VRE4EIC

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

Deliverable D5.2

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

Document version: Final

VRE4EIC DELIVERABLE

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

Mr Philippe Rohou t: +33 4 97 15 53 06 f: +33 4 92 38 78 22 e: philippe.rohou@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:	14.01.2015
Document	
Period covered:	M1-M18
Deliverable number:	D5.2
Deliverable title	Implications for the VRE end-users to handle security, privacy and trust issues – first version
Contractual Date of Delivery:	31.03.2017
Actual Date of Delivery:	31.03.2017
Editor (s):	Yi Yin (TU Delft), Anneke Zuiderwijk (TU Delft)
Author (s):	Laura Hollink (CWI), Cesare Concordia (CNR), Kevin Ashley (EuroCRIS), Angus Whyte (EuroCRIS)
Reviewer (s):	Valerie Brasse (EuroCRIS), Carlo Meghini (CNR)
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):	24

What is VRE4EIC?

VRE4EIC develops a reference architecture and software components for VREs (Virtual Research Environments). This e-VRE bridges across existing e-RIs (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 e-VRE provides a comfortable homogeneous interface for users by virtualising access to the heterogeneous datasets, software services, resources of the e-RIs 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

This document contains a description of the VRE4EIC project work and findings.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any responsibility for actions that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the VRE4EIC consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 28 Member States of the Union. It is based on the European Communities and the Member States cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors (http://europa.eu/).

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 e-VRE being developed by the VRE4EIC project, the security, privacy and trust requirements significantly vary amongst the potential end-users from various research domains and public. In deliverable D5.1, the VRE4EIC project has defined the strategies handling potential issues and risks regarding security, privacy and trust aspects. However, these strategies are mainly produced to help the technical developers better design the e-VRE and choose appropriate technologies for the implementation of the e-VRE. We assume that most of the potential end-users will have limited ICT knowledge so that they may 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, this deliverable clarifies these strategies in the early stage of e-VRE development. The user is a global actor representing any user accessing the e-VRE (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. After implementing an e-VRE prototype that will be tested by user groups, a second version of this deliverable will be completed in Month 33 (D5.4). 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 periodic result of Task 5.2, which identifies issues related to security, privacy, and trust for the use of the VRE, and which defines a strategy to handle these issues. The strategy will provide the end-users of the VRE insight in the various security, privacy and trust issues that they may face when they use the VRE. Since those issues may considerably hinder the use of VREs for excellent research, the strategy is expected to stimulate VRE uptake.

More specifically, the operations 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 WP6 project partners and define:
  - how VRE4EIC users can deal with security issues;
  - how VRE4EIC users can deal with privacy issues;
  - how VRE4EIC users can 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 training materials towards 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 in M36;
- Make the strategies publicly available so that users of other VRE related initiatives and VREs can also benefit.

This deliverable describes the implications of technical solutions and data governance strategies regarding the security, privacy and trust aspects implemented in the VRE4EIC project.

1.1 Potential security issues for users

Security is a rather wide concept; however we refer to information security. There are many security issues around the development and the use of the e-VRE. It is rather difficult to provide a complete detailed list of information security issues, but key security issues are discussed here. Sipoen et al. 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.

Potential Security-related issues		Potential risks
Access to the e-VRE	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 crisis events
secure development of e-VRE	Security requirements are not properly collected, misinterpreted or missing	Security requirements cannot be properly reflected in the development of e-VRE
	Security requirements cannot be fulfilled by certain technological solutions.	The security cannot be guaranteed, which damages service providers' reputations and reduce user trust

Table 1-1 Potential trust issues for end-users of the e-VRE

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). Table 1-2 shows the potential privacy-related issues for end-users.

Potential privacy issues	Potential risks
Researchers are forced or persuaded to be tracked or to provide personal information against their will (Pearson, 2009)	Individual information may be stolen, used nefariously or resold unauthorized
Big data analytics can be used to identify individuals (de 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 information 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 The government 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. The government might be blamed and lose credibility due to lack of proper governance

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

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.

