# VRE4EIC

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

# Deliverable D7.3

# Workshop Report

covering organisation, attendance and outcomes

Document version: 02

# VRE4EIC DELIVERABLE

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

Mr Philippe Rohou     t: +33 4 97 15 53 06     f: +33 4 92 38 78 22     e: philippe.rohou@ercim.eu

GEIE ERCIM, 2004, route des Lucioles, Sophia Antipolis, F-06410 Biot, France
Project website address: http://www.vre4eic.eu/

| Project | |
|---|---|
| Grant Agreement number | 676247 |
| Project acronym: | VRE4EIC |
| Project title: | A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration |
| Funding Scheme: | Research & Innovation Action (RIA) |
| Date of latest version of DoW against which the assessment will be made: | 14.01.2015 |
| **Document** | |
| Period covered: | M1-M8 |
| Deliverable number: | D7.3 |
| Deliverable title | Workshops Report |
| Contractual Date of Delivery: | 31.05.2016 |
| Actual Date of Delivery: | 14.06.2016 |
| Editor (s): | |
| Author (s): | Phil Archer (ERCIM/W3C), Anneke Zuiderwijk (TU Delft) |
| Reviewer (s): | Valérie Brasse (EuroCRIS), Laura Hollink (CWI) |
| Work package no.: | 7 |
| Work package title: | Dissemination, exploitation and sustainability of the VRE, including standardisation |
| Work package leader: | Pierre Guisset (ERCIM) |
| Distribution: | PU |
| Version/Revision: | 01 |
| Draft/Final: | Draft |
| Total number of pages (including cover): | 76 |

# Disclaimer

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

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

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

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

# Table of Contents

# 1 Introduction

The VRE4EIC project is using several channels through which to communicate with prospective users and developers of the eVRE. These include a series of workshops at different scales with different stakeholders held in a variety of settings. In the first 8 months of its operation, the project has held three workshops: two small scale meetings with its own partners and those from closely related projects, ENVRI$^{PLUS}$ and EPOS, and a third as a session during the highly regarded CeDEM conference. These have been important occasions for gathering use cases and detailed feedback. As outlined in the project's dissemination strategy, D7.2, these three are the beginning of an ambitious series of face to face interactions that are a characteristic of the project.

This document sets out the results of the three workshops held to date and looks forward to those already in planning, in particular, the W3C workshop, Smart Descriptions & Smarter Vocabularies (SDSVoc) scheduled for 30 November – 1 December.

# 2 Results of Past Workshops

## 2.1 Pisa, 15-17 February 2016

| Aim of workshop | Workshop title/topic | Target audience | Event | Specific (expected) outcome |
|---|---|---|---|---|
| Requirement collection and specification / architecture proposal | EPOS/ENVRI+ inputs | VRE developers | VRE4EIC project meeting | Overview of architectural proposals, interoperability plans and security, privacy and licensing proposals |

**Table 1 The plan for the Pisa workshop as stated in D7.2**

VRE4EIC is naturally integrated with many related activities and projects across Europe. As is often the case, there is an overlap in partners involved in this project and related ones. Keith Jeffery (ERCIM), and Zhiming Zhao (UvA) are both involved with ENVRI$^{PLUS}$ and Daniele Bailo (INGV) is also involved with EPOS. The project meeting in Pisa was attended by representatives from all VRE4EIC partners as well as Leonardo Candella representing ENVRI$^{PLUS}$ and Massimo Fares from EPOS from outside the project team. Thus the meeting was well placed to gather requirements and test out initial ideas for the architecture of the eVRE.

### 2.1.1 Workshop format

The workshop took the form of an extended project meeting with the afternoon of the second day dedicated to gathering requirements from the ENVRI$^{PLUS}$ and EPOS projects (the agenda is included in the appendices). Presentations were followed by extensive discussion that then fed into the discussion of the system architecture the following morning.

