *'kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed'*. This principle is intertwined with purpose and necessity as well. Thereby, in research data sharing contexts, such as the eVRE system, where data is collected from multiple and mixed-format sources, it usually depends on the context which data categories need to be stored only for a certain period of time.

Recommendation 5 : The end users may check whether the storage of the personal data is limited to a certain time span.

(f) Integrity and confidentiality

Integrity ensures that systems fulfil reliable functions and performance without corruption, damage or loss of personal data. Confidentiality of personal data means there is no access to the personal data or knowledge of the personal data by unauthorised third party

Recommendation 6 : The end users can frequently check the storage of the personal data is safeguarded and will be notified when there is data breaching incident.

(g) Accountability

Paragraph 2 of Article 5 of the GDPR allocates the legal responsibility to the data controller and demands that the controller must be able to demonstrate compliance with the rules laid down in the GDPR. This can be done e.g. by documenting the legal basis, the purposes and the means of a specific processing operation types, which includes:

- The categories of personal and data formats intended to be used
- The sources of these data categories
- The purposes for which it is intended be to used
- The legal ground on which the processing operation is based
- Technical systems involved (hardware, software and infrastructure)
- The processing entity's internal organisation and human resources involved when processing of data with the systems

Article 83 of the GDPR manifests the general conditions for imposing administrative fines, whereby under para. 5 (a), with reference to Article 5, infringements on the basic principles can lead to fines up to **20,000,000 EUR**, or in the case of an undertaking, **up to 4 % of the total worldwide annual turnover of the preceding financial year**, depending on which amount is higher.

In the section 2.3, we have presented three sources of data containing personal data. Processing these three sources of personal data may follow into two scenarios.

- In the scenario of processing data containing personal data completely generated by the eVRE system itself, it is eVRE's fully responsibility for safeguarding this personal data.
- In the scenarios of eRIs providing their user registration information or research datasets containing personal data to the eVRE, both eRIs and eVRE should be responsible for personal

data processing. Depending on how the contract defined between eVRE and ERIs or VREs, both of them can be data controller, or eVRE is with the role of data controller while ERIs are data processor who process user information on behalf of eVRE. No matter how the contract is defined between eVRE and ERIs, this only influence whether ERIs will be considered as the role of joint data processor together with eVRE or data controller. eVRE is always

Recommendation 7 : the end users should learn who is the responsible body for protecting their personal data and to which authority body they can refer when there is privacy dispute.

The principles and the rights of data subject defined in the GDPR provide guidelines for the end-users to review the design and operation of eVRE or other similar VRE systems and to take relevant strategies for their own privacy preserving.

4.1 Technical strategies

As stated in D5.1 and D5.3, "The eVRE should guarantee the protection of both personal research data that is accessed via the eVRE and personal data about the users of the eVRE and their actions on the system." The protection of research data is mostly related to proper authorization of end-users for access to the research data. This has already been discussed in section 3 of this deliverable regarding security.

Regarding the personal data of users of the eVRE, the eVRE will add an additional level above that of the e-RI level. An eVRE implementation may decide to partially duplicate or centralize personal data about users originally stored only at the e-RI level. On the other end of the spectrum, by using federated authentication and authorization infrastructures such as eduGAIN, both e-ERIs and eVREs may also delegate storage of user data to the federated infrastructure. Even in the fully federated case, users still need to learn where and by whom their data is actually stored. This is because user attributes may be stored by the identity provider of their home institute or on an additional attribute management platform provided by the federated authority infrastructure.

Several technical mechanisms are designed or recommended by the eVRE system. These mechanisms will be briefly mentioned in the following sections while more detailed information can be found in the deliverable D5.3.[1]

4.1.1 Authentication & Authorization

Personal data should only be accessible after the identity of the data requester is confirmed. Authentication is the process which establishes this identity confirmation (Nuffelen, 2016). Authorization is the process to confirm whether the identified user has the right to access to a certain resource and to execute a service (Nuffelen, 2016). Access control for web services has reached a maturity, supported by several initiatives, namely EGI36, EUDAT, AARC. The VRE4EIC project has designed two-factor authentication and role-based authorization components to control data access, especially to control personal data.

4.1.2 Encryption

A second recommended measure to safeguard privacy is the data encryption. Encryption is the process of encoding the information so that it is only understandable by trusted and authorized parties having the key to access it (Nuffelen, 2016). Secure HTTPS connections have already become the default in the process of data transmission. Providing data related services on the eVRE over HTTPS not only protects the personal data or metadata containing personal data, it also prevent other parties eavesdropping when personal data is being transferred over the Internet. In addition, cryptographic

[1] https://www.vre4eic.eu/images/Public_deliverables/D5.3_A_strategy_for_the_VRE4EIC_project_to_handle_security_privacy_and_trust_issues_V2.pdf

hashes are used by a wide variety of techniques to prevent data to be tampered with after publication, such as version control, nano-publications and distributed ledgers.

