VRE4EIC

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

Deliverable D2.1

State-of-the-art and user requirement analysis

Document version: 1.2

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:	14.03.2015
Document	
Period covered:	M1-M6
Deliverable number:	D2.1
Deliverable title	State of the Art and User Requirement Analysis
Contractual Date of Delivery:	31.03.2016
Actual Date of Delivery:	31.03.2016
Editor (s):	-
Author (s):	Yi Yin (TU Delft), Anneke Zuiderwijk (TU Delft)
Reviewer (s):	Laura Hollink (CWI), Carlo Meghini (CNR)
Participant(s):	CWI, EuroCRIS, INGV, UvA
Work package no.:	2
Work package title:	State-of-the-art and user requirements analysis
Work package leader:	TU Delft
Distribution:	PU
Version/Revision:	1.2
Draft/Final:	Final
Total number of pages (including cover):	51

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

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

The VRE4EIC project aims at making it easier for researchers to reuse heterogeneous scientific datasets from multiple disciplines. The project will deliver a VRE reference architecture, a reference implementation, and prototypes for two e-Research Infrastructures (e-RIs) (EPOS and ENVRI), removing barriers of existing e-RIs and providing a single point of homogeneous access to heterogeneous data and tools that support data reuse. VRE4EIC covers all European Union member states and European Free Trade Association countries, with a potential reach of 70,000 researchers all over Europe.

In order to meet the needs of VRE4EIC's end users, VRE4EIC has used various methods to collect functional and non-functional requirements.

The objectives of the VRE4EIC requirements collection and analysis are described in Task 2.1. They include:

- Conduct a literature review of the state-of-the-art in VRE-development and identify architecture, prototype and use case requirements;
- Identify existing VRE related initiatives and obtain user requirements for enhancing existing VREs;
- Conduct interviews with potential end-users of the VRE prototype and identify architecture, prototype and use case requirements;
- Conduct a comprehensive questionnaire among potential end-users of VRE4EIC and identify requirements;
- Integrate and prioritise the user requirements obtained through the above-mentioned methods and provide them to WP3.

This deliverable focuses on a part of these objectives. It provides the methods used for the requirements collection of the VRE architecture and prototypes, and presents the first findings from the user requirement analysis. It also includes the literature review of the state-of-the-art in VRE-development. The collection of user requirements will be a continuous effort throughout the project duration and will continue after the creation of this deliverable. For instance, the integration and prioritization of the user requirements will take place in the following months (based on the use case definition), and the comprehensive questionnaire will be sent out to potential end-users.

This deliverable is structured as follows. In the following section (section 2) the methodology used to collect requirements is described. Thereafter we describe the state-of-the-art and elicit requirements from the research lifecycle and from the literature (section 3). Subsequently, in section 4 we analyse the user requirements, and in section 5 and section 6 we discuss e-RI requirements collected from the EPOS project and the ENVRIPlus project. In section 7, conclusions are drawn and insight is provided in planned future activities regarding requirement collection and refinement.

2 Approach

Research on requirement engineering has received more and more attention as understanding user requirements is generally recognized as not only the most crucial but also the most difficult stage for the successful development, deployment and evolution of information systems (Browne & Rogich, 2001; Maguire & Bevan, 2002; Yu, 1997). In this chapter we review theories and practices in existing requirement engineering in the field of Information Systems (IS). The methodology of user requirements acquisition for the VRE4EIC project is based on existing theories and practices.

2.1 Requirement definition

There are many definitions on requirements from the literature (Zave & Jackson, 1997). IEEE 610.12-1990 standard defines a requirement as: *"1) A condition or capability needed by a user to solve a problem or achieve an objective, 2) A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed documents, 3) A documented representation of a condition or capability as in (1) or (2)", P62* (IEEE-STD, 1990). Aybuke and Claes claimed that requirements should include both user needs and those arising from other stakeholders like organizations, governmental bodies and industry standards (Aybuke & Claes, 2005). Aybuke and Claes categorized requirements in several ways (Aybuke & Claes, 2005):

- Functional versus non-functional: functional requirements describe the form of service and non-functional requirements describe the related constraints attributes to functional requirements such as performance, reliability or security, etc.
- Goal level requirements: business related goals
    - Domain level requirements: problem relevant area
    - Product level requirements: product related
    - Design level requirements: what should be built?
- Primary requirements which are collected from the stakeholders
    - Requirements which are derived from primary requirements
- Other classifications, e.g. business versus technical, product versus process and role based requirements.

What is common among requirement definitions is that they refer to describing what the proposed information system is supposed to do and how it should do this (Robinson, Pawlowski, & Volkov, 2003; Yu, 1997). However, the understandings regarding "what" and "how" differ a lot from various stakeholders, and it is not easy to identify the differences between different requirements classification in practise (Curtis, Krasner, & Iscoe, 1988).

Since this deliverable aims to describe requirements for the architecture, prototypes and use cases, it focuses on functional and non-functional requirements. Nevertheless, the architecture, prototypes and use cases should be developed from the perspective of numerous heterogeneous user needs from different disciplines while the requirements keep on changing and evolving during the research process. Therefore, the perspectives and requirements of multiple target groups are examined (see section 2.2).

2.2 Target groups

The individuals and organizations using an information system influence the basic requirements for the system. The users influence the system level complexity, the amount of necessary features and the way data needs to be presented. The VRE that will be developed aims to connect existing e-RIs

which may influence scientific researchers from different disciplines. As described in the Description of Work, the target groups of this VRE include:

- Researchers as *VRE users*, including academic and governmental researchers, research managers, educators, students, innovators, entrepreneurs and the interested citizen;
- *VRE developers*, including commercial (large IT companies, SMEs, entrepreneurs) and non-commercial (universities, not-for-profit organisations, foundations, VRE related projects) developers;
- *Scientific VRE researchers*, including academics who conduct research on VREs, for instance on VRE components and VRE communities;
- *VRE data publishers*, i.e. publishers who wish their data to be available to VRE users, including research institutions and archives, universities, governmental organisations, various researchers and other data publishers.
- *Other*. At the same time, we envision other potential target groups, such as journalists, educators and students, although these groups are not key to the project.

These target groups may overlap. For example, data publishers can also be VRE users. The target groups will be targeted especially in the domains of earth and environmental sciences related to other sciences (e.g. social sciences, humanities, life sciences, physics and other domains), as well as in the other domains mentioned in the Description of Work.

In this deliverable, the requirements are identified from the perspective of each of these target groups. In chapter 3 we discuss the requirements from the perspective of scientific VRE researchers. In chapter 4 we discuss the requirements from the perspective of researchers as VRE users. Chapters 5 and 6 contain the requirements for VRE developers and VRE data publishers.

2.3 Objective of the e-VRE

VRE4EIC develops a reference architecture and software components for VREs (Virtual Research Environments). This e-VRE bridges across existing e-RIs such as EPOS and ENVRIplus, both represented in the project, themselves supported by e-Is such as GEANT[1], EUDAT[2], PRACE[3], EGI[4], OpenAIRE[5]. 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.

2.4 Requirements engineering process

The requirements process involves the investigation and learning about the problem domain in terms of understanding the actual goals, needs and expectations of the users regarding a system (Koukias et al., 2013). Browne et. al (2001) stated three steps in the requirements determination process, including 1) information gathering, 2) representation and 3) verification (Browne & Rogich, 2001). Maguire et. al (2002) mentioned that the requirement analysis process encompasses 4 steps, namely 1) information gathering, 2) user needs identification, 3) envisioning & evaluation and 4) requirement specifications. Parviaien et. al (2003) stated three phases in the requirement engineering processes,

[1] http://www.geant.net/Pages/default.aspx
[2] http://www.eudat.eu/
[3] http://www.prace-ri.eu/
[4] http://www.egi.eu/
[5] https://www.openaire.eu/

including 1) requirements elicitation, 2) requirements analysis & negotiation and 3) requirements validation (Parviainen, Hulkko, Kaariainen, Takalo, & Tihinen, 2003). In reality, the requirements collection is a continuous and iterative process which needs to accommodate changes depending on the involved organisations, environment and stakeholders.

In the VRE4EIC project, we integrate the requirements engineering processes as mentioned above. Figure 1 depicts the requirements engineering process used in the VRE4EIC project.

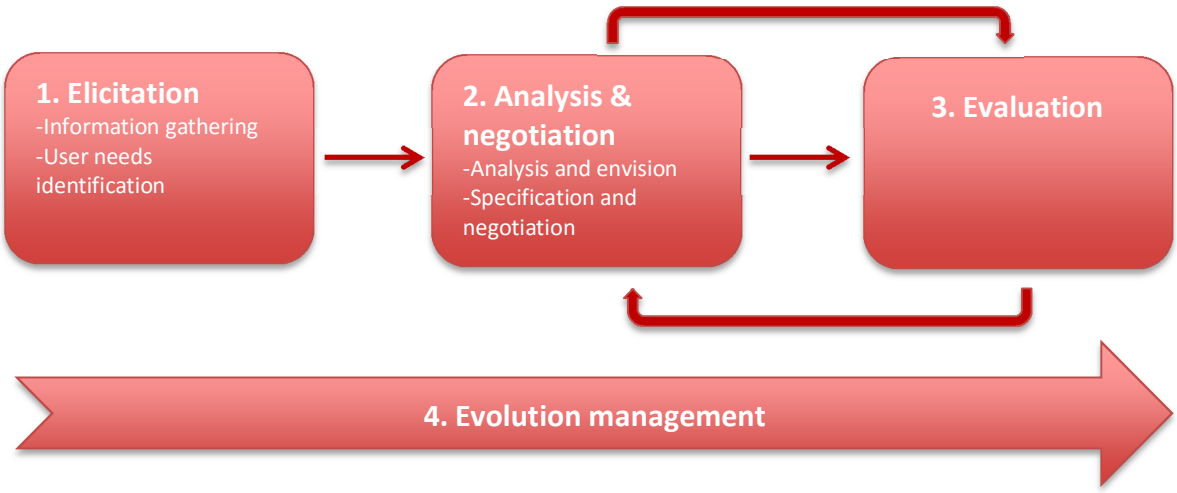


Figure 1 General process for user requirment engineering proces

Below we explain how we identified and elicited requirements for the VRE4EIC project through each of the steps shown in Figure 1.

2.4.1 Step 1: Elicitation

Requirements elicitation helps to discover requirements through information gathering and user needs identification:

- Information gathering. The first step in user requirements engineering process is to gather related background information from users, documentation or business market research by analysts. Several techniques can be used in this step, including stakeholder analysis, secondary market research, context of use analysis, task analysis, rich pictures, field study & observational methods, diary keeping and video recording (Maguire & Bevan, 2002). In this phase, various types of information are collected from different sources in terms of covering many user expectations and stakeholder interests.
- User needs identification. After the user information is collected, analysts can start to identify the real user needs and expectations. Several methods can be used in this step, including user surveys, focus group discussions, interviews, use case development, workshops and evaluating the existing systems (Maguire & Bevan, 2002). In this step, the initial user requirements are analysed and conceptualized.