### 2.1.2 Workshop participants

The full list of participants for the Pisa workshop and project meeting are shown in the table below. VRE4EIC partners who are also involved in ENVRI$^{PLUS}$ or EPOS are highlighted.

| Partner | Org | Participants |
|---|---|---|
| 1 | ERCIM | Philippe Rohou<br>Keith Jeffery (ENVRI$^{PLUS}$)<br>Pierre Guisset<br>Phil Archer (W3C) |
| 2 | CWI | Laura Hollink<br>Jacco van Ossenbruggen<br>Tessel Bogaard |
| 3 | CNR | Carlo Meghini<br>Cesare Concordia<br>Anna Molino |

| | | |
|---|---|---|
| 4 | FORTH | Maria Theodoridou |
| | | Theodore Patkos |
| 5 | TUDelft | Anneke Zuiderwijk-vanEijk |
| | | Yi Yin (ENVRI<sup>PLUS</sup>) |
| 6 | EuroCRIS | Valérie Brasse |
| | | Anne Asserson |
| 7 | INGV | Daniele Bailo (EPOS) |
| 8 | UvA | Zhiming Zhao (ENVRI<sup>PLUS</sup>) |
| | | |
| ® = Remote | Technical meeting | Massimo Fares from INGV and part of the IT team of EPOS |
| | | Martin Nayembil from BGS Keyworth and part of the IT team of EPOS ® |
| | | Leonardo Candella from CNR and part of ENVRI<sup>PLUS</sup> project team. |

**Table 2 Participants in the Pisa workshop, all but one of whom were able to attend physically.**

## 2.1.3 Results

In his presentation on behalf of ENVRI<sup>PLUS</sup>, (see appendices) Zhimming Zhao identified a number of requirements for the eVRE:

1. Requirements for common operations
   - o Identification/Citation, Processing, Optimization, Curation, Cataloguing, and Provenance
2. Requirements for generic functionality
   - o Community support

These are all very much in scope for the project. The requirements for common operations emphasise the need for rich metadata. Community support – interaction between different researchers who are working on the same datasets – is a key feature of eVRE. Zhimming Zhao further identified a number of challenges faced by Research Infrastructures and VREs that he believes both projects need to tackle:

1. **Knowledge challenge**: require knowledge from different domains: atmosphere, biosphere, hydrosphere and geosphere
2. **Modelling challenge**: these processes are difficult individually, and modelling their interactions are another order of complexity;
3. **Data challenge**: require observation or measurement data of the environment and earth system in large geo-spatial scale and long time duration;
4. **Resource challenge**: require large capacities of computing, storage and network resources

It is the first of these that the eVRE will tackle in particular. Combining datasets from different domains is always a challenge and, again, it is eVRE's rich metadata handling based on CERIF that provides the way forward.

In his presentation on behalf of EPOS (see appendices), Daniele Bailo highlighted a specific need to be able to browse datasets within a bounding box that describes a place and time of interest, that is:

- set a bounding box for a region

- set the time frame of interest
- select from a list of licences/use conditions
- select from a list of file types and retrieval types
- see available software for these file types
- see available publications of the research domain in the given region
- be able to contact the author of the data

These requirements are being met in the EPOS infrastructure but they have implied requirements for any VRE to be built on top of it:

- Unified access to resources, meaning that eVRE needs to handle different access control systems, different APIs and different metadata.
- Ancillary information (e.g. which software?)
- Information about the Research Infrastructure, i.e. provenance information, including contact details, information about the facilities and equipment available.
- Access/links to related information such as wikis, publications, grey literature etc.
- Discussions with other researchers
- Guidelines about projects, equipment and more.

In short, EPOS sees the VRE4EIC project as building a knowledge aggregator and one that is sufficiently flexible to be largely agnostic about how the original sources operate.

### 2.1.4 Conclusions

The presentations by ENVRI[PLUS] and EPOS complemented those by VRE4EIC partners on requirements gathering (WP2) and the technical architecture (WP3) which led to an animated discussion about what a Virtual Research Environment is, as distinct from a Research Infrastructure. The boundary is not well defined and is likely to vary depending who you ask. What was clear from the discussion, however, was that it is that role of an aggregator that is the key to a VRE's success. A single tool that provides seamless access to data and other elements via a single interface with a single sign on, and that allows easy communication between researchers. The discussion was wide ranging and left several questions unanswered. Therefore, it was decided to extend the scope of April's technical meeting to become a further workshop.

## 2.2 Delft 6-7 April 2016

| Aim of workshop | Workshop title/topic | Target audience | Event | Specific (expected) outcome |
|---|---|---|---|---|
| Requirement collection and specification / architecture proposal | Workshop with EPOS and ENVRI+ | VRE developers | VRE4EIC meeting with involved partners | Refinement of architectural proposals, interoperability plans and security, privacy and licensing proposals |

**Table 3 The plan for the Delft workshop as stated in D7.2**

The workshop in Delft differed from the Pisa event as it was attended only by the subset of representatives of VRE4EIC partners involved with technical work. The aim was to continue the discussion of the architecture of the planned eVRE with the help of external projects that are potential future users.