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 hinders the interoperability of data
	Secure storage of data	Data might be lost, stolen or leaked
	Data Confidentiality and auditability	The source changes of data can be 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 e-VRE

1.4 Structure of the document

First of all, this document provides an overview of current solutions with regards to security, privacy and trust issues in the existing e-RIs are illustrated. Section 2 recalls the requirements clarified in D5.1 on these three aspects. Section 4-6 describes the implications of security, privacy and trust strategies being developed by this project. Finally, the potential influence on the architecture design and related adjustment strategies are presented.

2 Current solutions in e-RIs and related initiatives

In the context of Task T2.1 and T5.1 we have analyzed the current solutions in e-RIs and related initiatives. Eight e-RIs were analyzed through interviewing the researchers involved in the development or operation of these e-RIs, including ICOS, EURO-ARGO, EPOS, ELIXIR, LIFEWATCH, CESSDA, ESS, CLARIAH. The results have been reported in the deliverable D5.1 of the VRE4EIC project. Hereby, we summarize the solutions which are currently being used or will be potentially used by other VREs.

The e-RI characterizations showed that one of the driving use cases - EPOS - is converging towards a technical solution that tackles some of the main issues of the AAAI.

Such a solution is already used in parallel projects (i.e. EPOS, ENVRI and is supported by other initiatives, namely EGI[1], EUDAT, AARC. AARC is at a different level than the two others though, as AARC wishes to provide an extension to IDPs such as EduGAIN, supporting user attributes and X.509 certificate based services. AAAI solutions basically enable a user to have single authentication to access all resources, and, under specific circumstances (depending on the protocol) enable a VRE to be delegated to act on user behalf.

Such a solution is the UNITY software,[2] which facilitates the establishment of solution for identity, federation and inter-federation management. Or, looking from a different perspective, it is an extremely flexible authentication service. Unity is a service that enables login to a web service using various protocols. It supports the LDAP protocol (e.g. OpenLDAP or Active Directory) and authentication can be performed with various identity providers, amongst others the EduGAIN federation previously mentioned. Unity is open source software licensed under the BSD licence[3].

In order to connect existing AAAI approaches from e-RIs into one VRE ecosystem, an AAAI hub is needed which will assure interoperability between existing technologies. HUB technologies passing logins, passwords, and such are deprecated and should no longer be used. Instead attribute based solutions are advised. An IDP should return a digitally signed document that states the identity of the user. Once a user is authenticated within the infrastructure all the authorizations can be done using the attributes only. The term attribute is used here to describe properties of the user, e.g. his/her name, email, affinity, role. A set of the VRE specific attributes will have to be defined on the hub.

Current IDPs include: LDAP, OpenID, X.509 certificates, EduGAIN (different from the others, it is more a political project using SAML as its core technology). In the EPOS AAAI Hub, and from the technological point of view, UNITY provides plug-ins for many IDPs (not just four mentioned above). For instance, a UNITY instance registered in EduGAIN is sufficient to enable EduGAIN authentication from any service attached to UNITY. Attribute management is flexible from the administrator point of view. REST APIs are also available. A solution like this, far from being the panacea, can indeed help to tackle at least the main problem, that is to say the federated identity management: in practice it will enable any user with almost any credentials (at least all those supported by UNITY) to access the VRE.

What still remains an open challenge is the delegation. This topic has been discussed in EPOS, EGI, EUDAT, AARC, and still remains a work in progress.

[1] https://access.egi.eu/

[2] http://www.unity-idm.eu/

[3] http://www.linfo.org/bsdlicense.html

3 Requirements of security, privacy and trust

D2.1 describes the steps that have been followed to elicit requirements for the e-VRE. The requirements regarding security, privacy and trust have been further described in D5.1, "A strategy for the VRE4EIC project to handle security, privacy and trust issues". We expanded the list of requirements related to these three aspects by adding the requirements which also implied these aspects. These requirements are used as input for the design of the AAAI component of the e-VRE. 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 e-VRE.