4.1.3 Anonymization

Anonymization is a technique turning a source dataset into an equivalent dataset with respect to some properties so that the real data subjects present in the source dataset cannot be derived from the processed dataset (Nuffelen, 2016). According to the recommendation of the GDPR, anonymized data can be used more freely(EU, 2016). However, in the current eVRE ecosystem, the application of anonymization will rely on the underlying eRIs and connected VREs.

4.1.4 Purpose based data storage & data access

The GDPR emphasizes that the data is only to be stored, used and shared for specific purpose with explicit consent from the data subject. Ideally a data processing environment should only request data for which it has the permission to get it at the time it needs (Nuffelen, 2016). However, current data access is grant during the whole time of a data processing. Therefore, data access based on more attributes such as subject, environment or source is new research field. Attribute Based Access Control (ABAC) and Context Based Access Control (CBAC) are the predominant works in this area. The Within these two data process mechanisms, purpose or context may be defined as an attribute to decide on when data processing right and data access right will be granted.

4.2 Governance strategies

As one element of the information security governance framework, we can also use the same strategies for handling security to deal with privacy enhancement. However, there are many regulations and legal enforcements regarding privacy protection, especially taking into account the new enforced GDPR in Europe, regulatory strategies deserves special attention. Regulatory strategies refer to governmental rules on the use of personal data or respective self-regulation efforts by industry. Because individual privacy preferences often differ in several ways, it is important to find general principles for privacy protection that fit the most common requirements. Those principles can be used to expand voluntary agreements or regulations enforced by law, but can also serve as important input for the design process of the eVRE system and the subsequent treatment of personal data.

The General Data Protection Regulation, being focused on the protection of the rights and freedoms of individuals, demands from the controllers and processors of personal data that they make the exercise of the eVRE end-users' rights possible. These are specifically laid down in the Articles 12 –22 GDPR (Shadish & Galindo, 2010). The SPECIAL project has summarized the data subjects rights in the GDPR, which includes:

(a) Overview of data subject rights

The data subject rights are:

- Transparent communication (Art. 12 GDPR)
- Information regarding the identity of the controller and the processing itself. This includes the means and purposes of the processing, whereas the law distinguishes between two cases:
    - Personal data are collected from the data subject (Art. 13 GDPR)
    - Personal data have not been obtained from the data subject (Art. 14 GDPR)
- Right of access (Art. 15 GDPR)
- Right to rectification of inaccurate data (Art. 16 GDPR)
- Right to erasure, 'right to be forgotten' (Art. 17 GDPR)

- Right to restriction of processing (Art. 18 GDPR)
- Right to receive a notification from the controller regarding rectification or erasure of personal data or restriction of processing (Art. 19 GDPR)
- Right to data portability (Art. 20 GDPR)
- Right to object (Art. 21 GDPR)
- Protection against automated decision-making, including profiling (Art. 22 GDPR)

(b) Transparent communication and information

For all communication with the data subject, Art. 12 para. 1 GDPR demands that:

> *'The controller shall take appropriate measures to provide any information […] relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child. The information shall be provided in writing, or by other means, including, where appropriate, by electronic means. When requested by the data subject, the information may be provided orally, provided that the identity of the data subject is proven by other means.'*

By following these principles and regional data protection regulation, the eVRE defines when and who can share what privacy information under which conditions in its project data management plan. The eVRE system will help end-users to define the sensitive level of the information they own and accessibility of this information. The eVRE also determines at what scale this information can be shared on the basis of a pre-defined privacy sensitivity level. However, the GDPR guidelines, as well as similar existing regulations, are often in conflict with VRE characteristics. For instance, the principles of purpose limitation and data minimization are conflicting with the active, pervasive, and continuous collection of data in VREs. Privacy Enhancement Technologies which try to enforce existing guidelines are therefore often a trade-off between privacy, and benefit or usability of the VRE system within the context of research purpose.

In the **Annex A**, we have provide an privacy compliance assessment guideline. The questions in the guideline are created by mapping the GDPR terms with the web privacy policy metric proposed by a group of researchers(Miller, Buck, & Tygar, 2012).The end users can use this as quick screening reference to check whether an online service not limited to eVRE complies with the GDPR.