### 2.2.1 Format

The workshop followed a similar pattern to Pisa with a combination of presentations and follow up discussions. The agenda is included in the appendices.

### 2.2.2 Participants

| Partner | Org | Participants |
|---|---|---|
| 1 | ERCIM | Keith Jeffery (ENVRI$^{PLUS}$) |
| | | Phil Archer (W3C) |
| 2 | CWI | Laura Hollink |
| | | Jacco van Ossenbruggen |
| | | Tessel Bogaard |
| | | Jan Weilermaker |
| 3 | CNR | Carlo Meghini |
| | | Cesare Concordia |
| 4 | FORTH | Nikos Minakakis |
| | | Theodore Patkos |
| 5 | TUDelft | Anneke Zuiderwijk-vanEijk |
| | | Yi Yin (ENVRI$^{PLUS}$) |
| 6 | EuroCRIS | Valérie Brasse |
| | | Anne Asserson |
| | | Laurent Remy |
| 7 | INGV | Daniele Bailo (EPOS) |
| 8 | UvA | Zhiming Zhao (ENVRI$^{PLUS}$) |
| | | Paul Martin |
| | | |
| Non-VRE4EIC | | James Martin (ENVRI$^{PLUS}$) |

**Table 4 Participants in the Delft workshop**

As in Pisa, the project benefited from cross-project coordination so that both ENVRI$^{PLUS}$ and EPOS were represented directly for example with James Martin from ENVRI$^{PLUS}$ from outside the project team.

### 2.2.3 Results

Since the Pisa meeting, several advances had been made that enabled a more detailed discussion.

1. ENVRI$^{PLUS}$ had completed an extensive review of requirements for Research Infrastructures and therefore Zhimming Zhao was able to more accurately pinpoint what, in his view, the eVRE should do and what it can expect of eRIs.
2. EPOS had completed two rounds of use cases and requirements collection for their own project.

3. The work on developing the architecture for the eVRE had advanced, allowing for discussion of more specific choices about things like access control mechanisms, metadata mapping, the role of Linked Data etc.

4. EuroCRIS had progressed on mapping Open Information Linking for Environmental RIs (OIL-E[1]) to CERIF and on testing an instance of FORTH's 3M tool[2] for use in VRE4EIC, highlighting the difficulty in linking two models (used in a Research Infrastructure and eVRE respectively) that are at different conceptual levels.

5. W3C had launched its Permissions and Obligations Expression Working Group[3] which is standardising how things like licences and terms of use can be expressed in a machine readable and potentially interoperable way.

6. TU Delft had made substantial progress with preparing VRE4EIC's survey to elicit use cases and requirements.

As an example of the inputs to the discussion, Daniele Bailo's presentation on the work undertaken by the EPOS project (see appendices) ended with a specific list of needs for the eVRE:

1. To provide general model for architecture and metadata.

2. To provide additional tools / guidelines to:

   a. export metadata (e.g. LOD),

   b. provide collaboration tools for scientists

   c. provide guidelines / ideas / tools for connecting data to publications

   d. harmonize EPOS with other e-RIs (API development)

During the workshop, there was an extensive discussion about the relationship between use cases and requirements. As Henry Ford famously almost said, if he'd just listened his customers, he'd have bred a faster horse rather than building a motor car. In the same way, asking potential users what they would like an eVRE to do will likely elicit suggestions for improving what is already possible but difficult, rather than new ideas. Therefore the project should be prepared to cite requirements that are implied, but not stated explicitly *and* develop a number of 'visionary use cases.' The latter will highlight new features and capabilities that make the eVRE attractive. However, that is not to say that VRE4EIC is able to work in a purely hypothetical universe.

## 2.2.4 Conclusions

A recurring theme in the workshop was the number of differences between EPOS, ENVRI[PLUS] and between other research infrastructures. Most, but not all, include a catalogue. Some include communication channels, visualisation tools and software, others do not, and so on. Exposing research infrastructure functionalities that do exist through eVRE, without making assumptions about what is available, means that the design needs to be very flexible. Having the external projects in the room provided VRE4EIC with important ground truth to complement the idea of visionary use cases.