3.1 Requirements on security

The requirements elicitation process has resulted in a number of concrete functionalities that improve the security of the e-VRE (Table 3-1).

Requirement ID	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 3-1: Identified requirements related to security.

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-RIs connect to the VRE and (3) allow unrestricted access to open data. In the latter case, the e-RIs should be warned for 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 at the eight currently characterized e-RIs, while it was identified as a requirement. Individual e-RIs 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-RIs. 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.

3.2 Requirements on privacy

The VRE should guarantee the privacy of both users of the e-VRE 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 CERIF.

Differential privacy: The fact that the VRE bridges across several e-RIs 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 3.2.

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

Table 3-2: Identified requirements related to privacy.

3.3 Requirements on trust

The elicited requirements (see Table 3-3) 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 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 3.1 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 3-3: Identified requirements related to trust.

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.

4 Implications of security

In order to help VRE4EIC users deal with security issues when using the e-VRE system, the VRE4EIC consortium should help making users aware of information security issues and understand the reasons for the technological solutions offered in the e-VRE. Since the e-VRE system is built on many e-RIs and connected with many other VREs and e-RIs, users need to understand the complexity of e-VRE and adapt their using behavior to the characteristics of the e-VRE and develop their own security strategies when using the e-VRE. 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 (Veiga & 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.

4.1 Technological implications

The e-VRE reference architecture has been designed such that the AAAI component consists of three parts: a Unity server, a Role Based Access Control (RBAC) component and a CERIF database. Login to the e-RIs and access to the e-RI resources goes through the AAAI component of the VRE.

Single sign-on using Unity login

An advantage for end-users is that this design enables single sign-on for access to multiple associated e-RIs. It does mean that end-users need an account on the Unity server. In most cases, this will be federated to the research institutes to which they are affiliated. Users who are not part of a research institute, or who do not wish to register with those credentials, can create an account on the Unity server directly. This, however, makes the validation of the user-provided attributes more difficult.

AAAI component as a secure environment

A second implication is that users need to trust that the VRE AAAI component is a secure environment and that the information in it (e.g. the roles in the CERIF database) is correct. An advantage is that it limits the amount of user information that needs to be collected at the e-RI level, thus reducing security risks at that level. In the implementation of the AAAI component, there is room for several variants regarding the amount of information that is disclosed to the e-RI by the VRE. At least, the VRE provides the e-RI with a verdict about whether or not the end-user is authorised for a certain resource. In some cases, it is desirable that the e-RI receives additional information, such as the real identity and/or roles of the end-user. In this case, users may need to give explicit permission for disclosure of these attributes to the e-RI.

4.2 Governance implications

Besides the technological solutions within the e-VRE, governance strategies regarding security are also important. The governance strategies imply not only the governance on the development and operation of the e-VRE from the VRE4EIC research group, but also self-management of end users' using behavior. Therefore, we will explain the governance strategies on handling security from the role of the system developer's perspective. In addition, we will provide recommendations for end-users to deal with potential information security issues.

Security Governance by e-VRE

To achieve effective information security management, the VRE4EIC consortium defines strategic governance, managerial and operational governance which contains a series of activities. From security governance model proposed by Veiga and Eloff, the information security governance framework generally consists of:

- **Strategic governance**:
    - **Leadership and governance**:
        - Executive level sponsorship such as reporting structure, authority, responsibilities, policy enforcement,
        - A comprehensive information security strategy explicitly linked with project and IT objectives,
        - IT governance which defines the way of controlling the use of technologies to protect information security,
        - Risk assessment and mitigation
        - Metrics to define the information security level
- **Managerial and Operational governance:**
    - **Security management and organization:**
        - Program organization which defines the information security organizational design, composition and reporting structure;
        - Legal and regulatory which refers to the legislation related to information security,
    - **Security polices:**
        - Security policies, procedures, standards and guidelines which talk about the value of information protected and delivered and which directs the behavior of the stakeholders within the system. These security policies have been developed in the project deliverable D5.1.
    - **Security programs management:**
        - Monitoring, auditing and compliance management which refer to the management of security programs in order to measure and enforce the technology and users' behaviors and to ensure continued evaluation and update of security policies, standards, procedures and risks,
    - **User security management:**
        - User awareness regarding information security,
        - Education and training
        - Ethical conduct,
        - Trust and privacy