2.4.2 Step 2: Analysis and negotiation

Once an initial set of user requirements has been formulated, requirements are detailed and discussed and agreed by stakeholders in the analysis and negotiation phase, including two main steps:

- Analysis and envision. When analysing and describing the requirements, it is essential to fully document “the design element or its interfaces in terms of requirements (functional, performance, constraints and design characteristics)” (IEEE-STD, 2007), P9. The Unified Modeling Language (UML) is widely used to describe the user requirements and related use

cases. After describing the requirements, it is also necessary to develop a prototype to illustrate the requirements and get feedback from the stakeholders. On the basis of the feedback, the requirements are evaluated and may be modified. The possible techniques in this step including brainstorm, card sorting, affinity diagramming, scenarios presentation, prototyping, function allocation, design guidelines & standards and parallel design sessions (Maguire & Bevan, 2002).

- Specification and negotiation. During the analysis step of user requirements, the following should be discussed with all stakeholders and documented within the specification: identification of the range of relevant users, clear design goals, the requirements with prioritized levels and evaluation criteria to test the requirements whether will be fulfilled and evidence of acceptance of the requirements by stakeholders. The following methods can be used for specification and negotiation: function mapping, requirements categorisation, prioritisation, evaluation criteria setting (Maguire & Bevan, 2002).

2.4.3 Step 3: Evaluation

The evaluation of requirements mainly checks the consistency and completeness of the requirements (Parviainen et al., 2003). The phase is concerned with the examination of the requirement document to ensure that it defines the system in an accurate and complete way.

2.4.4 Step 4: Evolution management

Requirements are the starting point for the system design phase (Parviainen et al., 2003). However, we cannot wait for complete requirements before starting the design. As the content and the priority of the initial requirements from different viewpoints may evolve and change during the development process, requirements engineering processes need to accommodate and manage changes in the requirements. Successful information systems always evolve as the environment in which these systems operate changes and stakeholders' requirements change (Nuseibeh & Easterbrook, 2000). Therefore, evolution management is considered as a parallel support process in the whole requirements engineering processes (Parviainen et al., 2003). The requirement evolution management or change management contains "documenting and evaluating the change justification, change impact assessment, decision-making to approve or reject the change and implementation of the change"(Hooks & Farry, 2001).

2.5 Methods for requirements engineering

This deliverable reports on the findings from the elicitation and initial analysis phase in the requirement engineering processes of the VRE4EIC project. First, a literature overview is created to identify the state-of-the-art and to elicit requirements (chapter 3). This non-exhaustive overview of challenges is derived from the literature, and has been selected based on VRE experiences of the partners in this consortium.

Semi-structured interviews from different disciplines are used to analyse user requirements (chapter 4). Ten interviews have been conducted to elicit requirements. The interviews mainly involved scientific researchers and IT developers who had experience with e-RI and VRE-related projects before. Even though the interviews mainly involved scientific researchers and IT developers, the interviewees a) were domain-experts and representative of huge communities, b) they had also studied the requirements from the other target groups, and c) they were asked to answer the interview questions from the perspective of each of the VRE4EIC user groups.

The interviewees covered scientists with diverse culture background located in The Netherlands, Germany and Italy, ranging from material science, environmental science, geo-science, health and social science. Besides the scientist, several librarians and IT developers have been also interviewed.

An interview protocol has been created to guide and structure the interviews (see Appendix A). The interview protocol was created through many iterations in collaboration with various VRE-experts. The questions of the interviews were created using the Reference Model of Open Distributed Processing (ODP) (Linington, Milosevic, Tanaka, & Vallecillo, 2011). The questions cover each of the five ODP viewpoints: enterprise (science), information, computation, engineering and technology, since the e-VRE should account for the needs of heterogeneous user groups. In addition, the questions concerning activities of VRE users addressed user activities in line with those mentioned in the literature (Buddenbohm et al., 2015; De Roure, Goble, & Stevens, 2009). We reviewed questions that have been used in the EPOS, ENVRIPlus and EUDATA2020 projects before, as well as the FURPS+ model (Grady, 1992), and reused questions where appropriate. All interviews have been recorded and are transcribed for further analysis purposes. The transcripts are shared with the partners, so that they can obtain detailed information concerning e-VRE requirements.

3 State-of-the-art and requirement elicitation

3.1 Research lifecycle

Tenopir et al. (2001) stated the research lifecycle includes discovering and generating ideas, finding research collaborators, finding grant opportunities and writing (research) proposals, conducting research, publishing research findings and so forth (Tenopir et al., 2011). Traditional VREs are designed to support the researchers' basic research activities (see Figure 2). In contrast, the e-VRE developed in the VRE4EIC project provides additional support for research activities, including support and related services regarding finding collaborators, finding grants, writing proposals and finding and writing publications by scientific researchers. The e-VRE will cover the whole range of research lifecycle activities. The research lifecycle elaborated on in Section 3.1 guides the requirements elicitation. However, VREs are still facing a lot of challenges which will be elaborated in section 3.2.

Figure 2 Stages of the research and data lifecycle (Tenopir et al., 2011)

3.2 Requirements from the literature

In this section we describe the requirements elicited from the literature. These requirements have also been described in a paper by Zuiderwijk, Jeffery, Bailo, and Yin (Forthcoming) that will be presented at the Conference for E-Democracy and Open Government (CeDEM). Section 3.2.1 and 3.2.2 below are based on this conference paper.

3.2.1 Key elements of VREs

"VREs have become critical to modern research processes (Buddenbohm et al., 2015). They are one of three major components or layers, namely 1) e-Infrastructures (e-Is) providing Information and Communication Technology (ICT) facilities (e.g. EUDAT, PRACE, EGI and underlying them GEANT), 2) e-Research Infrastructures (e-RIs) providing for the end-user homogeneous access over heterogeneous data but also over software, resources (of the e-Infrastructure), and 3) the VRE with its users, who can cooperatively work through the VRE. VREs support research by interconverting between the multiple underlying e-RI) supported by e-Is, while the VRE user neither knows nor cares about the underlying e-Is. While VREs depend on e-RIs, they are on a higher level of hierarchy than e-RIs and underlying e-Is, and provide more advanced functionalities for their end-users (researchers). The perspective of the user, i.e. the researcher, is central to VREs.

Compared to the early research supporting environments, such as Problem Solving Environments (PSE) and Virtual Laboratories (VL), Virtual Research Environments (VRE) have very high focus on the research collaborations, besides the support for all aspects of research activities (Sinnott & Stell, 2011). Such research collaborations include a) access to data, tools, resources from different research infrastructures, b) co-operation or collaboration between researchers at the same or different institutions, c) co-operation at the intra- and inter-institutional levels, and/or d) preserving or taking care of data and other outputs (Carusi & Reimer, 2010)" (Zuiderwijk, Jeffery, Bailo & Yin, forthcoming).

There is much confusion over what is a VRE. Some researchers claim a simple portal to access datasets is a VRE, others suggest a VRE is a composition of software services – with a user interface - accessing datasets for a particular purpose. The terms VRE, VL (mainly in Australia) and RG (Research Gateway, mainly in USA) tend to be used interchangeably. However, as indicated above VL is a term used for older systems and RGs can be confused with systems providing access to information on research activities at a university or a funding organisation.

A definition of a VRE[6] from JISC in UK is "A VRE helps researchers from all disciplines to work collaboratively by managing the increasingly complex range of tasks involved in carrying out research" and is accompanied by the vision: "The purpose of a Virtual Research Environment (VRE) is to help researchers from all disciplines to work collaboratively by managing the increasingly complex range of tasks involved in carrying out research on both small and large scales. The concept of a VRE is evolving. The term VRE is now best thought of as shorthand for the tools and technologies needed by researchers to do their research, interact with other researchers (who may come from different disciplines, institutions or even countries) and to make use of resources and technical infrastructures available both locally and nationally. The term VRE also incorporates the context in which those tools and technologies are used. The detailed design of a VRE will depend on many factors including discipline, context, and security requirements. The intention of this programme is therefore not to produce a complete VRE, but rather to define and help to develop VRE frameworks and associated standards, and to encourage the development and population of these frameworks with applications, services and resources to create VREs appropriate to particular needs" (see [6]).

3.2.2 Challenges for multidisciplinary VREs

Various challenges for the development and use of multidisciplinary VREs have already been described in the literature. The requirements for the e-VRE need to include these challenge categories. The main challenges are as follows (this overview of challenges will also be published in Zuiderwijk et al. (Forthcoming).

[6] http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx

- "Data context issues. The reuse of research data is challenging outside the borders of an enterprise or organization (De Roure & Goble, 2007), and especially beyond scientific disciplines. Accurate metadata associated with documents (Connaway & Dickey, 2010) and datasets (Zuiderwijk, Jeffery, & Janssen, 2012) is essential to understand the context in which data have been created. Such metadata is also required for software, users/researchers and research services (laboratories, equipment, facilities including computing) in order for VREs to represent the complete research environment (Jeffery, Asserson, Houssos, Brasse, & Jörg, 2014).
- Data heterogeneity issues. Large amount of data are generated and captured by sensor networks, simulations and instruments from various sources (Hey, Tansley, & Tolle, 2009). Open data from different domains can be heterogeneous (Reichman, Jones, & Schildhauer, 2011). This makes large scale integration and interoperability in VREs challenging (Candela, n.d.). VREs typically integrate existing systems and resources (Jeffery & Asserson, 2010), including a variety of (open) datasets. At the same time, systems and resources within certain domains are bound by standards, terms and practices within this domain (Candela, n.d.), and there is a lack of flexibility for the reuse of these components for multidisciplinary research.
- Fast-changing data issues. Large amounts of data are generated and captured by sensor networks, simulations and instruments from various sources (Hey et al., 2009). These datasets are used by scientists, however, their usage is challenging due to the fast updates and changes. This complicates citing datasets, since many different versions are released, and the large datasets can be difficult to use because of their volume.
- Data quality issues. The quality of information is important for researchers to determine whether they can use a dataset for a particular purpose [ref hidden for review]. A basic set of data quality dimensions includes completeness, accuracy, timeliness and consistency (Batini, Cappiello, Francalanci, & Maurino, 2009). Information quality can be high on one dimension, but low on another dimension. A researcher's trust in data might decrease by finding datasets which have poor quality, and thus insight in various quality dimensions is required.
- Privacy issues. Datasets often require removing privacy sensitive variables from it before publication, but at the same time it is often not clear which variables need to be removed exactly. Privacy and data protection legislation prescribe how one should deal with privacy sensitive data only on a high level, as the guidelines need to give sufficient space for the interpretation of privacy sensitivity (Zuiderwijk & Janssen, 2014). There is no complete list of variables that are assessed as privacy sensitive, since the privacy sensitivity of these variables also depends on their combinations and the context in which they are used (idem). 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. Some datasets cannot be published in an open environment but require a more secure space for usage, or different levels of openness. Furthermore, regulation on privacy and security may be different across countries (Zuiderwijk & Janssen, 2014), which makes sharing and using data across country borders challenging.
- User experience issues. User satisfaction is critical for gaining the benefits of using a VRE for research activities (Crosas, 2011). Connaway and Dickey (2010) identified the ease of use as a major theme for VRE and digital repository projects, and state that "ease of use and the need to embed the systems into the scholars' workflows are critical, yet can be difficult to accomplish" (p. 2). Many challenges nowadays also require the collaboration between researchers from multiple disciplines, and VREs can be used to handle the complex tasks that this multidisciplinary user collaboration demands (Edwards et al., 2014).
- Technological issues. VREs need to provide access to data, tools, and services (Carusi & Reimer, 2010; Joint Information Systems Committee, 2011), through a technical framework that is embedded in a wider research infrastructure (Carusi & Reimer, 2010; Jeffery &