The theme of the variety of metadata schemas and capabilities of online platforms, such as research infrastructures, is one that has resonance in many communities. It was during the Delft workshop that partners agreed the scope for a much larger scale workshop to be held later in 2016 (see section 3.2).

---

[1] https://confluence.egi.eu/display/EC/OIL-E
[2] http://139.91.183.3/3M/
[3] https://www.w3.org/2016/poe/

## 2.3 CeDEM 18-20 May 2016

| Aim of workshop | Workshop title/topic | Target audience | Event | Specific (expected) outcome |
|---|---|---|---|---|
| Requirement collection and specification / dissemination | Virtual Research Environments: Obtaining new Insights by sharing Open Data for interdisciplinary Research Purposes | VRE users (researchers) | Conference on e-Democracy and Open Government (CeDEM) | Elicitation and refinement of requirements for a multidisciplinary VRE that integrates Open Government Data and open research data |

**Table 5 The plan for the CeDEM workshop as stated in D7.2**

The workshop at CeDEM was very different in organisation and participation from the previous two. Whereas the discussions in Pisa and Delft were among peers in related projects seeking common ground in developing what the eVRE should do, the CeDEM workshop was organised along traditional lines with external attendees. The Conference on e-Democracy and Open Government (CeDEM)[4] is an annual event organised by and held at the Danube University in Krems, just outside Vienna. Although it is not within the scope of the PSI Directive, discussions of government data often include a discussion of research data since they are usually both funded from public funds. Therefore, this increasingly prestigious event provided an excellent opportunity to solicit requirements and to promote the project.

The workshop description and objectives were published as a paper submitted to the organisers, the key text of which is included below.

"Researchers can access and use more and more research data opened by the government and by publicly-funded research organizations (Anneke Zuiderwijk, 2015). They can use this data to obtain new insights, especially by combining datasets with other data. Various projects are already producing e-Research Infrastructures to give researchers access to publicly funded research and open government research data, and are developing towards Virtual Research Environments (VREs). VREs provide access to data, tools, resources from different research infrastructures, co-operation or collaboration between researchers at the same or different institutions, co-operation at the intra- and inter-institutional levels, and/or preserving data and other outputs (Carusi & Reimer, 2010). They consist of three major components:

1) e-Infrastructures providing Information and Communication Technology (ICT) facilities (e.g. EUDAT, www.eudat.eu/, and PRACE, www.prace-ri.eu);
2) e-Research Infrastructures providing for the end-user homogeneous access over heterogeneous data but also over software, resources (of the e-Infrastructure) (e.g. LifeWatch, www.lifewatch.eu/), and
3) the VRE with its users, who can cooperatively work through the VRE (A. Zuiderwijk, Jeffery, Bailo, & Yin, 2016).

Examples of projects reaching towards VREs are EPOS[5] for earth/geo-physical sciences, ENVRI+[6] for environmental sciences and EXCELERATE[7] for biological/biomedical sciences. Nevertheless, it is already clear from these existing projects that researchers who want to conduct multidisciplinary

---

[4] http://www.donau-uni.ac.at/en/department/gpa/telematik/edemocracy-conference/

[5] http://www.epos-eu.org/

[6] http://www.envriplus.eu/

[7] http://www.elixir-europe.org/about/eu-projects/excelerate

research with open research data often face various problems in existing research environments, such as issues related to data heterogeneity, user experience, and fast changes to datasets. Researchers are often willing to share their data with others under certain conditions, however, no VREs exist that meet the requirements for multidisciplinary research. This complicates the reuse of open government data by researchers in other disciplines. For instance, studying the evolution of volcanoes requires information from deep sea, solid earth and eco-system research, while these types of data are currently fragmented.

The objective of this 1,5 hour workshop is to elicit, refine and discuss requirements for a secure and trusted VRE that integrates OGD and open research data for researchers from multiple disciplines. The workshop is relevant for participants of the international Conference for e-Democracy and Open Government (CeDEM), since it focuses on topics that are key to this conference, including open data, open access, and open and collaborative government. The workshop is of interest to experts that CeDEM brings together in the area of open government, e-participation and e-democracy. The workshop is aimed at (potential) users of (open) (government) research data. (A. Zuiderwijk, Janssen, et al., 2016)"