This framework in turn provides the basis for the development of a cost-effective information security programme that supports the organization's goals and provides an acceptable level of predictability for operations by limiting the impacts of adverse events. The overall objective of the programme is to provide assurance that information assets are given a level of protection commensurate with their value or the risk their compromise poses to the organization.

During the development and operation of e-VRE, the leadership of the information security governance is defined as shown in Figure 4-2. The consortium defines the VRE4EIC development strategy which includes the project objectives. The executive management of the VRE4EIC 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.

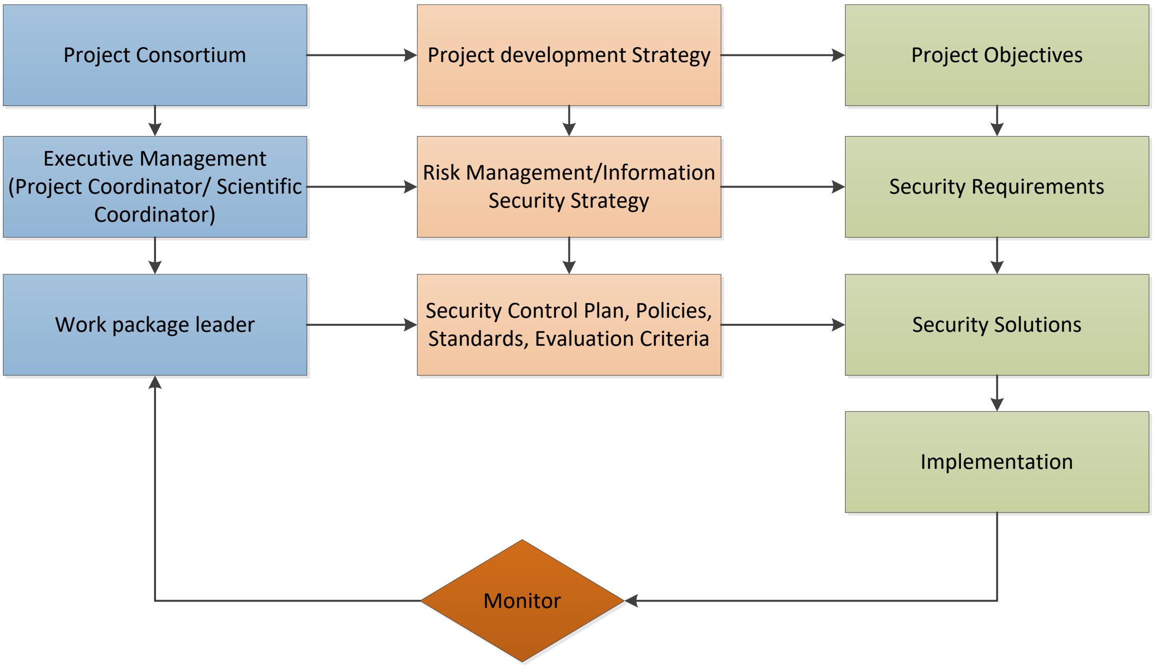
Project Consortium
Project development Strategy
Project Objectives
Executive Management (Project Coordinator/ Scientific Coordinator)
Risk Management/Information Security Strategy
Security Requirements
Work package leader
Security Control Plan, Policies, Standards, Evaluation Criteria
Security Solutions
Implementation
Monitor

Figure 4-1 Organizational management structure on information security

Security Governance by end-users

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.

5 Implications of privacy

In general, privacy protection mechanisms can be divided into the four main categories of regulatory strategies, policy matching, prevention and control, and detection (Könings, Schaub, & Weber, 2016). The regulatory strategy belongs to governance solutions while the last three mechanisms belong to privacy enhancement technologies, or technological solutions.