5 Strategies to handle trust aspects

In order to realize a trustworthy eVRE system, different kinds of trust need to be considered. In the eVRE system, data are provided by researchers or data publishers and shared on the eVRE and underlying research infrastructure. We distinguish between technical trust in system components and data, and social trust in other persons who use the eVRE system.

In this section, the implications of trust on people and system will be discussed. These implications will help end-users develop their own strategies to establish their trust on other users of the eVRE, data provided in the eVRE, and eVRE system itself, and be trusted by others.

5.1 'Trust on people' strategies

Trust is also part of the information security governance framework, and thus we can also use the same governance strategies for handling security to deal with trust establishment and enhancement. In addition, there are two specific mechanisms for trust enhancement, namely **Credential-based Trust** and **Reputation-based Trust**. Credential-based computational trust refers to "cryptographic solutions

for establishing trust by obtaining and verifying credentials of an entity" (Könings, Schaub, & Weber, 2016, p. 153). Reputation-based trust "uses the history of an entity's past behaviour or recommendations and experiences about this entity provided by other entities to compute a certain trust level" (Könings et al., 2016, p. 153) . Therefore, we use the entities of the user profile, for instance affiliation, publication history, and number of citations to compute the trust level of the users in the eVRE. In additional, the feedback system of the eVRE is equipped to receive feedback regarding individuals or service providers. This will increase the transparency of the eVRE which will also improve the trust on the system. VRE4EIC will regularly review the trust assessment results and update the trust information regarding the user of the eVRE.

5.2 'Trust on data' strategies

5.2.1 'Trust on data quality' strategies

In deliverable D5.2, we discussed about the implications of trust on data quality for a VRE. In this section, we will focus on the strategies that VRE users should take to increase their trust on data quality if they want to use a dataset they searched, or other's trust on data quality if they want to register a dataset in the VRE. These two objectives have common keys, as the information that a user will look for to check for data quality of a dataset is the same than the one a user should provide when registering a dataset. Thus the following discussion is applicable either to standard users, as to data providers, in its large sense, i.e. any user (human, organisation or automate) that provides dataset through a VRE. Note that the same could also be applied to eRIs, as one of the objective of a VRE is to allow users to search for information coming from various communities, which implies eRIs from these communities.

In deliverable D5.2, we said that several factors have a role on the trust of users regarding data: provenance information or contextualisation of the data for example. The various factors showed that the more the data is described, the better the trust of the users will be. So the first step when thinking about the quality of the data, is to think about the metadata.

In (Davis, 2004), six factors have been described for an organization to build a strategy on data quality:

- Context: the type of data being cleansed and the purposes for which it is used
- Storage: where the data resides
- Dataflow: how the data enters and moves through the organization
- Workflow: how work activities interact with and use the data
- Stewardship: people responsible for managing the data
- Monitoring: processes for regularly validating the data

These six factors can be used to determine the trust level on data quality about a dataset.

The description of the dataset should contain information about the context of the data. The context will help users understand the data, the various cleansing algorithm to be used –or that have been used- and what the data have been harvested for, so the purpose for which the data collection has been done.

The **storage** is another information that reveals the level of trust. This factor includes the physical storage of the dataset, the way it can be accessed, and also information about the data model used. The trust in data quality might be better if the data model used is a standard one, thus having some support and documentation.

Data- and **work-flow** can be considered the same factor, when looking data quality from a user perspective. The objective here is to search for a description of the workflow around the data. Is that raw data directly coming from the collecting equipment? Or is that consolidated data? Are other

versions of the dataset available, corresponding to the other steps of the workflow? Are the various cleansing or consolidation operations well described?

The **stewardship** represents the organization or humans that are responsible for the dataset. Having information about maintainers, creators, hosts, etc. improves the trust in data quality, by being tagged as having a specific responsibility regarding the dataset.

The description of the **monitoring** factor in Davis (2004) applies to live data that should be monitored and checked regularly to be up-to-date. Regarding scientific data, the problem is a bit different in most of the case as data represent the status of the studied objects at a certain point in time. But the monitoring of data can be derived, and users may want to search for related dataset obtained at a different period by applying the same methodology for example, or by having some information about how to reproduce the methodology to check the results.

In this section we studied the various factors that can help a user defining a level of trust for data quality. We mainly focused on the description of the dataset, but other social factors can also play a role in the trust of data quality. For example, the reputation of the organization or persons linked to the dataset will change the level of trust placed in data quality by users. The feedback of other users can also have an impact on the trust on data quality. This social factor is highly dependent on the user's opinion and on his or her specific beliefs.