## 2.3.1 Workshop format

The 1.5 hour workshop consisted of the following elements:
- Introduction (10 minutes). The topic and objectives of the workshop were introduced by Dr. Anneke Zuiderwijk (TU Delft). The workshop aimed to elicit, refine and discuss requirements for a secure and trusted VRE that integrates OGD and ORD for researchers from multiple disciplines.
- Using multidisciplinary research data: VRE4EIC project (15 minutes). Anneke Zuiderwijk presented the VRE4EIC project. The key objectives of a project addressing challenges related to the use of research data were presented. The initial results of the requirement analysis and ideas for the VRE reference architecture were presented.
- Challenges for information sharing of Open Data by researchers (15 minutes). Prof. Marijn Janssen (TU Delft) presented trust and privacy as key challenges that exist for sharing Open Research Data with others. Trade-offs and considerations were discussed.
- Brainstorming (all - 20 minutes). In order to elicit, refine and discuss requirements for a secure and trusted VRE, participants were asked to discuss the needs of (potential) providers and users of (open) research data in groups of three to five persons. Each group chose a different topic of discussion, such as funding requirements, collaboration requirements, data and service management requirements, and computational tools or service requirements. Each group developed a scenario concerning how researchers can share and use data in a VRE if all the identified requirements would be met.
- Results, interaction and discussion (all - 30 minutes). The group discussion functioned as a basis for a plenary discussion thereafter. Each group presented the developed scenario. Subsequently, a group discussion took place in which the participants were asked to discuss the requirements for a VRE that their scenario pointed at. The discussion provided participants with insight in the way that open data might be used by researchers in the future and it showed the key requirements that researchers have for realizing these scenarios. Participants were told that the results of this discussion will be used to develop and further specify the requirements of the VRE4EIC research environment. At the end of the workshop, the participants were given leaflets concerning the VRE4EIC project.

## 2.3.2 Workshop participants

The workshop was aimed at researchers as VRE users and attracted the expected audience. In total, 16 people participated in the workshop. Participants of the workshop (potentially) used open

government data for research purposes and were therefore an important target group for the VRE4EIC project. The table below provides information concerning the 16 workshop participants. It shows that the participants come from 9 different countries and mainly work for universities, but also for companies, NGOs and a non-university research organization.

| Group | Gender | Position | Type of organization | Country |
|---|---|---|---|---|
| Group 1 | Female | Research fellow at the Centre for E-Governance at Danube University Krems | University | Austria |
| | Male | Professor at University of Jerusalem | University | Israel |
| | Male | Professor at Zeppelin University | University | Germany |
| | Female | Researcher AIT Austrian Institute of Technology | Non-university research institute | Austria |
| Group 2 | Male | Doctoral researcher at Centre for development informatics, University of Manchester | University | United Kingdom |
| | Male | Professor at Bern University of Applied Sciences | University | Switzerland |
| | Male | Postdoctoral researcher at Centre for Social Innovation | NGO | Austria |
| | Female | Postdoctoral researcher at Delft University of Technology | University | The Netherlands |
| Group 3 | Male | Lecturer (Dr) at Chiang Mai University, Chiang Mai | University | Thailand |
| | Female | Employee at Contesi srl, Rende, Italy | Company | Italy |
| | Female | Employee at Contesi srl, Rende, Italy | Company | Italy |
| | Male | Professor at Bern University of Applied Sciences | University | Switzerland |
| | Male | Employee at Contesi srl, Rende, Italy | Company | Italy |
| Group 4 | Male | Professor at Delft University of Technology | University | The Netherlands |
| | Female | Head of research at mySociety | NGO | United Kingdom |
| | Female | PhD Candidate at National University of Ireland Galway | University | Ireland |

## 2.3.3  Workshop results

During the workshop, two people took notes of the scenarios and the requirements that were discussed. All workshop participants could see the notes taken by one person and could immediately respond to those and add requirements. The following scenarios and requirements were presented:

**Scenario by group 1**

Topic: Mass migration in Europe

Description: VREs can be used to investigate the movement/wave of immigrants through Europe.

Data domains: The data that can be used to investigate patterns of immigrant movements included combinations of datasets concerning 1) the use of transportation, 2) the number of smart phones used in certain areas, 3) the amount of garbage produced in certain areas, 4) weather data, 5) asylum processes and 6) combined with other data.