The information security governance framework shown in the section 4 shows that privacy enhancement is part of information security management. In order to help VRE4EIC users deal with privacy issues when using the e-VRE system, the VRE4EIC consortium should help users understand the technological solutions offered in the e-VRE systems and regulation on privacy, and on European regulation in particular.

5.1 Technological implications

As stated in D5.1, "The e-VRE should guarantee the protection of both personal research data that is accessed via the e-VRE and personal data about the users of the e-VRE 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 the Section 4 regarding security.

Regarding the personal data of users of the e-VRE, a shift will take place from the e-RI level to the e-VRE level. The e-VRE architecture means that less information about a person is stored at the e-RI level. In principle, the e-RIs only need to have opaque handles for each end-user that cannot be traced back to their identity. This means there is less concern about the privacy aspects of e.g. access logs of the e-RI resources. The architecture, however, allows for cases where the e-RI will receive more information about the end-users. The fact that these differences exist may be unclear for end-users, as they are partly obfuscated by the single sign-on principle. The burden on users to learn where their data is stored is therefore increased.

Users should be able to trust that the e-VRE keeps their credentials secure. The RBAC mechanism means that a considerable amount of information about a person is present at the e-VRE-level (e.g. project or institution membership, or level of seniority).

5.2 Governance implications

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 General Data Protection Regulation[4] (GDPR) in Europe in 2016, regulatory strategies deserves special attention. Regulatory strategies refer to governmental rules on the use of personal information 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 e-VRE system and the subsequent treatment of personal information.

The new GDPR updates and modernizes the principles in the EC 95 Directive (Bird&Bird, 2017) , including:

[4] http://www.eugdpr.org/

- **Lawful, fair and transparent processing** – emphasizing transparency for data subjects,
- **Purpose limitation** – having a lawful and legitimate purpose for processing the information in the first place,
- **Data minimization** – making sure data is adequate, relevant and limited and organizations are sufficiently capturing the minimum amount of data needed to fulfil the specified purpose,
- **Accurate and up-to-date processing** – requiring data controllers to make sure information remains accurate, valid and fit for purpose,
- **Limitation of storage in a form that permits identification** – discouraging unnecessary data redundancy and replication,
- **Confidential and secure** – protecting the integrity and privacy of data by making sure its secure (which extends to IT systems, paper records and physical security),
- **Accountability and liability** – demonstrating compliance,

By following these principles and regional data protection regulation, the e-VRE will define when and who can share what privacy information under which conditions. The e-VRE system will help end-users to define the sensitive level of the information they own and accessibility of this information. The e-VRE will also automatically decide 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. PETs which try to enforce existing guidelines are therefore often a trade-off between privacy, and benefit or usability of the VRE system.

6 Implications of trust

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

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

6.1 ’Trust on people’ implications

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 et al., 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 e-VRE. In additional, the feedback system of the e-VRE is equipped to receive feedback regarding individuals or service providers. This will increase the transparency of the e-VRE 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 e-VRE.

6.2 ‘Trust on data’ implications

6.2.1 ‘Trust on data quality’ implications

D5.1 describes a number of techniques in section 3 for promoting trust in data or systems that may be otherwise unfamiliar to the end user. They include the provision of provenance information and, in some cases, explicit assertions about the measurement of particular quality attributes in metadata fields. Work such as Wang and Strong (1996) tells us that data quality is a many-dimensioned property, and hence that what is high quality to one consumer may be low quality to another (Wang & Strong, 1996). Thus, the strategy of making directly visible to the consumer assertions about a number of properties of data that are related to quality has the best chance of satisfying the trust issues in data quality.