> Asserson, 2010). One key requirement of VREs is that they allow for carrying out research on various levels and across boundaries, such as on an international level (Sinnott & Stell, 2011), across countries and institutions (Joint Information Systems Committee, 2011), and across disciplines (Edwards et al., 2014). Platforms, software and services across all these levels are often heterogeneous. Semantic Web technologies are essential to "provide a common framework to allow the creation of intelligent applications and services that can be integrated with data resources, people and other objects in a VRE" (Edwards et al., 2014, p. 70)." (Zuiderwijk et al., Forthcoming)

The literature has mainly focused on generic challenges for data use in a single discipline. In the VRE4EIC project, these generic challenges were not sufficiently detailed and in-depth to identify requirements to fulfil the VRE4EIC objectives. While a generic impression of the state-of-the-art can be obtained through the literature, this does not provide sufficient detail. More detailed information about requirements for multidisciplinary VREs was derived through interviews, as is described in the next section.

4 User requirements analysis

In the following sections, we present a set of tables with the functional and non-functional requirements for the e-VRE. Each requirement has a number and a description, in order to understand the requirement itself.

4.1 Functional requirements analysis

The functional requirements describe the expected functionalities and related services in the e-VRE to support the researchers' scientific research activities. In the following sections we describe functional requirements in the categories of general requirements, query requirements, funding requirements, collaboration requirements, data and service requirements and computational data requirements. These requirement categories were created to align with the clustering of the research lifecycle activities that the e-VRE will support (see section 3.1).

4.1.1 General requirements

The general requirements describe the basic user interface of the e-VRE, listed in Table 1.

General requirements of the e-VRE (GRQ)		
No.	**Requirement**	**Description**
GRQ1	Login	Login with a user account and password
GRQ2	Continuous access	Access to the software, services and datasets anywhere with internet connection
GRQ3	Single login	Ability to gain multiple accesses to the system with one login
GRQ4	Interface customization	Ability to customize the interface and functionalities of the e-VRE
GRQ5	Wizard configuration	Ability to set default options for user-interface and create user accounts through the wizard
GRQ6	User instruction	Providing training and support services for using the system
GRQ7	Multilingual interface	Multilingual support in the user-interface
GRQ8	Update alert	Having An area displaying What's New in the Virtual Research Environment Platform
GRQ 9	Online dataset editing	Ability to (co)edit dataset online
GRQ10	Notification	Sending notification when certain information available to the users
GRQ11	Additional services interfaces	Third parties being able to provide additional services based on the e-VRE

Table 1 General requirements of the e-VRE

4.1.2 Query requirements

The query requirements specify the query functionalities supporting users to browse and search for datasets and related information from various scientific research domain, listed in Table 2.

Querying for research data requirements of the e-VRE (QRQ)		
No.	**Requirement**	**Description**
QRQ1	Simple search	A single Google-style search box to search for research data
QRQ2	Multiple format support	Ability to search for dataset in different formats (including text, image, video, or domain specific file, e.g. GPS data, satellite data, programming project file)
QRQ3	Cross searching	Ability to search for data across many research infrastructures
QRQ4	Advanced search	An advanced search form enabling searching in specific data fields
QRQ5	Spelling checking	Spelling checker of search terms
QRQ6	Query suggestion	Search suggestions displayed as you type in the query
QRQ7	Query filter	Filter options for finding the data (e.g. by language, by time, by date, by size, by file type, by video time of duration, etc.)
QRQ8	Datasets viewing	Ability to view all the available datasets resources, and select which are included in each search
QRQ 9	Datasets pre-selection	Relevant resources and databases to be pre-selected by user's department
QRQ10	Dataset customization	Build user's own lists of resources and databases to search
QRQ11	Linking external resources	Direct links to external resources and datasets to allow user to query each one separately

Table 2 Querying for research data requirements of the e-VRE

4.1.3 Funding requirements

The funding requirements specify the functionalities supporting users to get research funding related information and apply for research grants, listed in Table 3.

Funding requirements of the e-VRE (FRQ)		
No.	**Requirement**	**Description**
FRQ1	Search for funding	Ability to search and view calls for research funding
FRQ2	Funding proposal	Ability to search, view, and download funding proposals that have been submitted
FRQ3	Electronic funding bid	A structured electronic funding bid template with electronic distribution and sign-off by funding body
FRQ4	Funding body information	funding agency's information to be included in research proposals
FRQ5	Funding alert	Automated alerts about new funding opportunities based on researcher's preferences

Table 3 Funding requirements of the e-VRE

4.1.4 Collaboration requirements

The collaboration requirements specify how the e-VRE supports users to set up a collaboration team and assist researchers' collaboration activities (see Table 4 and 5).

Research collaboration network requirements of the e-VRE (CRQ)		
No.	**Requirement**	**Description**
CRQ1	Research team setup	Ability to easily set up research teams
CRQ2	Finding collaborators	Ability to locate previous collaborators and potential collaborators
CRQ3	Expertise finding	Mechanism to find researchers with specific expertise
CRQ4	Forum tool	Forum tools to discuss with other researchers
CRQ5	SNS integration	A collaboration network integrated with public social media network

Table 4 Research collaboration network requirements of the e-VRE

Collaboration assistance requirements of the e-VRE		
No.	**Requirement**	**Description**
CRQ6	Group newsletter	Group periodic newsletter
CRQ7	Meeting organizer	Tools to arrange team meeting, book rooms, catering booking, etc.
CRQ8	Digest email	Collation/digest of e-mail groups
CRQ9	Tele-conferencing	Tools for tele-conferencing
CRQ10	Instant message	Instant message tool
CRQ11	Project monitoring	Mechanism to monitor involved research project progress

Table 5 Collaboration assistance requirements of the e-VRE

4.1.5 Data and service requirements

The data and services requirements specify how the e-VRE will help researchers process data concerning data retrieval, access and curation. The main research data requirements are adapted from the ENVRI[7] project, listed in Table 6 (Chen et al., 2013).

Research Data requirements

Data acquisition requirements of the e-VRE (DRQ)		
No.	**Requirement**	**Description**
DRQ1	Instrument Integration	Ability to create, edit and delete an instrument or sensor which will generate data
DRQ2	Instrument Configuration	Ability to set-up an instrument or a sensor in the network
DRQ3	Instrument Calibration	Ability to control and record the process of aligning or testing a sensor against dependable standards or specified verification processes
DRQ4	Instrument Access	Ability to read and/or update the state of an instrument
DRQ5	Configuration Logging	Ability to collect configuration information or (run-time) messages from a sensor (or a sensor network) and outputs into log files or specified media which can be used by routine troubleshooting and in incident handling
DRQ6	Instrument Monitoring	Ability to check the state of a sensor or a sensor network which can be done periodically or when triggered by events

[7]For more details regarding ENVRI, see https://envrireferencemodel.atlassian.net/wiki/display/ERM/Start

DRQ7	(Parameter) Visualisation	Ability to output the values of parameters and measured variables a display device.
DRQ8	(Real-Time) (Parameter/Data) Visualisation	A specialisation of (Parameter) Visualisation which is subject to a real-time constraint.
DRQ9	Process Control	Ability to receive input status, apply a set of logic statements or control algorithms, and generate a set of analogue / digital outputs to change the logic states of devices.
DRQ10	Data Collection	Ability to obtain digital values from a sensor instrument, associating consistent timestamps and necessary metadata.
DRQ11	(Real-Time) Data Collection	A specialisation of Data Collection which is subject to a real-time constraint.
DRQ12	Data Sampling	Ability to select a subset of individuals from within a statistical population to estimate characteristics of the whole population.
DRQ13	Noise Reduction	Ability to remove noise from scientific data.
DRQ14	Data Transmission	Ability to transfer data over communication channel using specified network protocols.
DRQ15	Data Transmission Monitoring	Ability to check and report the status of data transferring process against specified performance criteria.
DRQ16	Data backup	Ability to backup datasets according to specified policies
DRQ17	Data archive	Ability to archive datasets according to specified policies
Data curation requirements of the e-VRE		
No.	**Requirement**	**Description**
DRQ18	Data Quality Checking	Ability to detect and correct (or remove) corrupt, inconsistent or inaccurate records from data sets.
DRQ19	Data Quality Verification	Ability to support manual quality checking.
DRQ20	Data Identification	Ability to assign (global) unique identifiers to data contents.
DRQ21	Data Cataloguing	Ability to associate a data object with one or more metadata objects which contain data descriptions
DRQ22	Data Product Generation	Ability to process data against requirement specifications and standardised formats and descriptions. (optional/may be null)
DRQ23	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.
DRQ24	Workflow Enactment	Ability to interpret predefined process descriptions and control the instantiation of processes and sequencing of activities, adding work items to the work lists and invoking application tools as necessary.
DRQ25	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.
DRQ26	Data Replication	Ability to create, delete and maintain the consistency of copies of a data set on multiple storage devices.
DRQ27	Replica Synchronisation	Ability to export a packet of data from on replica, transport it to one or more other replicas and import and apply the changes in the packet to an existing replica.
Data access requirements of the e-VRE		
No.	**Requirement**	**Description**
DRQ28	Access Control	Ability to approve or deny the access requests based on specified access policies.
DRQ29	Resources Annotation	Ability to create, change or delete a note that reading any form of text, and associate them with a computational object.

DRQ30	(Data) Annotation	A specialisation of Resource Annotation which allows to associate an annotation to a data object.
DRQ31	Metadata Harvesting	Ability to regularly collect metadata (in agreed formats) from different sources.
DRQ32	Resource Registration	Ability to create an entry in a resource registry and insert resource object or a reference to a resource object in specified representations and semantics.
DRQ33	(Metadata) Registration	A specialisation of Resource Registration, which registers a metadata object in a metadata registry.
DRQ34	Data Conversion	Ability to convert data from one format to another format
DRQ35	Data Compression	Ability to encode information using reduced bits by identifying and eliminating statistical redundancy
DRQ36	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.
DRQ37	Data Citation	Ability to assign an accurate, consistent and standardised reference to a data object, which can be cited in scientific publications.
DRQ38	Semantic Harmonisation	Ability to unify similar data (knowledge) models based on the consensus of collaborative domain experts to achieve better data (knowledge) reuse and semantic interoperability.
DRQ39	Data Discovery and Access	Ability to retrieve requested data from a data resource by using suitable search technology.