Requirements: The scenario requires the availability of tools to easily visualize patterns and datasets from multiple disciplines. This also requires knowledge about the value of visualizations. For instance, certain types of visualizations are very effective and helpful for representing certain data, but may not be that helpful for other purposes. One of the researchers in group 1 stated that it would be helpful for researchers if some general calculations or data analysis results would be shown in the dataset overview. For instance, in the data catalogue, for each dataset some basic data profiling should take place to show average values, standard deviations and other relevant information to save time for the researcher who needs to browse through many datasets that might potentially be relevant for his/her research. Such data profiling tools are very common in the industry/commercial areas.

Some of the datasets involved might contain sensitive information (e.g. smart phone information) which emphasizes the need of a privacy compliant VRE. The researchers also pointed at the requirement that the VRE should make the researcher aware of possibilities of finding other related data, as the original keywords that the researcher has thought by him or herself may not help the researcher in finding data that he/she has not thought of originally. For instance, the use of weather data may not be obvious to locate immigrant flows, but research has shown that immigrants have moved away from countries like Finland because they found it too cold, so the weather may have an influence here. One researcher called such a recommendation service the "Amazon approach for research data", as recommendations for other datasets and services/tools to be used should be provided. The first scenario concerns explorative research and therefore the researchers mentioned the importance of being able to reset the changes that have been made to a dataset and the VRE should allow for 'playing around'. Furthermore, data versioning has a prominent role – it should be clear what is the original data file and what has changed in modified files and by whom.

In the discussion, the researchers of group 1 mentioned the need for brainstorming tools and bringing together people from different disciplines, as well as funding opportunities. Since the data involved in this scenario may come from different countries, the need for a multilingual VRE was made explicit. In addition, the researchers of group 1 mentioned the importance of the quality of the data and the metadata and to make a distinction between those two. The question was raised whether we can develop tools that will tag data itself and propose additional usages of data. Metadata tools are required. Another requirement concerned making clear which licenses are applicable and how the data can be used compliant with legislation. Collaboration is required for this scenario, such as online brainstorming tools (e.g. MeetingSphere.com, collaboration all over the world) and video conferencing/ audio conferencing tools. Discussion should be embedded in the VRE.

Challenges: An important challenge identified by the researchers of group 1 was that immigrant data can be sensitive and might be misused.


**Scenario 2**

Topic: The organization of a music festival in Vienna.

Description: VREs can be used to identify the risks of incidents at music festivals.

Data domains: The data used in this scenario is 1) event specific reference data (data about the temperature and other weather data, the type of music played at the festival, the type of audience of this festival) 2) general contextual data (e.g. data about health conditions of a certain population, the number of crimes in a certain area, the number of visitors of the festival) and 3) data from comparable other festivals.

Requirements: This scenario requires the availability of large and diverse datasets from various sources. Some datasets are generally available as open data (e.g. the type of music played at a certain festival), while other datasets need to be requested from governments and from the organizers of other festivals. This requires the availability of a data request function in the VRE. Moreover, this scenario does not only involve researchers as data users, but group 2 also mentioned the importance of making the VRE available for citizens in general, as important insights can be obtained for the festival data analysis. Another requirement mentioned was the possibility to trace how datasets are being used in the VRE and where they go. Thus, provenance information was seen as important. Finally, various skills for the data analysis are needed, which may require training of and support for the data users.

Challenges: The data may be misinterpreted as large datasets from different disciplines are involved. Specific skills for the data analysis are needed.


## Scenario 3

Topic: Sustainable and economic well-being in Italy.

Description: In this scenario a researcher studies the effects of pollution on lung cancer in Italy.

Data domains: Health data (e.g. data about the number of deaths from long cancer and environment data (e.g. data about air quality).

Requirements: The researcher in this scenario needs to be very aware that a correlation between high pollution and high numbers of long cancer deaths does not necessarily mean that a causal effect exists. The researcher needs to be pointed at the difference between correlations and causal effects. Training can be used to make the researcher aware of this. In addition, data providers could help the data users with interpreting the data and therefore a collaboration environment for these actors is required. Since policies may influence the number of long cancer death, information about policy interventions on top level need to be available. Other requirements mentioned during the workshop were those of high quality data and clear metadata to describe this data. The metadata should support the correct interpretation of the data. Other data quality aspects that are important are its completeness and a good coverage of the population.

Challenges: The researcher in this scenario needs to be very aware that a correlation between high pollution and high numbers of long cancer deaths does not necessarily mean that a causal effect exists. In addition, in this scenario various factors can influence the number of long cancer deaths, such as policies that make healthcare cheaper and/or more available so that long cancer might be detected earlier and numbers of deaths might be influenced.