In the context of the eVRE all the information we talked about can be stored in the catalogue using CERIF, particularly the relations between entities. Obviously, the quality of the metadata depends on the source metadata schema and the level of completion of the metadata about datasets. For example, ISO19139 metadata standard has some elements to deal with the data quality like gmd:dataQualityInfo. This elements has been matched to CERIF so that the information from any source based on the ISO19139 metadata standard will be available in the eVRE through CERIF.

5.2.2 'Trust on metadata' strategies

One of the main objectives of VRE4EIC is *to improve the contextual awareness and interoperability of the metadata across all layers of the resources in the VRE*. Contextual metadata leads to understanding the context in which the scientific data has been created, the semantics, and how the data can be reused. Contextual metadata is crucial to promote data sharing, data use, and to enhance **trust** of researchers in the data and in the data provider.

Can scientific data be trusted? What is its quality? Trust is a term with many definitions and uses, but in many cases establishing trust in scientific data involves analyzing its origins and authenticity, understanding the context in which the data has been created. Scientific data cannot be understood without knowledge about the meaning of the data and the ways and circumstances (used software, equipment etc.) of their creation. This knowledge compeRIses the provenance of the data. So, trust is directly associated to provenance information. The more complete provenance information we have, the higher trust on the data we achieve.

For instance, the catalogue of active faults from a region (from EPOS) can be used for assessing the seismic shaking and tsunami hazard to infrastructure and people in the vicinity, which can be abused by certain commercial bodies. It is important for researchers to know the quality of the data cited or provided by the others and obtain correct vision on the results. This requires the supply of provenance data and contextual metadata. However, the provided information is often very limited and is not sufficient to assess data quality and whether it can be trusted.

Trust is derived from provenance information, and typically is a subjective judgment that depends on context and use. The trust in a scientific dataset is directly related to the trust in the provider - "trusted source" - of that data or the systems used to produce and process that data. The same holds for other types of resources shared through the eVRE, such as services, tools, or software. By "trusted source",

we mean that someone gives me information he himself trusts because of further knowledge, and I trust that it is his best knowledge, regardless whether it turns out to be wrong later.

If I get a catalogue of earthquakes with fault plane solutions, I have no idea if it is a correct or not. I would not assume that it is wrong, if I can trace its provenance to the producer, the methods and its publication record.

The provenance chain should go back to primary evidence. For this catalogue, I should be able to trace:

- the complex analyses and other data products and/or equipment used,
- the publication record and the publishing body,
- the producer and the human curator of the data.

The reference to a human in the metadata is crucial in the provenance chain and further supports trust. In particular, for data in the eVRE, which is not (easily) verifiable in itself as a simple mathematical proof, the connection to a human curator is necessary and provides trust in the resource.

Quite often, special equipment is involved in the production of scientific datasets and faults might be associated with them, influencing the produced data. Again, trust is established as long as detailed documentation is available.

An example that shows the importance of following the history of a device is presented in [Doerr2011]. *Within weeks of the Hubble Space Telescope launch, the returned images showed that there was a serious problem with the optical system. Nonetheless, during the first three years of the Hubble mission, before the optical corrections, the telescope still carried out a large number of productive observations. The error was well characterized and stable, enabling astronomers to optimize the results obtained using sophisticated image processing techniques such as deconvolution. In December 1993 the First Servicing Mission restored Hubble's Vision. Once Hubble received its corrective "eyeglasses," it began seeing more clearly.*

Thus we suggest that contextual metadata should describe not only the actual scientific results but, firstly, the resources at hand for making research, how they are disposed and what condition they are in, who uses and cares for their quality and validity. This will provide means to track if data is reliable and well maintained.

In this direction, CERIF is a well suited common reference model since it models datasets, software, services, projects and actors as well as, most importantly the contextual relations that exist between them. This conceptual model provides the means to represent resources in an accurate and manageable way. Moreover, the Semantic Layer in CERIF provides a classification system and allows for the efficient and meaningful management of controlled vocabularies by the communities.

The VRE4EIC architecture has at its centre a CERIF metadata catalogue that documents what actors and resources exist, how they are related and who has provided them. For each resource basic provenance information is recorded when it is ingested in the catalogue. This aims to meet the basic trust criteria for the scientific community, giving the provenance of knowledge for data and software.

Related work is done in the EU HORIZON2020 project PARTHENOS (http://www.parthenos-project.eu/). The project proposes the Parthenos Entities (PE), an ontological model and RDF schema to encode data of use in supporting the activities and aims of research infrastructures to pool and connect services, software, datasets and to enable users of such services to reach the actors and understand the knowledge generation processes which generated the offered datasets, thus meeting the basic trust criteria. Moreover, the project proposes a minimal set of metadata that should be collected with regards to the entities referred to in the Parthenos Entities model for the provisioning of the Parthenos Registry.