Table 6 Research data requirements of the e-VRE adapted from ENVRI Common requirements analysis report (Chen et al., 2013)

Service requirements

The service requirements specify the services which the e-VRE can offer for the users.

Service requirements of the e-VRE (SRQ)		
No.	**Requirement**	**Description**
SRQ1	Computing resource connection	Access to software tools and High Power Computing facilities for data mining and feature extraction for big data sets.
SRQ2	Education support	Ability to make elements of the research process and outcomes available for the educational purpose
SRQ3	Financial information	Ability to publish financial information on Research Infrastructures, cost, subscription fee, operation and maintenance costs, etc.
SRQ4	Accounting	Accounting services for data and services provider
SRQ5	Workflow engine	The provision of a workflow to link together the software services as they access appropriate data
SRQ6	API	Application programming Interface for third parties to provide additional computational data-driven services

Table 7 Service requirements of the e-VRE

4.1.6 Computational data requirements

Computational data requirements specify how scientific data needs to be aggregated from different sources and be processed and analysed in terms of scientific experiments (see Table 8).

Computational data requirements of the e-VRE (PRQ)		
No.	**Requirement**	**Description**
PRQ1	Data assimilation	Ability to combine observational data with outputs from a numerical model

		to produce an optimal estimate of the evolving state of the system
PRQ2	Data analysis	Ability to inspect, clean, transform data, and to provide data models with the goal of highlighting useful information, suggesting conclusions, and supporting decision making
PRQ3	Data mining	Ability to supports the discovery of patterns in large datasets
PRQ4	Data extraction	Ability to retrieve data out of (unstructured) data sources, including web pages ,emails, documents, PDFs, scanned text, mainframe reports, and spool files
PRQ5	Scientific Modelling and Simulation	Ability to support of the generation of abstract, conceptual, graphical or mathematical models, and to run an instance of the model
PRQ6	Scientific workflow enactment	A specialisation of Workflow Enactment, which support of composition and execution a series of computational or data manipulation steps, or a workflow, in a scientific application. Important processes should be recorded for provenance purposes
PRQ7	Visualisation	Ability to graphically illustrates scientific data to enable scientists to understand, illustrate and gain insights from their data
PRQ8	Data Processing Control	A functionality that initiates the calculation and manage the outputs to be returned to the client
PRQ9	Data Processing Monitoring	A functionality that checks the states of a running service instance

Table 8 Computational data requirements of the e-VRE

4.2 Non-functional requirements analysis

The non-functional requirements specify the attributes related to functional requirements for the e-VRE. According to the interviews with the researchers, a quick-accessible, reliable, easy-to-use, low-cost virtual research environment is expected by them. Besides the performance related requirements, we need also consider the ethical, legal and privacy and security perspective according to the guidelines and principles defined by the "*European Charter for Access to Research Infrastructures*[8] ".

4.2.1 Ethical requirements

Ethical requirements specify the acceptable behaviours of the stakeholders in the e-VRE system, such as users, system developers and services providers.

Computational Data requirements of the e-VRE (ERQ)		
No.	**Requirement**	**Description**
ERQ1	Real expectations	Inform users' rights and responsibilities concerning the use of the data and services provided by e-VRE
ERQ2	Usage acknowledge	Users should acknowledge the contribution of the Research Infrastructure in any output (i.e. publication, patent, data, etc.) deriving from research conducted within its realms
ERQ3	Code of conduct	"Research Infrastructures and Users should undertake the necessary actions to adhere to the standard codes of conduct and ethical behaviour in scientific research when conducting research and using and disseminating research data and findings"
ERQ3	No discrimination	Research Infrastructures shall not discriminate on any personal grounds in granting Access to Users.

Table 9 Ethical requirements of the e-VRE

[8] https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=access_ri

4.2.2 Legal requirements

Legal requirements specify the whole development of e-VRE should comply with all legislations, especially the new General Data Protection Regulation[9] agreed by the European Parliament, the Council and the Commission in 2015.

Legal requirements of e-VRE (LRQ)		
No.	**Requirement**	**Description**
LRQ1	Confidentiality	Non-disclosure of sensitive information
LRQ2	Use log	Logs of the system usage for auditing and legal compliance
LRQ3	Licencing	Specific definition regarding software service authorization in compliance with legislation
LRQ4	Data management policy	Research Infrastructures should have a research data management policy. Research Infrastructures and Users should have an agreement on how to use the data.
LRQ5	Data and IP ownership	Specified service authorization contract
LRQ5	Less administration	The administration connected to requesting and granting access to Research Infrastructures should be kept to a minimum(European Commission 2016) .

Table 10 Legal requirements of the e-VRE

4.2.3 Privacy and Security requirements

Privacy and security requirements specify how the use of the e-VRE should be robust against cyber-attack in terms of protected privacy and security .

Privacy and Security requirements of e-VRE (PSRQ)		
No.	**Requirement**	**Description**
PSRQ1	Data dispersion	Data storage with data dispersion
PSRQ2	Secure storage	Secure storage of data, especially sensitive data
PSRQ3	Credentials protection	Ability to protect the users' digital identities and credentials
PSRQ4	Anti-fraud	Protection against phishing, data loss, and malware
PSRQ5	Physical access control	Identity control of the access to the physical infrastructure

Table 11 Privacy and security requirements of the e-VRE

[9] http://ec.europa.eu/justice/data-protection/reform/index_en.htm

5 e-RI requirements analysis: EPOS

In chapter 4 we described the requirements for users of VREs. Besides these user requirements, we need to specify the requirements that are needed to allow for connecting the e-VRE to existing e-RIs. We need to know the functions which the existing e-RIs provide and metadata that describe the information to connect the e-RIs to e-VRE.

The European Plate Observing System (EPOS) developed in a EU FP7 project is a VRE to "facilitate integrated use of data, data products , and facilities from distributed research infrastructures for solid Earth science in Europe " (EPOS, 2016). EPOS covers >250 RIs (Research Infrastructures which may or may not be e-RIs) each with their own assets of data, software, users and resources (such as computers or equipment).

Metadata

The EPOS-PP project has already described the structure and content of the assets of each RI in their EPOS metadata model, RIDE. A RIDE metadata model[10] containing three layers has been developed to specify the details of the RIs (See Figure 3) (Ulbricht, 2013).

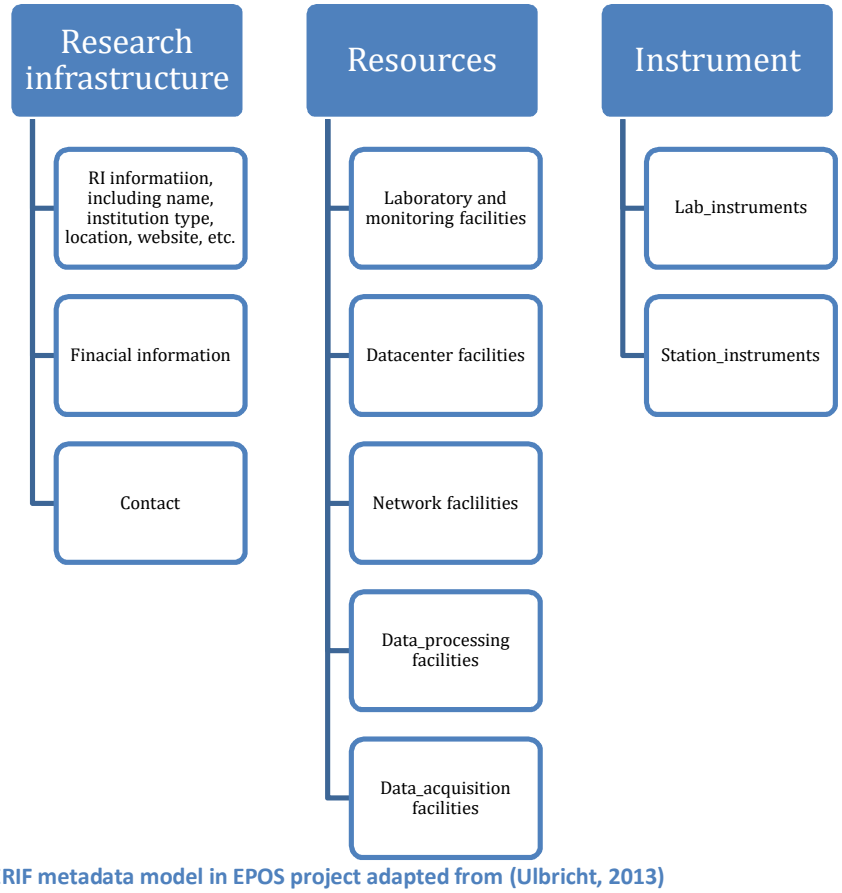


Figure 3 RIDE-CERIF metadata model in EPOS project adapted from (Ulbricht, 2013)

Within this model, the first layer contains general information about the research infrastructure, as the Institution the RI belongs to, the place where it is located, the name of the contact and other

[10] EPOS RIDE metadata model definition: https://drive.google.com/drive/folders/0B71ZSSekfEgkRWlrZFVWTWVOQUk

financial details. The second layer is a collection of resources belonging to the RI layer. This resources layer contains one or more resources. Each resource can be an item like: laboratory facility, datacenter, network facility, data processing facility, data acquisition facility, etc. The third layer is a collection of instruments in a single resource, e.g. lab instruments, monitoring instruments, stations, laboratory instruments, high performance computing resources, ICT resources (Ulbricht, 2013) . Within the subsequent EPOS-IP project, this datastructure is being converted to CERIF to form the EPOS-IP catalog.

Thematic Metadata

The EPOS-PP project defined the data products that researchers expect in the EPOS system, including 4 layers[11], namely

1. Raw data (e.g. sesmograms, accelerograms),
2. processed data from nearly automated procedures,
3. processed data by researchers' investigations,
4. processed data from complex analysis or community shared products.

In order to clearly describe the data produts, EPOS defines the following attributes of the data products, including:

- Product name
- Description (any necessary details for clarification)
- Format
- Product-specific attributes
- Institute (the offering organiztion)
- Connection protocols (the way how the data products can be accessed, by visulalization or download)
- Legal framework (restriction of use or licisensing) (Ulbricht, 2013)

At the time of writing this deliverable the major requirements collection exercise of EPOS-IP is underway. Specially designed questionnaire forms used remotely or as a basis for interview have been distributed and the information is being returned. An initial assessment of early results indicates that the generic requirements in section 4 of this deliverable match those of EPOS-IP although there are differing priorities across the individual RIs.