## Scenario 4

Topic: Overlaying geographic data with social data.

Description: In this scenario a researcher uses local geographic data (e.g. maps) and adds various datasets from the social sciences to this geographic data. Social science datasets can be combined which may not have been combined before, so that new insights and correlations can be found. The example of the application 'Fix your street' was given to illustrate how geographic and social data can be combined and used.

Data domains: Geographical data and social data (e.g. the number of crimes in a certain area, the number of schools, data concerning social deprivation).

Requirements: One requirement mentioned by the researchers of group 4 was that the system needs to make the researcher aware of interesting data that is available and that could be related to each other. The VRE could provide recommendations of best practices to support the researcher. In order to combine datasets from different domains, the researcher needs to have the skills needed to do this, or needs to be able to learn in the VRE how this can be done. Another requirement mentioned

concerns the maintenance of the data and the VRE system. The data needs to be up to date or even real time.

Challenges: The researchers may not have the skills to visualize and combine various datasets. How to match geographic data and overlay this with social data?

### 2.3.4 Conclusions

This workshop aimed to elicit, refine and discuss requirements for a secure and trusted VRE that integrates OGD and ORD for researchers from multiple disciplines. The workshop attracted the intended audience (researchers as VRE users) and four scenarios were developed. Based on the scenarios, requirements were elicited and refined. The four developed scenarios confirmed most of the requirements that were already identified through previous tasks in VRE4EIC, but they also helped to refine the requirements, for instance by discussing examples of these requirements. In addition, a number of new requirements were mentioned that need to be considered in VRE4EIC. The scenarios can potentially be developed into use cases. The outcomes of this workshop will be used to further refine the user requirements overview and may be used in the use case development.

### 2.3.5 Reference list for the CeDEM workshop

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

Zuiderwijk, A. (2015). *Open data infrastructures: The design of an infrastructure to enhance the coordination of open data use*. 's-Hertogenbosch: Uitgeverij BOXPress.

Zuiderwijk, A., Janssen, M., Yin, Y., Jeffery, K., & Bailo, D. (2016). *Virtual Research Environments: Obtaining new Insights by sharing Open Data for interdisciplinary Research Purposes.* Paper presented at the International Conference for E-Democracy and Open Government, Krems an der Donau, Austria.

Zuiderwijk, A., Jeffery, K., Bailo, D., & Yin, Y. (2016). *Using Open Research Data for Public Policy Making: Opportunities of Virtual Research Environments*. Paper presented at the Conference for E-Democracy and Open Government, Krems an der Donau, Austria.

# 3 Future workshops

As the previous sections have shown, the three workshops already undertaken were planned as part of the VRE4EIC project's dissemination, described in deliverable D7.2. That document shows that two further workshops are planned for the first year of the project. One will be a continuation of an event that took place in September 2015, *Interoperable infrastructures for interdisciplinary big data sciences* (IT4RIs 15)[8]. This event benefits very much from the continuing relationship between VRE4EIC and both ENVRI[PLUS] and EPOS. At the time of writing, details of that event have yet to be finalised but current indications are that it will take place in December 2016. Plans for a workshop on 7 September, however, are fully developed

## 3.1 EGOV 2016, 7 September

On behalf of the project, TU Delft's Anneke Zuiderwijk and Marijn Janssen submitted a proposal to the Dual EGOV 2016 and ePart 2016 conference[9]. Hosted by the UN's University in Guimarães, Portugal, the annual international IFIP EGOV conference is the top-2 ranked core conference in the domain of e-government or ICT in the public sector and the public sphere. As such it receives many submissions and, to be accepted, the VRE4EIC paper had to be merged with another from TU Delft. This reduces the proportion of the workshop that will be devoted to VRE4EIC of course but the impact remains high due to the high profile of the event itself. The full paper is included in the appendices but the key element for VRE4EIC is described in the following extract from the description of the workshop objectives:

> "In the beginning of the workshop we will share the findings of our previous research outlining the proposed taxonomy of data collaboratives, alongside several examples from practice. One of the examples concerns a project about Virtual Research Environments (the VRE4EIC project) in which researchers can collaborate in data analysis and discussions about datasets concerning different societal challenges. The taxonomy will provide participants with insights into the different types of data collaboratives found in various sectors and into the dimensions and characteristics distinguishing them. Using the taxonomy as a point of reference, we will then present our initial framework of influential factors for data collaboratives. The workshop objective is to validate and refine this initial framework of influential factors by asking participants to assess the importance of different factors based on their expertise"

From a VRE4EIC point of view, the workshop is designed to elicit and refine of use cases for a multidisciplinary VRE that integrates open government data and open research data, as well as to promote the project and the potential of VREs to a high level audience.