For some users, accurate provenance information will be sufficient. Those users place implicit trust in the quality of material that comes from certain sources and require no other information to be confident of data quality. Other users will require more information to be reassured, either because they are not familiar with the sources documented in the provenance information or because they do not have confidence in those sources. These other users will seek information about the data’s timeliness, or its coverage, accuracy or precision. They may wish to know what processes (if any) have been applied to the data or they may wish to know who else has used it and for what purpose. The mechanisms proposed by D5.1 permit all these attributes of a dataset or system to be asserted and more. However, since they cannot be required, the user’s willingness to trust some data sources will be greater than that of others. There are also skills implications for the end users of data where some of these attributes may be expressed via metadata which is not necessarily visible to the casual user.

Provenance information, for example, is often made visible and is generally comprehended by the typical research dataset user. But technical metadata which makes quality assertions using W3C standards for data quality is not so well known and may not be so well understood by the average user without training or documentation which explains how to make use of it to judge quality.

6.2.2 'Trust on metadata' implications

Section 6.2.1 has described how metadata can be used to establish trust in data quality. This has implications for what metadata is ideally present in the e-RIs and how it is presented to end users in the VRE. Users also need to be able to manipulate this metadata or, ideally, to have it manipulated for them in accordance with their instructions, whenever they are themselves generating data products or other outputs using data within the VRE. At minimum they will want to be able to cite their data sources. When many data sources are used, this can be an error-prone task if carried out manually, but the VRE should be capable of doing this automatically for the data user. This will increase the trust of that user in the accuracy and completeness of the citation metadata and the trust of others reading about that user's research. Users will also want to be reassured that their intended use of the data is permitted by the agreements or licences which pertain to the source data and systems that they are using. The VRE will be drawing on information from multiple e-RIs which do not necessarily all apply the same licences to data products. It is important that such information, which has implications for permitted use, is expressed in a consistent fashion to end users. Potentially the VRE could warn end users if they attempt to combine data within the VRE which has incompatible licensing conditions attached to it.

As well as automatically generating citation information for the underlying data products consumed by a user in the VRE, the potential exists to automatically generate assertions about matters such as data accuracy, completeness, precision etc. using the metadata indicators described in 6.2.1. This is a trivial task if a user is utilising a single data source in the VRE and one might say that it isn't worth the effort to have the system assist with this task. But when multiple data sources from multiple e-RIs are being combined for analysis within the VRE, then constructing a metadata expression of the quality of the resultant derived dataset is a non-trivial task even for the expert end-user. A degree of assistance with this by the VRE would reduce effort for the end user and increase trust in results.

There are trust implications in the heterogeneity of e-RIs and metadata about the research resources they provide to the VRE, as this is likely to lead to inconsistent granularity and homonymous use of basic terms such as 'research data'. They might be mitigated by employing the VRE capability to express relationships between resource types (e.g. research data, context information, code, articles), and producing effective guidance on resource cataloguing through engagement with the data providers, service catalogue providers, and VRE users.

6.3 'Trust on e-VRE system' implications

The e-VRE system has a micro services architecture: e-VRE services can be distributed over different servers and to implement functionalities they need to communicate with each other. In general the security of micro services architecture involves three main perspectives:

- **Secure Development and Test**: an important advantage in a micro service architecture is the possibility of implementing a service, testing it and instantly deploying it in production. This development/test/deployment path is also valid for service update and service replacement. To reduce the possibility of introducing security (and trust) vulnerabilities at the code level, the project has defined a security policy for development and a set of integration tests that will be automatically implemented via a Continuous Integration (CI) framework.
- **Hosting security**: Micro service architectures are distributed; possibly we can have a configuration where every service runs in a separate host. It is important to define a security

deployment policy and carefully decide if this policy is implemented by a candidate host for a service. To try to reduce this kind of security issues the project plans to distribute e-VRE service using containers; we are currently analysing available containers technology to understand which proposal could give us the highest possible level of security implementation.

- **Application level security**: the secure communication channels between micro services are critical issues in a micro service architecture. In e-VRE, two interaction paradigms are implemented:

    a. Request driven: every e-VRE service will publish a number of integration entry points via a Web Service interface, so that other services can use this interface to send/request data or to activate service functionalities.

    b. Event driven: e-VRE services will use a communication bus to asynchronously exchange messages reporting the occurrence of events.