[11]EPOS thematic metadata definition:
https://drive.google.com/drive/folders/0B71ZSSekfEgkTkhFUFluRm8tWnc

6 e-RI requirements analysis: ENVRIPlus

At the time of writing this deliverable a major RI requirements collection exercise is underway in ENVRIplus (Review of existing RIs: requirements, achievements, technologies and gaps leading to characterisation) covering 21 RIs. ENVRIplus aligns its activities to a core strategic plan where sharing multi-disciplinary expertise will be most effective. The project aims to improve earth observation monitoring systems and strategies, including actions to improve harmonization and innovation, and generate common solutions to many shared information technology and data related challenges. It also seeks to harmonize policies for access and provide strategies for knowledge transfer amongst RIs. ENVRIplus develops guidelines to enhance trans-disciplinary use of data and data-products supported by applied use-cases involving RIs from different domains. The project coordinates actions to improve communication and cooperation, addressing Environmental RIs at all levels, from management to end-users, implementing RI-staff exchange programs, generating material for RI personnel, and proposing common strategic developments and actions for enhancing services to users and evaluating the socio-economic impacts.

The information collected from the Research Infrastructures (RIs) participating in ENVRIplus shows, after analysis, that there are the common issues and technological opportunities that were anticipated when ENVRIplus was proposed. The primary examples are :

- The need to achieve data harmonisation, i.e. consistency of representation, interpretation and access, both within and between RIs.
- The need to learn from one another and pool efforts in order to accelerate delivery of data services and working practices that support well each stage of the scientific data life cycle from data acquisition to delivery of actionable derived information. This implies sharing of common software components and composed workflows for re-use and the provision of packaged services for end-users.
- Help with facing the challenge of delivering data services immediately to meet current RI priorities while taking into account longer-term issues and technology trends.

However, care must be taken not overestimate the pervasiveness of these similarities; for example,

- differences in maturity lead to substantially different priorities, e.g. many RI currently face setting up internal collaborative support for the early stages of data acquisition, whereas a long-established RI, such as EuroARGO, prioritises improved access to existing data products,
- and differences in the internal diversity and prevailing collaborative arrangements between RIs, e.g. EPOS incorporates more than 600 independent organisations with different priorities and practices whereas EISCAT-3D has a comparatively small number of participants all focused on studying the upper atmosphere, lead to significant differences in requirements and working practices.

Prior investments, particularly in the more mature RIs have to be considered. These are not just the capital investment in equipment, software and services; they are also the training and development of working practices that become manifest in cultures and collaborative arrangements that have long-term and substantial value.

The requirements were classified into the following ICT areas: Identification and citation, Curation, Cataloguing, Processing, Provenance, Optimisation and Community support as preparation for the design phase.

An initial assessment of early results indicates that the generic requirements in section 4 of this deliverable match broadly those of ENVRIPlus although there are differing priorities across the individual RIs.

7 Conclusions and future development

In the VRE4EIC project we used the key steps of the requirements analysis processes as described by Maguire et. al (2002), Browne et. al (2001) and Parviaien et. al (2003) to analyse the state-of-the-art and user requirements. The requirements engineering process consists of four steps, namely: 1) information gathering, 2) user needs identification, 3) envisioning and evaluation and 4) requirement specifications. In this deliverable we described the results of the first two steps of this process (see section 7.1). Section 7.2 about 'Future development' will describe how steps three and four will be completed in the project.

7.1 Conclusions

Regarding the first step of the requirements analysis processes, information and user needs were derived from the literature, from interviews and from documentation of the EPOS and ENVRIPlus projects. In chapters 3-6, we discussed the following.

- Chapter 3: The state-of-the-art as described in the literature. This section provided insight in which challenges exist for the development of VREs. The challenges related to data context, data heterogeneity, fast changes to data, data quality, privacy, user experience and technology. This section showed that the literature has mainly focused on generic challenges for data use in a single discipline. In the VRE4EIC project, these generic challenges were not sufficiently detailed and in-depth to identify requirements to fulfil the VRE4EIC objectives. More detailed information about requirements for multidisciplinary VREs was therefore derived through interviews.
- Chapter 4: More information about the state-of-the-art and about user requirements was collected through interviews. From the 10 conducted interviews, we collected functional requirements in the following categories: general requirements, query requirements, funding requirements, collaboration requirements, data and service requirements and computational data requirements. Non-functional requirements were collected in the categories of performance related requirements, ethical perspective, legal perspective and privacy and security perspective.
- Chapters 5 and 6: The interviews provided insight in what users require for the e-VRE, but do not provide insight in how the e-VRE can be connected to existing e-RIs. Those requirements were therefore collected from the EPOS and ENVRIPlus projects, since these projects had already reported and are currently reporting those types of requirements and the RI requirements collected through the EPOS and ENVRIPlus projects. The metadata defined in the EPOS-PP project provides insight of the functions of existing RIs.

7.2 Future development

The requirements in this deliverable were discussed on a relatively high level of abstraction, and they need to be refined further in the context of the VRE4EIC project. The following months will be devoted to:

- refining and detailing the requirements;
- prioritizing the requirements;
- collecting lacking information.

The process of refining and detailing the requirements will start with a project meeting on April 6 and 7 in Delft, in which the implications of the collected requirements for each of the work packages will be discussed. The project partners will receive the audio recordings and the transcripts of the

interviews, which allows for detailing the user requirements for particular topics. Delft University of Technology will explain the analysed user requirements and provide input for the development tasks in the other work packages, especially WP3. The refinement and detailing of the user requirements will also be done through the development of use cases in task 2.2, and through the organization of four workshops. Deliverable 7.2 concerning the dissemination plan shows which workshops will be organized to refine the user requirements and when and where they will be organized.

In addition, the collected requirements need to be prioritized. A user questionnaire has been developed that will be disseminated in Month 7 (April 2016), so that users can complete it in the following months. The networks of the involved partners (e.g. ERCIM and euroCRIS) will be used to reach as many potential users as possible, and to let them indicate which requirements they would like to be fulfilled by the VRE4EIC project.

Furthermore, additional information needs to be collected to allow for connecting the e-VRE to existing e-RIs. In this deliverable we have described initial information that we obtained from the EPOS and ENVRIPlus projects on which functions the RIs provide and on how we can connect the e-VRE to these functions. This information was still limited and more information about RI requirements will be collected in the following months.

The requirements collection does not stop when the architecture or prototypes have been developed. Requirements collection will be a continuous effort in the VRE4EIC project. In the evaluation workshops and other evaluation sessions (see deliverable 7.2) particular attention will be paid to evolving user requirements. For instance, there might be requirements that have not been described by the users or by the related EPOS and ENVRIPlus projects, but that may appear to be important when people actually start working with the e-VRE in practice. Such requirements will be identified in sessions with actual users of the e-VRE system.

8 References

Aybuke, A., & Claes, W. (2005). Engineering and Managing Software Requirements: Springer-Verlag Berlin.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys, 41*(3), 1-52. doi: 10.1145/1541880.1541883

Browne, G. J., & Rogich, M. B. (2001). An empirical investigation of user requirements elicitation: Comparing the effectiveness of prompting techniques. *Journal of management information systems, 17*(4), 223-249.

Buddenbohm, S., Enke, H., Hofmann, M., Klar, J., Neuroth, H., & Schwiegelshohn, U. (2015). Success Criteria for the Development and Sustainable Operation of Virtual Research Environments. *D-Lib Magazine, 21*(9/10).

Candela, L. (n.d.). Virtual Research Environments. Retrieved January 5, 2015, from http://www.grdi2020.eu/Repository/FileScaricati/eb0e8fea-c496-45b7-a0c5-831b90fe0045.pdf

Carusi, A., & Reimer, T. (2010). Virtual Research Environment Collaborative Landscape Study. Retrieved January 2, 2015, from http://www.jisc.ac.uk/publications/reports/2010/vrelandscapestudy.aspx#downloads

Chen, Y., Hardisty, A., Preece, A., Martine, P., Atkinson, M., Schentz, H., . . . Zhao, Z. (2013). Analysis of Common Requirements for ENVRI Research Infrastructures. from https://envrireferencemodel.atlassian.net/wiki/display/ERM/Appendix+A+Common+Requirements+of+the+ESFRI+Environmental+Research+Infrastructures

Connaway, L. S., & Dickey, T. J. (2010). Towards a profile of the researcher of today: what can we learn from JISC projects? Common Themes Identified in an Analysis of JISC Virtual Research Environment and Digital Repository Projects Retrieved January 5, 2015, from http://repository.jisc.ac.uk/418/2/VirtualScholar_themesFromProjects_revised.pdf

Crosas, M. (2011). The dataverse network®: an open-source application for sharing, discovering and preserving data. *D-lib magazine, 17*(1), 2.

Curtis, B., Krasner, H., & Iscoe, N. (1988). A field study of the software design process for large systems. *Communications of the ACM, 31*(11), 1268-1287.

De Roure, D., & Goble, C. (2007). myExperiment–a web 2.0 virtual research environment. Retrieved December 1, 2014, from http://eprints.soton.ac.uk/263961/1/myExptVRE31.pdf

De Roure, D., Goble, C., & Stevens, R. (2009). The design and realisation of the Virtual Research Environment for social sharing of workflows. *Future Generation Computer Systems, 25*(5), 561-567.

Edwards, P., Pignotti, E., Mellish, C., Eckhardt, A., Ponnamperuma, K., Bouttaz, T., . . . Gotts, N. (2014). Lessons learnt from the deployment of a semantic virtual research environment. *Web Semantics: Science, Services and Agents on the World Wide Web, 27–28*(0), 70-77.

EPOS. (2016). What is EPOS. from https://www.epos-ip.org/about/what-epos

European Commission (2016). *European Charter for Access to Research Infrastructures*.

Grady, R. (1992). Practical Software Metrics for Project Management and Process Improvement: Prentice Hall.

Hey, A. J. G., Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: Data-intensive scientific discovery*. WA: Microsoft Research Redmond.

Hooks, I. F., & Farry, K. A. (2001). *Customer-centered products: creating successful products through smart requirements management*: AMACOM Div American Mgmt Assn.

IEEE-STD. (1990). IEEE Standard Glossary of Software Engineering Terminology. *IEEE Std 610.12-1990*, 1-84. doi: 10.1109/IEEESTD.1990.101064

IEEE-STD. (2007). ISO/IEC Standard for Systems Engineering - Application and Management of the Systems Engineering Process. *ISO/IEC 26702 IEEE Std 1220-2005 First edition 2007-07-15*, c1-88. doi: 10.1109/IEEESTD.2007.386502

Jeffery, K., & Asserson, A. (2010). e-Science, Cyberinfrastructure and CRIS. In D. J. Farace & J. Schöpfel (Eds.), *Grey Literature in Library and Information Studies*: De Gruyter.

Jeffery, K., Asserson, A., Houssos, N., Brasse, V., & Jörg, B. (2014). *From open data to data-intensive science through CERIF*. Paper presented at the 12th International Conference on Current Research Information Systems, Rome, Italy.