5.3 'Trust on eVRE system' strategies

As described in Deliverable 3.3, the eVRE adopted the microservices approach to implement the eVRE system. In the eVRE system, every building block is implemented as a Microservice (i.e. a standalone component) that publishes its functionalities via REST APIs.

Interactions between external software agents and eVRE are mediated by eVRE Web Services and occurs using secure protocols, in particular HTTPS, implementing asymmetric cryptography scheme.

The eVRE microservices, in order to coordinate their behaviour, cooperate with each other by exchanging *messages*. A message is produced by a microservice when a significant event occurs and is asynchronously consumed by all building blocks (or microservices) that are interested in that event.

The eVRE building blocks needs to communicate with each other in a secure manner: the infrastructure exchanging messages between building blocks must be reliable and secure. This means not only that the software used as communication channel must provide security features, but also that messages content need to be verifiable and traceable. In the next section we'll explain, with an example, the approach adopted to implement a trusted communication infrastructure in eVRE.

One of the challenges in a microservice-based architecture is identity propagation. After the authentication, the identity of the authorized user needs to be propagated to all microservices in a trusted way. As explained in Deliverable 3.3 eVRE implements a token-based authentication: users (or software agents) enter their credentials in order to obtain a token which allows them to access the functionalities of eVRE. The token is created by a specific eVRE building block, the AAAI, and distributed to other building blocks in the system (beside being sent back to the user client).

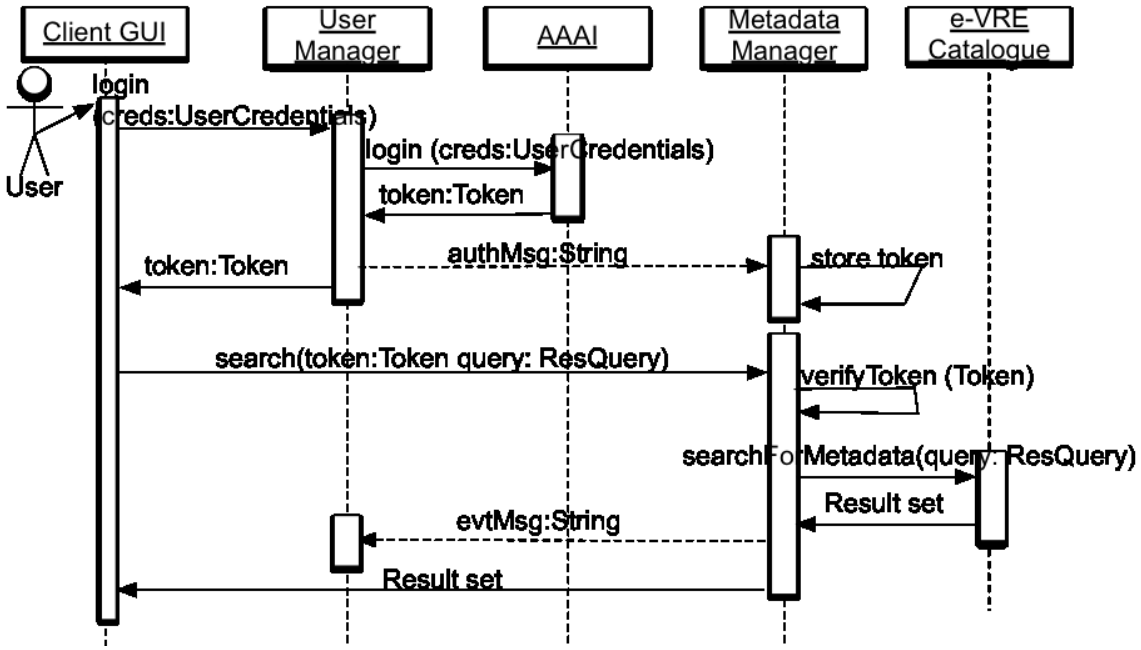

Figure 1 UML sequence diagram: a user authenticates and executes a query on the catalogue

The figure above shows a sequence diagram that describes how eVRE building blocks interacts in the use-case of a user that authenticates and executes a query on the catalogue:

1. during the authentication the AAAI creates a token and sends it synchronously to the client and asynchronously to the Metadata Service (and to any other eVRE building Blocks).
2. the Metadata Service stores locally the token.