In both cases secure communication channels will be used, adopting asymmetric encryption algorithms to encode messages. This can be done in many different ways, the approach that the project plans to adopt is to build a private Certificate Authority (CA) and to provide services with the certificates granted by this authority. As technical solutions we plan to use SSL (HTTPS) for request driven interactions and JSON Web Token (JWT) for asynchronous messages.

7 Potential influence on architecture design and adjustment strategy

7.1 Potential influence on architecture design

The user requirements collected in Deliverable 2.1 (State-of-the-art and user requirement analysis report) that have directly influenced the design of the AAAI component of the e-VRE system are listed and briefly described in Table 3-1, 3-2, 3-3. The user interacting with the e-VRE system can be a person or a software agent, for instance a program that controls an instrument that is integrated in e-VRE (see requirement CLRQ1 in Table 3.1).

The e-VRE authentication protocol will be based on scoped credentials assigned to a user or an agent and controlled by authenticators. Please note that we are referring here to the authentication process between a user and the e-VRE system. This process will be based on the protocols adopted by the AAAI component. The scoping of the credentials will be enforced jointly by a User Agent implementing the E-VRE authentication API and an authenticator that holds the credential, by constraining the availability and usage of credentials. Scoped credentials are located on authenticators, which can use them to perform operations subject to user consent. Authenticators can run as separate services from the user agent, this configuration is used to implement requirements about instruments/devices integration or logging. The resources accessed by a user can belong to the e-VRE system or can belong to e-RIs. In this case the authorization to operate on the resource can require interactions with the e-RIs security management systems.

7.2 Adjustment strategy

The improvement of strategies for the VRE end-users to handle security, privacy and trust issues does not stop when the architecture or prototypes have been developed. There will be a continuous effort in the VRE4EIC project improving the strategies regarding these three aspects. In the evaluation workshops and other evaluation sessions (see deliverable 7.2) particular attention will be paid to these evolving strategies. For instance, there might be strategies that have not been clearly defined or that are not suitable for the operation of e-VRE prototype, but that may appear to be important when people actually start working with the e-VRE in practice. After several experiments for testing the prototype of e-VRE or operating the e-VRE certain periods, these proposed strategies will be reviewed. Such strategies will also be discussed in sessions with end-users of the e-VRE system. Feedback will be sent to the VRE4EIC consortium. The system design might be updated, and related security, privacy and trust strategies may be adapted accordingly (see deliverable2.4).

8 Conclusions

Security, privacy and trust can be enhanced not only by the proper design and operation of the e-VRE, but also by appropriate use of the e-VRE by the end-users. In this deliverable, we have discussed the strategies of existing VRE-related projects for users to handle security, privacy and trust issues. 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 insights in how the VRE4EIC project deals with the issues from both the perspective of the technological solution and the perspective of information governance.

The described strategy reflects the solutions and suggestions made by the VRE4EIC consortium regarding security, privacy and trust issues. This document will be updated in M33 based on the prototype developed in the WP3 and the evaluations in WP2 and WP6. The second version of this deliverable will contain details regarding how the end users develop their own strategies regard those three aspects and adapt their use behaviors to the implemented e-VRE. Upon acceptance of this strategy document, it will be made publicly available and especially distributed to potential user groups. 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 of handling these issues.

9 References

Bird&Bird. (2017). *Guide to the General Data*

Protection Regulation. Retrieved from https://www.twobirds.com/~/media/pdfs/gdpr-pdfs/bird--bird--guide-to-the-general-data-protection-regulation.pdf?la=en

de Montjoye, Y.-A., 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

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

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

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. P. (2007). An Information Security Governance Framework. *Information Systems Management, 24*(4), 361-372. doi:10.1080/10580530701586136

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems, 12*(4), 5-33.

10 Annexes

10.1 Training materials

The training materials are developed in the working package 6, the access to the training materials can be found in Deliverable D6.3 “Engagement and Training Events and their Evaluation – First version”. More training materials will be generated along with the development of e-VRE.