Joint Information Systems Committee. (2011). Virtual Research Environment Programme. Retrieved December 1, 2014, from http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx

Koukias, A., May, G., Vasyutynskyy, V., Nadoveza, D., McCarthy, J. C., Taisch, M., & Kiritsis, D. (2013). *Approach on analysis of heterogeneous requirements in software engineering.* Paper presented at the 11th IFAC Workshop on Intelligent Manufacturing Systems, IMS 2013, Sao Paulo.

Linington, P. F., Milosevic, Z., Tanaka, A., & Vallecillo, A. (2011). *Building Enterprise Systems with ODP. An Introduction to Open Distributed Processing*. Washington: Chapman & Hall/CRC Press.

Maguire, M., & Bevan, N. (2002). User requirements analysis *Usability* (pp. 133-148): Springer.

Nuseibeh, B., & Easterbrook, S. (2000). *Requirements engineering: a roadmap.* Paper presented at the Proceedings of the Conference on the Future of Software Engineering.

Parviainen, P., Hulkko, H., Kaariainen, J., Takalo, J., & Tihinen, M. (2003). Requirements engineering inventory of technologies. *VTT PUBLICATIONS*.

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *331, 11*, 703-705

Robinson, W. N., Pawlowski, S. D., & Volkov, V. (2003). Requirements Interaction Management. *ACM Computing Surveys, 35*(2), 132-190. doi: 10.1145/857076.857079

Sinnott, R. O., & Stell, A. J. (2011). Towards a Virtual Research Environment for International Adrenal Cancer Research. *Procedia Computer Science, 4*, 1109-1118.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., . . . Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *Plos One, 6*(6). doi: 10.1371/journal.pone.0021101

Ulbricht, F. H. E. D. B. K. G. J. T. L. H. D. (2013). D6.2 Report on EPOS e-Infrastructure Requirements.

Yu, E. S. K. (1997). *Towards modelling and reasoning support for early-phase requirements engineering.* Paper presented at the Proceedings of the 1997 3rd International Symposium on Requirements Engineering, Los Alamitos, CA, United States, Annapolis, MD, USA.

Zave, P., & Jackson, M. (1997). Four Dark Corners of Requirements Engineering. *ACM Transactions on Software Engineering and Methodology, 6*(1), 1-30.

Zuiderwijk, A., & Janssen, M. (2014). *The negative effects of open government data - investigating the dark side of open data*. Paper presented at the Proceedings of the 15th Annual International Conference on Digital Government Research, Aguascalientes, Mexico.

Zuiderwijk, A., Jeffery, K., Bailo, D., & Yin, Y. (Forthcoming). *Using open research data for public policy making: Opportunities of Virtual Research Environments*. Paper presented at the Conference on E-Democracy and Open Government (CeDEM), Krems an der Donau, Austria.

Zuiderwijk, A., Jeffery, K., & Janssen, M. (2012). The potential of metadata for linked open data and its value for users and publishers. *Journal of e-Democracy and Open Government, 4*(2), 222-244.

Annexes

A. Interview protocol

VRE4EIC

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

Interview protocol

WP2

Document version: 0.6 (20151120)

Interview template

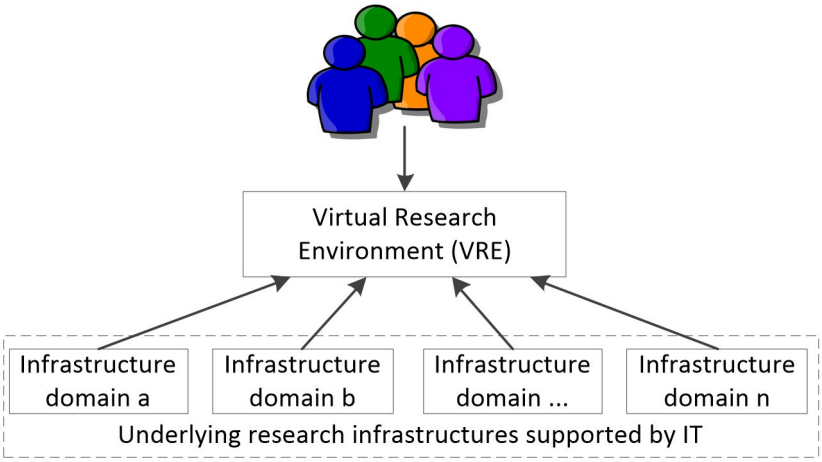
1.1 Introduction of the interview by the interviewer

Thank interviewee for being available for the interview.

At the start of the interview, provide the information below to the interviewee.

1.1.1 Definition of VRE in relation to Research Infrastructures

In this interview we will refer to Virtual Research Environments, or VREs, many times. I will now briefly explain to you what we mean with VREs. VREs support research by interconverting between multiple underlying Research Infrastructures (RIs) supported by Information Technology, while the VRE user (i.e. researcher) neither knows nor cares about the underlying RIs. While VREs depend on infrastructures, they are on a higher level of hierarchy, and provide more advanced functionalities for their end-users (researchers). The perspective of the user, i.e. the researcher, is central to VREs (see Figure below).



Compared to the early research supporting environments, such as Problem Solving Environments (PSE), Virtual Laboratories (VL), Virtual Research Environments (VREs) have very high focus on the research collaborations, besides the support for all aspects of research activities [12] . Such research collaborations include:

a) access to data, tools, resources from different research infrastructures,
b) co-operation or collaboration between researchers at the same or different institutions,
c) co-operation at the intra- and inter-institutional levels, and/or
d) preserving or taking care of data and other outputs[13].

The development of Grid, Cloud and Internet technologies enable collaborations among distributed resources from a very large scale; however, also face challenges in heterogeneity of data and metadata, interoperability between platforms and services, security and trust on remote resources, usability and community support, and other challenges.

1.1.2 Purpose of VRE4EIC project in general

VRE4EIC is a Horizon2020 project that develops a reference architecture and prototypes to be used for future VREs including building blocks that can be used to improve existing VREs. The project addresses the key data and software challenges in supporting multidisciplinary data driven sciences. VRE4EIC develops one canonical prototype that interoperates across existing VREs (to which we refer as e-VRE), as well as two prototypes in the domains of earth and environmental sciences. The e-VRE developed in the VRE4EIC project interconverts between infrastructures and VREs from multiple domains (earth sciences, environmental sciences, social sciences, humanities, material sciences, astronomy, etc.).

[12] R. O. Sinnott and A. J. Stell, "Towards a Virtual Research Environment for International Adrenal Cancer Research," *Procedia Computer Science,* vol. 4, pp. 1109-1118, // 2011.

[13] A. Carusi and T. Reimer. (2010, January 2). Virtual Research Environment Collaborative Landscape Study. Available: http://www.jisc.ac.uk/publications/reports/2010/vrelandscapestudy.aspx#downloads

i The interview

The purpose of this interview is to obtain requirements for the development of the e-VRE and the prototypes, as well as to obtain information about real-world use cases of VREs. The requirements and the real-world use cases in the interdisciplinary domains of environmental, earth, social, material, and other sciences will be developed and supported by the architecture and toolset. We start with a first interview session concerning general questions. The first session addresses the RI/VRE project leader and consists of 28 questions in different categories. The second session goes more in (technical) depth and is aimed at an IT expert involved in the project and consists of 33 questions. Each session will take about one hour. Additional session(s) might be needed later on. We envision continuous improvements and iterations of our requirement collection throughout the project.

ii Confidentiality and recording of the interview

With your permission, we would like to record the interview. The recording will only be used to create an overview of VRE requirements and use cases and not for any other purposes. The information that you provide during this interview will be treated confidentially. Findings from the interview will not be reported on individual level but only on VRE or VRE project level.

Participant signature [Place]:

[Date]

We will now start the interview.

Session 1: Questions from the interviewer to the VRE/RI project leader (1 hour)

1) General interview information

Name of interviewee: _____

Name of organization: _____

Name of related project: _____

Name of related VRE/ infrastructure: _____

Date of the interview: _____

To which of the following communities does the interviewee belong in his/her opinion?

1. VRE developers / IT experts (commercial and non-commercial)
2. Academics conducting research on VREs
3. Researchers as VRE users
4. VRE data publishers
5. Other, namely_____

2) Usage of the VRE/RI

1. Could you describe the main purpose of the RI/VRE that your project develops?
2. Which communities are involved in the development of the RI/VRE of your project?
3. Which communities and research domains are involved in the use of the RI/VRE in your project? (e.g. researchers from social sciences, material sciences, environmental sciences, etc.)
4. How many users does the RI/VRE have in relation to the total size of the potential user group?
5. How successful is the publicity work of a RI/VRE? (e.g. visibility within potential user groups, promoting the VRE to funding bodies, and the extent to which the VRE is known by the general public or within other academic disciplines)
6. Which components of the VRE are used most often?
7. Could you describe how a researcher can interact with the VRE/RI that your project develops, with regard to the following user activities? (use cases) *Please use the use case template in Appendix A.*
    a. Data access and querying
    b. User identification
    c. Data citation
    d. Data curation
    e. Data cataloguing
    f. Data processing (e.g. analysis, visualization)
    g. Data provenance
    h. Data optimization
    i. Data quality analysis
    j. User collaboration
    k. User training and support
    l. Other community support
    m. Other activities

3) VRE relevance/community acceptance

8. How intensively and how comprehensively does the target group use the RI/VRE?
9. How does this group assess the usability of the RI/VRE within its academic field?
10. Is the RI/VRE seen as a potentially useful addition for the group's research purposes?

4) Collaboration, transfer of knowledge and dissemination of expertise

11. What kinds of collaboration with other researchers does the infrastructure enable, and to what extent are these put into practice?
12. Aside from concrete research results, to what extent does the RI/VRE contribute to an accumulation of knowledge within the discipline, for example through an exchange of ideas about methods and objectives, publicly available project documentation and dissemination of expertise within a user community?
13. How many and what kind of publications and results are published in the context of the RI/VRE?

5) Questions about interactions with other RIs/subsequent use of infrastructure

14. Do you have any use case involving interactions with other RIs/VREs (in the same or different domains?) *(Use the template in Appendix A)*
15. Does your RI/VRE also allow for sharing the following elements:
    a. Datasets? Under what conditions are they available?
    b. Software/services? In this case, is it open source?
    c. Computing resources (for running datasets through your software or software on your datasets)?
    d. Access to instrumentation/detectors or lab equipment? If so, what are the open-access conditions? Are there any bilateral agreements?
    e. Users/expertise to provide advice on various topics?
    f. Access to related scholarly publications?
    g. Access to related grey literature (e.g. technical reports)?
    h. Other elements?
16. Are these elements shared in practice?

6) Performance/Usability

17. What are the key challenges for researchers interacting with the RI/VRE that your project develops?
18. Does the RI/VRE fulfil users' requirements in terms of learnability and usability?
19. Does your RI/VRE have non-functional constraints for data handling and exploitation? For example:
    a. Capital costs
    b. Maintenance costs
    c. Operational costs
    d. Security
    e. Privacy
    f. Computational environment in which your software runs
    g. Access for scrutiny and public review

    If so please provide the documentation/links
20. What part of your RI/VRE needs to be improved in order:
    a. For the RI/VRE to achieve its operational goals?

b. For you to be able to do your work?