3. at this point the user is logged in and he/she will include the token in the query request
4. the Metadata Service checks the token for validity and executes the query in the catalogue and before sending the result generates a message (evtMsg) with information about the operation

The main security/trust issues in this use case are:

- the Metadata Manager needs to know that the token it receives asynchronously in the step 1 has been created by a software component belonging to eVRE and having the authority of creating it, and that it has not been tampered with.
- the Microservices consuming the evtMsg message generated by the Metadata Service (for instance the Node Service that registers system operations in a log file) need to be sure that the message is actually created by the Metadata Service

The solution adopted to solve these issues has been to sign and encrypt token and messages exchanged by microservices among them. In the use case above the AAAI component creates the token, digitally signs it and encrypts it, when the Metadata Service receives the token it decrypts the token and checks the claim, if the signature is correct the token is locally stored for being used in clients authentication. In the same way the evtMsg is signed and encrypted before being published by Metadata Services, the consumers will check the signature before actually consuming the message. In essence every microservice in eVRE is able both to sign/encrypt and to decrypt/check messages to be sure that eVRE internal interactions occur in a trusted environment.

We decided to encode our token and messages using the JSON Web Token (JWT) standard. The JWT is "an open standard […] that defines a compact and self-contained way for securely transmitting information between parties as a JSON object. This information can be verified and trusted because it is digitally signed. JWTs can be signed using a secret (with the HMAC algorithm) or a public/private key pair using RSA" (JWT, 2018) .

In the current release of eVRE, messages and tokens are signed using a secret shared among microservices, we will adopt the more reliable public/private key pairs scheme in the future releases.

6 Conclusions

Security, privacy and trust can be enhanced not only by the proper design and operation of the eVRE, but also by appropriate use of the eVRE by the end-users. Therefore, the end users of the eVRE should understand their rights regarding the personal data and proactively undertake strategies with respect to these three aspects. This deliverable aims to educate the end users and help them understand security, privacy and trust strategies from different aspects.

We took the requirements regarding security, privacy and trust issues from WP2 and analyzed the gaps between the requirements and currently implemented solutions by other projects. Then we provided recommendations for the potential end users of eVRE to deal with the issues from both the perspective of the technological solution and the perspective of information governance.

The described strategies reflect the solutions and suggestions made by the VRE4EIC consortium regarding security, privacy and trust issues.

Security strategies

- The eVRE reference architecture supports federated authentication and authorization. The end users are encouraged to log onto the eVRE using the credentials of their own organization by using their organization's Identity Provider through federated service such as EduGAIN.
- Multi-factor authentication is recommended for the eVRE when a user accesses sensitive data.
- The end users should adapt their daily behaviors to information security in daily practices.

Privacy strategies

- The end users should always check that techniques such as authentication, authorization, encryption and anonymization, purpose based data storage & data access are used by the online service provider.
- The end users may check whether they can easily get access to the privacy policy of the eVRE and connected datasets, software and services.
- The end users may check whether the use of their personal data is only limited to the delivery of the services provided by eVRE and other connected eRIs and VREs.
- The end users may check how eVRE collects and stores the personal data concerning him or her. They can ask for the full records of the data stored by eVRE and check whether limited data is collected.
- The end users may frequently check the accuracy of the data concerning themselves on the eVRE.
- The end users may check the storage of the personal data is limited to a certain time span.
- The end users may frequently check the storage of the personal data is safeguarded and will be notified when there is data breaching incident.
- The end users should understand who is the responsible body for protecting their personal data and to which authority body they can refer when there is privacy dispute .

Trust strategies

- The end users should check the entities of the user profile to see the research history of data publishers and use feedback system to provide information concerning individual activities on the eVRE system.
- The end users may track source metadata schema and check the level of completion of the metadata about datasets defined in the CERIF.
- The users group of service providers may use Microservices approach to implement their software and services on the eVRE.

Upon acceptance of this strategy document, this document will be made publicly available together with training materials in the Annex B and especially distributed to potential user groups. This work related to dissemination has been described in the deliverable of WP7. The awareness of security, privacy and trust challenges will help the end users understand the complexity of these issues and encourage them to develop their own strategies to handle these issues.

References

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free

movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016).

JWT. (2018). Introduction to JSON Web Tokens. Retrieved from https://jwt.io/introduction/

Könings, B., Schaub, F., & Weber, M. (2016). Privacy and Trust in Ambient Intelligent Environments. In S. Ultes, F. Nothdurft, T. Heinroth, & W. Minker (Eds.), *Next Generation Intelligent Environments: Ambient Adaptive Systems* (pp. 133-164). Cham: Springer International Publishing.

Leino-Kilpi, H., Välimäki, M., Dassen, T., Gasull, M., Lemonidou, C., Scott, A., & Arndt, M. (2001). Privacy: a review of the literature. *International Journal of Nursing Studies, 38*(6), 663-671. doi:http://dx.doi.org/10.1016/S0020-7489(00)00111-5

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review, 20*(3), 709-734. doi:10.2307/258792

Miller, B., Buck, K., & Tygar, J. D. (2012, 10-12 Dec. 2012). *Systematic analysis and evaluation of web privacy policies and implementations.* Paper presented at the 2012 International Conference for Internet Technology and Secured Transactions.

Montjoye, Y.-A. D., Radaelli, L., Singh, V. K., & Pentland, A. S. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science, 347*(6221), 536-539. doi:10.1126/science.1256297

Nuffelen, B. B. V. (2016). *Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance Technical Requirements V1*. Retrieved from https://www.specialprivacy.eu/images/documents/SPECIAL_D1.4_M8_V1.0.pdf

Panteli, N., & Sockalingam, S. (2005). Trust and conflict within virtual inter-organizational alliances: a framework for facilitating knowledge sharing. *Decision Support Systems, 39*(4), 599-617. doi:http://dx.doi.org/10.1016/j.dss.2004.03.003

Pearson, S. (2009). *Taking account of privacy when designing cloud computing services*. Paper presented at the Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust1. *Journal of Personality, 35*(4), 651-665. doi:10.1111/j.1467-6494.1967.tb01454.x

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*(5), 443-452. doi:10.1037/h0031464

Schlehahn, E., & Zwingelberg, H. (2017). *Legal requirements for a privacy-enhancing Big Data V1*. Retrieved from https://www.specialprivacy.eu/images/documents/SPECIAL_D1.2_M6_V1.0.pdf

Shadish, W. R., & Galindo, R. (2010). Quasi-Experimental Design. In *The Corsini Encyclopedia of Psychology*: John Wiley & Sons, Inc.

Siponen, M. T., & Oinas-Kukkonen, H. (2007). A review of information security issues and respective research contributions. *SIGMIS Database, 38*(1), 60-80. doi:10.1145/1216218.1216224

Sun, D., Chang, G., Sun, L., & Wang, X. (2011). Surveying and Analyzing Security, Privacy and Trust Issues in Cloud Computing Environments. *Procedia Engineering, 15*, 2852-2856. doi:http://dx.doi.org/10.1016/j.proeng.2011.08.537

Veiga, A. D., & Eloff, J. H. (2007). An information security governance framework. *Information Systems Management, 24*(4), 361-372.

Veiga, A. D., & Eloff, J. H. P. (2007). An Information Security Governance Framework. *Information Systems Management, 24*(4), 361-372. doi:10.1080/10580530701586136

Annexes

Annex A: Privacy policy compliance assessment guideline

This privacy policy compliance assessment guideline provide a reference for the end users to conduct a quick screening concerning the privacy policy stated by an online service provider.

1. Completeness and content of the privacy policies

1.1. Does the system lawfully collect and process personal data?

1.2. Does the system offer a single, comprehensive document referred to as the "privacy policies" which describes the privacy practices of the system?

1.3. Does the system homepage use the word "privacy" in a link to either the privacy statement or a dedicated section of the system which contains any privacy policies?

1.4. Does the main privacy document begin with a summary of the entire policies and a section of links which help the reader to navigate the content?

1.5. Do the privacy policies include definitions that use layman terms to explain privacy relevant aspects of any technical or legal terms?

1.6. Do the privacy policies indicate the last date they were updated?

1.7. Does the system offer to notify users of changes to the privacy policies?

2. Privacy policies scope

2.1. Do the privacy policies contain clear explanation of the purpose of collecting personal data?

2.2. Do the privacy policies contain an unambiguous and clear explanation of the personal data collected by the system limited to realization of the purpose?

2.3. Do the privacy policies provide an exhaustive list of methods the system uses to collect information, such as HTTP cookies, web bugs and Flash Locally Stored Objects (LSOs), etc.?

2.4. Do the privacy policies state how collected information will be used?

2.5. Do the privacy policies state which information will be shared with third parties, and for what purposes?

2.6. Do the privacy policies state any measures taken to protect data in transmission and in storage?

2.7. Do the privacy policies state how long data will be stored after collection by the system or deletion by the user?

2.8. Do the privacy policies contain an email address and a postal addresses for contacting the data protection officer of the organization?