21. What are the big open problems for your RI/VRE pertinent to handling and exploiting your data?
22. Does the RI/VRE integrate seamlessly into the day-to-day work of researchers?
23. How is the usability of the RI/VRE judged in comparison to other (generic) tools?

7) Questions about your expectations of the e-VRE

As mentioned, the VRE4EIC project develops an architecture and a canonical prototype that interoperates across existing VREs from multiple domains. We refer to this with the term e-VRE. At the moment there are several challenges for VREs, such as data context issues, data heterogeneity issues, fast-changing data issues, data quality issues, privacy issues, user experience issues and software issues.

The following questions concern your expectations of an e-VRE that handles these challenges. We will focus on the situation that you envision in three to ten years' time.

24. What do you think is needed in the e-VRE to harmonize and integrate data from different RI-sources and different domains? Please describe a use case *(using Appendix A).*
25. What objectives would you like the e-VRE to attain? Please describe a scenario.
26. How would you expect the e-VRE to handle security? Please describe a scenario.
27. How would you expect the e-VRE to handle privacy? Please describe a scenario.
28. How would you expect the e-VRE to handle trust? Please describe a scenario.

Thank you for your participation in the first session of this interview. We appreciate your valuable contributions.

Session 2: Questions from the interviewer to the VRE/RI IT-expert (1 hour)

Provide introduction. Explain what the previous interview session was about as well as its key outcomes. Explain that we would like to have a second interview session about the technical aspects of the RI/VRE.

1) Questions about interoperability, data management, standards and software services

1. Could you describe how data is acquired, curated and made available to users in the RI/VRE that your project develops? *Please use the use case template in Appendix A.*
2. Does your RI/VRE provide any of the following software services for end-users?
   - ☐ Software to orchestrate and harmonize resources, data, users and processing facilities involved in the process of discovery, access, visualization and computation of models
   - ☐ Advanced services for the exploration of the infrastructure information space
   - ☐ Tools to keep track of changes made to datasets – provenance - (e.g. for real-time data)
   - ☐ Identifiers and citations for fast-changing data
   - ☐ Data mining tools
   - ☐ Text mining tools
   - ☐ Visualisation tools
   - ☐ Modelling tools
   - ☐ Efficient data transfer tools to deliver data for computing workflows
   - ☐ Improved retrieval tools
   - ☐ Networking tools integrated with existing and dedicated social media
   - ☐ Interoperable workflows
   - ☐ Application Programming Interfaces (APIs) for developers to include new RIs to the research environment
   - ☐ Efficient deploy tools for computing tasks in the infrastructure
   - ☐ Efficient mechanism for customizing network, computing and storage services, and for optimizing data and computing tasks placement
   - ☐ Models for the description of data and service resources according to the Semantic Web standards
   - ☐ Information models and software component for the long-term archiving of and access to research data
   - ☐ Other services, namely ….
3. Are your data, software and computational environment subject to an open-access policy?
4. Could you describe how the RI/VRE deals with fast changing data?
5. Could you describe how the RI/VRE deals with data citations?
6. Could you describe how the RI/VRE deals with data quality?
7. Could you describe the computational environments involved?
8. Do the VRE's research processes adhere to established standards? What are its strengths and weaknesses?
9. Which metadata standards do you use in your VRE/RI? (e.g. DC, DCAT)
10. Are you considering changing the current standard(s), software or working practices as part of a future plan?

Ask interviewee to provide documentation/links for the standards, software and working practices that apply. If the interviewee is not the right person to reply to some questions from the above, please ask the interviewee to suggest the right person to contact from your RI/VRE for those questions.

2) Questions about privacy, security and trust

11. Do you have an overall approach to privacy and access?
12. Could you describe a use case of how the RI/VRE deals with privacy issues? *(use the template in Appendix A)*
13. Do you have an overall approach to security?
14. Could you describe a use case of how the RI/VRE deals with security issues? *(use the template in Appendix A)*
15. Do you have an overall approach to create trust of end-users?
16. Could you describe a use case of how the RI/VRE deals with trust issues? *(use the template in Appendix A)*

3) Questions about the architecture and technology

17. Could you describe how the RI/VRE system was designed (architecture)?
18. Could you describe which standards were used in the design of the system? *(apart from the standards already mentioned in your answer to question 8)*
19. Could you describe how an external system (machine) can interact with your system (e.g. APIs, RESTful web services, direct access to databases or catalogues)?
20. Are you using any standard interface to communicate with your system?
21. Could you describe which technology was used for:
    a. The programming languages?
    b. The interfaces
22. Could you specify the metadata used in you VRE/RI catalog? For instance, do you have metadata 'elements' corresponding with the proposed RDA elements (see Appendix C)
23. What are the key challenges from the VRE/RI developer perspective?

4) Questions about scalability/modularity

24. Could you describe the size of the datasets that your project deals with and whether this causes any constraints to the use of the RI/VRE?
25. How adaptable is the RI/VRE in terms of expanding the user group or the number of research topics?
26. Has the architecture of the RI/VRE been designed in such a way that at least elements of it can be used subsequently in a different context?

5) Questions about other non-functional requirements

27. Are there any special requirements that relate to Usability? For example:
    a. accessibility
    b. aesthetics ("Look and feel" of the system)
    c. consistency (e.g. of interfaces)
28. Are there any special requirements that relate to Reliability? For example:
    a. accuracy of any calculations
    b. availability of the system
    c. recoverability from a system failure
29. Are there any special requirements that relate to Performance? For example:
    a. the time the system takes to respond to particular events, such as user interaction
    b. start-up time

c. shutdown time
d. throughput of data supported by the system

30. Are there any special requirements that relate to Supportability? For example:
    a. adaptation of the software (including upgrading)
    b. system and its compatibility with previous versions of this system or legacy systems providing the same capability
    c. configurability
    d. system installation
    e. system maintenance
    f. volumes of users and data the system will support
    g. testability of the system

6) Questions about your expectations of the e-VRE

As mentioned, the VRE4EIC project develops an architecture and a canonical prototype that interoperates across existing VREs from multiple domains. We refer to this with the term e-VRE. At the moment there are several challenges for VREs, such as data context issues, data heterogeneity issues, fast-changing data issues, data quality issues, privacy issues, user experience issues and software issues.

The following questions concern your expectations of an e-VRE that handles these challenges. We will focus on the situation that you envision in three to ten years time.

31. What services would you expect the e-VRE to provide? Please describe a scenario.
32. Could you explain for each of the following software services to which extent you think they should be available in the e-VRE and provide an example of how they can be used?
    - Software to orchestrate and harmonize resources, data, users and processing facilities involved in the process of discovery, access, visualization and computation of models
    - Advanced services for the exploration of the infrastructure information space
    - Tools to keep track of changes made to datasets – provenance - (e.g. for real-time data)
    - Identifiers and citations for fast-changing data
    - Data mining tools
    - Text mining tools
    - Visualisation tools
    - Modelling tools
    - Efficient data transfer tools to deliver data for computing workflows
    - Improved retrieval tools
    - Networking tools integrated with existing and dedicated social media
    - Interoperable workflows
    - Application Programming Interfaces (APIs) for developers to include new RIs to the research environment
    - Efficient deploy tools for computing tasks in the infrastructure
    - Efficient mechanism for customizing network, computing and storage services, and for optimizing data and computing tasks placement
    - Models for the description of data and service resources according to the Semantic Web standards
    - Information models and software component for the long-term archiving of and access to research data
    - Other software services, namely…
33. Are there any other aspects that you would consider to be essential for the e-VRE?

Thank you for your participation in the second session of this interview. We appreciate your valuable contributions.

Introduction e-mail to request interview

- General introduction to the VRE4EIC project
- Mention the name of the partner in our project that recommended contacting this person
- Request interview (max. 1 hour per session)
- Clarify purpose of the interview: to collect requirements and use cases for the e-VRE
- Ask for face-to-face interview – if not possible a Skype interview
- Propose date for the interview
- Send questions to the interviewees in advance

B. Online survey

VRE4EIC

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

User requirement survey

Document version: Draft_01
Authors: Yi Yin, Anneke Zuiderwijk
Date: 21/01/2016

1. Introduction

The VRE4EIC project aims to develop reference architecture and prototypes for a Virtual Research Environment (VRE) that allows researchers to easily reuse heterogeneous scientific datasets from multiple disciplines.

This survey is targeted at anyone involved with research data. The purpose of this survey is therefore to find out your needs with regard to the use of research data from multiple disciplines. You are asked to participate in this survey, because you might (potentially) use (open) research data. The results of this survey will be used to develop and further specify the requirements of the Virtual Research Environment developed in the VRE4EIC project, and will benefit your own research activities. Completion of this survey is voluntary and the information provided by you participating in this survey is treated in a confidential way. Completing the survey will take about 10−20 minutes of your time. You may receive the survey results in a few months after we analyze the data from all participants.

The VRE4EIC project is funded under the European Commission H2020 Programme (http://www.vre4eic.eu). Among the project partners are Delft University of Technology, W3C, euroCRIS, ERCIM, CNR, INGV, CWI and University of Amsterdam. For further questions about this survey you may contact Yi Yin, Delft University of Technology (Y.Yin@tudelft.nl).

Thank you very much in advance for participating in this survey. We appreciate your time and input!

2. Questionnaire

If you want to complete the survey online, please use the following link : *****.

2.1 Background information

1) What is your gender? (Only choose one option)

Male	○
Female	○
Other	○

2) What is your age? (Only choose one option)

Under 18	○
18-29	○
30-39	○
40-49	○
50-59	○
Above 59	○

3) What is your nationality? (Only write one nationality)

4) Which research fields are you primarily associated with? (Categorized based on the World University Ranking of Time Higher Education, multiple choices)

Life Sciences	□
Environmental Sciences	□
Materials and Analytical Facilities	□
Physics and Astronomy	□
Social sciences and Humanities	□
Energy and Engineering	□
Other, namely:	□

5) Which of the following items best describes your role when using a Virtual Research Environment?

Researcher	□
Data and/or service provider	□
VRE system administrator	□

General citizen	□
Other, namely:	

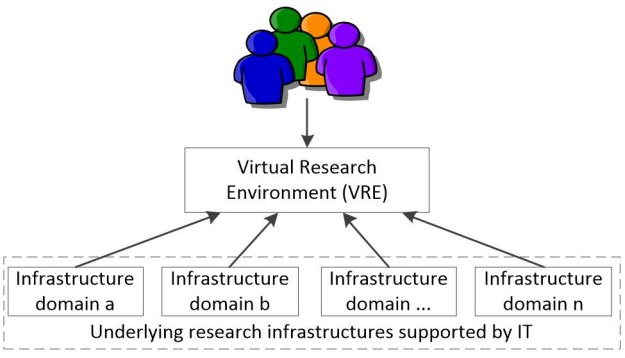
6) In daily life, how often are you involved in the following activities?

	Never	Yearly or a few times per year	Monthly or a few times per month	Weekly or a few times per week	Daily or multiple times per day
Find and acquire published information such as articles, conference proceedings, literature, research datasets	○	○	○	○	○
Find out about funding opportunities; apply for funding; manage funded projects	○	○	○	○	○
Collaboration with internal or external partners from the universities or research institution	○	○	○	○	○
Collect, share, manage, store and archive research results such as preprints, postprints, technical reports, software, or datasets	○	○	○	○	○
Process, analyze or visualize research data with online supporting tools or services	○	○	○	○	○
Other activities					

7) To which extent do you have experience with using research data?

	1	2	3	4	5	6	7	8	9	10	
No experience	○	○	○	○	○	○	○	○	○	○	Experienced

In this survey, we will refer to Virtual Research Environments, or VREs, many times. VREs support research by interconverting between multiple underlying Research Infrastructures (RIs) supported by Information Technology, while the VRE user (i.e. researcher) neither knows nor cares about the underlying RIs. While VREs depend on ICT infrastructures, they are on a higher level of hierarchy, and provides more advanced functionalities for their end-users (researchers). The perspective of the user, i.e. the researcher, is central to VREs (see Figure right).



A VRE should not be confused with a (research) infrastructure, which is underlying the VRE. A VRE provides access to multiple (research) infrastructures, while the end-user is not confronted with the complexity of accessing multiple infrastructures through one homogeneous Virtual Research Environment.

8) To which extent do you have experience with using Virtual Research Environments? (VREs)

	1	2	3	4	5	6	7	8	9	10	
No experience	○	○	○	○	○	○	○	○	○	○	Experienced

This VRE reference architecture developed in the VRE4EIC project is called Excellent VRE, or e-VRE. The e-VRE, potentially benefiting more than 70,000 European researchers, will remove barriers of existing VREs and provide a single point of homogeneous access to heterogeneous data and an architecture and tools that support data reuse. The e-VRE developed in the VRE4EIC project aims to support the research activities mentioned in the previous questions by providing various functionalities. In the following questions, we ask you to indicate the importance of these functionalities in the following categories: (1) general VRE requirements, (2) funding requirements, (3) collaboration requirements, (4) data and service management requirements, (5) Computational tools or services. Even if you have never used a Virtual Research Environment, please answer the questions from the perspective of functionalities that might be important if you would use such an environment.

2.2 General requirements

9) To which extent are the following functionalities concerning the general behaviour of a Virtual Research Environment critical for your daily activities?

The general behaviour of the Virtual Research Environment		1	2	3	4	5	6	7	8	9	10		I don't know
Login with a user account and password	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Access software and information anywhere	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Be able to gain multiple accesses with one login	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Customize the interface and functionalities of the e-VRE	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Be able to set default options for user-interface and create user accounts through the wizard	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Have a service that is easy to use, with little need for help	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Provide training and support services	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Multilingual support	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
An area displaying **What's New** in the Virtual Research Environment Platform	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

Notification containing information that I specify when it becomes available via email, mobile phone message, phone call	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
API for third parties to provide additional service	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

** An API is software that exposes its functionality to other software or programs, but not to people*

10) To which extent are the following requirements concerning searching for research data useful for your daily activities?

Searching for research information		1	2	3	4	5	6	7	8	9	10		I don't know
A single Google-style search box to find data	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Be able to search for data in different formats (including text, image, video, domain specific file)	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Be able to search for data across many resources	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
An advanced search form enabling search on specific data fields	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Direct access to the native search interface for each resource to maximize retrieval capability	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Spelling checker of search terms	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Search suggestions displayed as you type in the query	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Filter options for finding the data (e.g. by language, by time, by date, by size, by file type, by video time of duration, etc.)	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

11) To which extent are the following requirements concerning the default screen for searching critical for your daily activities?

Default screen for searching		1	2	3	4	5	6	7	8	9	10		I don't know
View all the available resources and databases, and select which are included in each search	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Relevant resources and databases to be pre-selected by my department	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Build my own lists of resources and	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

databases to search													
Direct links to external resources and databases to allow me to query each one separately	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

12) To which extent are the following requirements concerning the online editing tools critical for your daily activities?

Online editing tool		1	2	3	4	5	6	7	8	9	10		I don't know
Online document editing	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Online reference management	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

2.3 Funding requirements

13) To which extent are the following requirements concerning writing and submitting funding proposals useful for your daily activities?

Writing and submitting funding proposals		1	2	3	4	5	6	7	8	9	10		I don't know
Ability to search, view, and download funding proposals that have been submitted previously	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
A structured funding bid template with electronic routing/distribution and sign-off by research administration	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Full financial information to be included in research proposals	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Automated alerts about new funding opportunities based on my preferences	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

2.4 Collaboration requirements

14) To which extent are the following requirements concerning searching for collaboration network useful for your daily activities?

Searching for collaboration network		1	2	3	4	5	6	7	8	9	10		I don't know
Easily set up teams and groups	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Ability to locate previous collaborators and potential collaborators	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Mechanism to find researchers with specific expertise	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

Support for building new interest groups	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Forum tools to discuss with other researchers	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
A collaboration network integrated with public social media	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

15) To which extent are the following requirements concerning collaboration assistance network useful for your daily activities?

Collaboration assistance		1	2	3	4	5	6	7	8	9	10		I don't know
Group periodic newsletter and meeting organizer	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Collation/digest of e-mail groups	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Tele video conferencing tool	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Instant message tool	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Online document co-editing tool	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Mechanism to monitor involved research project progress	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

2.5 Data and service requirements

16) To which extent are the following requirements concerning sharing data useful for your daily activities?

Data sharing		1	2	3	4	5	6	7	8	9	10		I don't know
Uploading and handling large datasets	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Be able to handle massive concurrent service requests	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Datasets with access control so that they can be changed safely under multiple level of authorization	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Control over who can view or modify shared files	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Provide feedback or rating to the existing data	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Enforce collaborators to use the same application in order to view or manipulate data or files	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Support multiple metadata model to archive datasets in my	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

own data warehouse													

17) To which extent are the following requirements concerning managing the research outputs useful for your daily activities?

Managing the research outputs		1	2	3	4	5	6	7	8	9	10		I don't know
Access to documents and hyperlinks should be maintained even if staff transfer to new jobs or organizations	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
A single place where I can submit research results for long-term storage, or archive and easy access	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Data citation	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Ability to monitor who has viewed and used the research data, or usage statistics	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

18) To which extent are the following requirements concerning options to help disseminate information critical for your daily activities?

Options to help disseminate information		1	2	3	4	5	6	7	8	9	10		I don't know
The list of my publications should include links to the full documents	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
People should be able to find a list of my previous and current projects	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
My publication list should be automatically generated from a repository	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
My on-line CV should be automatically updated with new projects and publications	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

19) To which extent are the following requirements concerning administrative research activities support useful for your daily activities?

Administrative research activities support		1	2	3	4	5	6	7	8	9	10		I don't know
Transforming a dataset from one standard to another	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Processing by linking metadata	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Mechanisms to make elements of	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

the research process and outcomes available for the educational purpose													
Support for peer-review activities (find reviewers, distribute articles to reviewers, track progress)	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Support for peer-review activities (find reviewers, distribute articles to reviewers, track progress)	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Accounting services for data and services provider	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

Starting from this part, the questions will focus on non-functional requirement, such as ethical, legal, privacy and security perspective.

20) To which extent are the following terms concerning ethical aspect critical for your daily activities?

Ethical		1	2	3	4	5	6	7	8	9	10		I don't know
Confidentiality towards informants and participants	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Non-disclosure of sensitive information	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Inform users about the product real possibilities	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Reliability of data	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Equality of using data and services	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

21) To which extent are the following requirements concerning legal aspect critical for your research?

Legal		1	2	3	4	5	6	7	8	9	10		I don't know
In compliance with data protection laws and regulations	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Inform users' rights and responsibilities concerning the use of the data and services provided by e-VRE	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Logs of the system usage for auditing and legal compliance	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

22) To which extent are the following requirements concerning privacy aspect critical for your research?

Privacy		1	2	3	4	5	6	7	8	9	10		I don't know
Secure transmission of sensitive data between VRE and use side, as well as, between VRE components	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Secure storage of data, especially sensitive data	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Protect the users' digital identities and credentials	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

23) To which extent are the following requirements concerning security aspect critical for your research?

Security		1	2	3	4	5	6	7	8	9	10		I don't know
Robustness against Denial-of-Service/Distributed-Denial-of-service attacks	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
The e-VRE shouldn't allow unauthorized applications to abuse the system	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Data storage with data dispersion	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
User preferences and administrative user information, such as contracts, passwords should be stored separately in two different components.	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Protection against phishing, data loss, and malware	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Administrative interface should be separated from ordinary entry points	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
User side components protection against malicious use or manipulation of its codes	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Identity control of the access to the physical infrastructure	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

2.6 Computational data requirements

24) To which extent are the following requirements concerning processing the research data useful for your research?

Processing the research data		1	2	3	4	5	6	7	8	9	10		I don't know
Analytic tools (e.g. …)	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Visualization tools (e.g. tools to create tables, graphs and maps)	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Computational tools (e.g. …)	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Automatic data tagging tool to inform the data user on what level this data can be shared	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○
Application programming Interface API for third parties to provide additional computational data-driven services	Uncritical	○	○	○	○	○	○	○	○	○	○	Critical	○

3. Comments on previous questions

25) How applicable are the features identified earlier in this questionnaire to your research needs?

Spot on	○	Good coverage	○	Meets some needs	○	Irrelevant	○

26) Does the presented features set seem complete?

Does the features set presented seem complete?	Yes	○
	No	○
	N/A	○

27) If your answer to above question is no, could you please list other features that need to be included in a Virtual Research Environment (our e-VRE) in your opinion?

4. Wrap up

This is the end of this online survey. The results of this survey will be used to find out your needs regarding the use of (open) research data in a Virtual Research Environment. The collection of user requirements for such a research environment will be used to develop the reference architecture and prototypes for the excellent VRE (e-VRE) developed by the VRE4EIC project.

Your response is very valuable to us. Thank you very much for participating in this survey.

If you are interested in the outcomes of this survey and/or the progress of the VRE4EIC project, please tick the boxes below and leave your contact details (your personal details will only be used for this project and they will not be forwarded to others). You may also go to our project portal for more information.

☐ I would like to receive the results of this survey

☐ I would like to be informed about the progress of the VRE4EIC project

Name: ……

E−mail address: ……..

LinkedIn: ……(to join the VRE4EIC−community)

Twitter: ……(to join the VRE4EIC−community)

For any further questions or suggestions you may contact Yi Yin, Delft University of Technology (Y.Yin@tudelft.nl; +31152782271) or Pierre Guisset (pierre.guisset@ercim.eu)