3. Client Storage Practice and Policy

3.1. Does the system use HTTP cookies?

3.2. Does the system use HTML5 Local Storage or other non-cookie browser storage?

3.3. Does the system use Flash LSOs?

3.4. Does the system become significantly less useful if HTTP cookies or other browser storage is disabled?

3.5. If the system uses Flash LSOs, does the system become significantly less useful if Flash is disabled?

3.6. If the system uses cookies, do the privacy policies accurately describe the use of HTTP cookies?

3.7. If the system uses Flash LSOs or browser storage other than cookies, do the privacy policies accurately describe these practices?

3.8. Do the privacy policies accurately explain the extent to which the user can avoid client storage based tracking and guide the user through that process?

4. Third parity use practice and policies

4.1. Does the system use any resources or services provided by third parties, such as cloud computing, data analytics, web beacons or Javascript libraries?

4.2. Does the system generate any requests to third parties which include a persistent cookie?

4.3. Does the system continue to function if browser extensions blocking third party tracking are used?

4.4. Does the system make any requests to third parties which are not blocked by browser extensions blocking third party tracking, such as Javascript libraries?

4.5. Does the organization associated with the system offer any web plug-ins designed to be used on third party sites?

4.6. Do the privacy policies guarantee the user the right to opt-out of non-critical information sharing with third parties and explain any technical measures necessary to do so, such as opt-out options?

5. Data processing Practice and Policy

5.1. Does the privacy policies allow targeting either advertising or content to the user?

5.2. Does the entire system support HTTPS?

5.3. Does the entire system use HTTPS by default?

5.4. Does the entire system store the consent by the users regarding the collection and processing the personal data?

5.5. Does the system provide features for customizing the access or restriction to the data which is stored about the user?

5.6. Does the system provide features for modifying or deleting data which is stored about the user?

5.7. Do the privacy policies guarantee the user rights to access, modify or delete data about the user and provide guidance to do so?

5.8. Do the privacy policies allow data collected from or deleted by the user to be retained longer than is legally necessary?

5.9. Do the privacy policies allow users to easily access the structured, machine-readable, personal data concerning him or her and transmit to another system?

5.10. Do the privacy policies allow users to object the results of automated individual decision-making, including profiling?

5.11. Does the system use any techniques to ensure the secrecy of the personal data?

5.12. Does the system maintain a record of data processing activities regarding users' personal data, including the content stated in the GDPR?

6. Personal account information

6.1. Does the system request the user's phone number for an activity where the user's primary goal is not to receive a phone call or use two factor authentication?

6.2. Does the system request the user's physical address for an activity where the user's primary goal is not to receive physical mail?

6.3. Do the privacy documents allow any contact information to be used for marketing purposes?

6.4. Do the privacy documents allow third parties to use any contact information for marketing purposes?

6.5. Do the privacy documents allow the user to opt out of non-critical mailings?

6.6. Do the privacy documents state what user information will be publicly visible, either by default or at minimum?

6.7. If the organization offers plug-ins for use by third party systems, do the privacy documents guarantee that any data obtained from plug-ins will not be linked to data entered at the organization's system?

6.8. If the system facilitates sharing data with other users and the public, does the system provide mechanisms for defining access policies at the granularity of individual users as opposed to differentiating between groups of users or using the same policies for all users?

6.9. If the system facilitates sharing data with other users and the public, does the system allow different access policies to be associated with different types of data, or preferably with different instances of the same type of data?

7. Usability of privacy setting customization

7.1. Does the account setup process include selecting privacy settings?

7.2. Do all privacy settings default to the most private setting available?

7.3. Does the system have a section dedicated to privacy settings and containing all privacy settings applicable to the whole system?

7.4. Does either the system homepage, a menu on the site homepage, or the account settings page contain a link to the privacy settings page?

7.5. Does the system describe each privacy option in detail?

7.6. Are descriptions of what each privacy setting entails directly accessible from where the user would alter the privacy settings?

7.7. Does the system provide active tips or guidance the first time each privacy setting or privacy sensitive feature is used or adjusted?

7.8. Does changing an aspect of the privacy settings ever require more knowledge of computer security than a lay person has?

7.9. Does the system provide any information about computer security necessary to help the user select privacy settings?

8. Privacy breaching incident response

8.1. Do the privacy policies clear define the actions should be undertaken when a privacy breaching incident happens?

8.2. Does the privacy policies clear state that the organization will notify the users when the personal concerning him or her is leaked?

Annex B: Training materials

The training materials are developed in Work Package 6. The training materials can be accessed on the VRE4EIC website through this link: https://www.vre4eic.eu/tutorials.