## 3.2 W3C Workshop: Smart, Interoperable Data Descriptions, Amsterdam, 30 November – 1 December 2016

Plans are also well advanced for a workshop to be held in the project's second year, on 30 November and 1 December at CWI's premises on the Amsterdam Science Park. This will be a full W3C workshop and will therefore focus on the standardisation aspects of the project.  It directly addresses the objectives cited in the DoW:

---

[8] http://zhiming.zh-ao.net/workshop/it4ris/
[9] http://www.egov-conference.org/egov-2016

- collect experiences with relevant technologies to highlight problems, particularly data integration problems, and how these can be overcome
- engage VRE developers, whose software can be enhanced by use of the standards developed by VRE4EIC
- form the core of the W3C group that will continue throughout VRE4EIC

The discussion during the second workshop in Delft concluded that in order to provide a single point of access to multiple research infrastructures, the eVRE will need to be very flexible in the way it operates, handling research infrastructures that do and do not include a catalogue, that do and do not include software and visualisation tools, and that use different metadata schemas (section 2.2.4). These challenges are common in a number of different circumstances in addition to research data such as open government data and cultural heritage. This kind of cross-disciplinary application is essential for if the outputs of the VRE4EIC project are to influence and inform standardisation efforts. Work done directly under the auspices of the European Commission to define application profiles for metadata in data catalogues across Europe (DCAT-AP[10] and related efforts) is also highly relevant therefore.

At the time of writing, detailed plans for the workshop are being finalised but the date and venue are agreed. The call for participation (below) was developed by the VRE4EIC partners with minor modification following feedback from W3C staff. It will be formally published in June 2016. Previous W3C workshops of this type typically attract around 100 participants.

## 3.2.1 Call for Participation

The need to describe data with metadata is well understood: the problem is how best to do it. There are many answers to that question which in itself creates a further problem: with so many standards to choose from, which one should I use to describe my data? With so many in use, which one(s) should I build my application to look for?

The Data Catalogue Vocabulary, DCAT, became a W3C Recommendation in January 2014[11]. Making use of Dublin Core wherever possible, DCAT captures many essential features of a description of a dataset: the abstract concepts of the catalogue and datasets, the realisable distributions of the datasets, keywords, landing pages, links to licenses, publishers etc. But it's clear that DCAT is not a full solution. For example, it doesn't cover versioning or time and space slices; it does not relate semantically the dataset to organisations, persons, software, projects, funding; it describes datasets, not APIs or equipment, and so on. Other well-established and widely used schemas for describing data include CKAN's native schema[12], schema.org[13], DDI[14], SDMX[15], CERIF[16], INSPIRE, and the Healthcare and Life Sciences Interest Group's Dataset Description vocabulary[17]. These provide for discovery of datasets and - in some cases - contextualization (to ascertain relevance and quality) and action (access). Of the above only CERIF provides for provenance, although the W3C Recommendation PROV[18] is also clearly relevant here. To emphasise the variety, the UK's Digital

---

[10] https://joinup.ec.europa.eu/asset/dcat_application_profile/description

[11] https://www.w3.org/TR/vocab-dcat/

[12] http://ckan.org/

[13] http://schema.org

[14] http://www.ddialliance.org/

[15] https://sdmx.org/

[16] http://www.eurocris.org/cerif/main-features-cerif

[17] https://www.w3.org/2001/sw/hcls/notes/hcls-dataset/

[18] https://www.w3.org/TR/prov-o/

Curation Centre - jointly with the RDA's Metadata Standards Catalogue group - manages an extensive catalogue of metadata standards[19] used in different scientific disciplines.

The VRE4EIC project is building a Virtual Research Environment that draws datasets from multiple Research Infrastructures, many of which have data catalogues, and tries to cope with this diversity of methods of data description and data access. It does this partly by drawing on the CERIF metadata schema that provides a mapping of many other schemas. But what is the scalable solution? Where is the line between being flexible enough to meet the needs and preferences of different communities and predictable enough to allow meaningful communication between data publishers and data users?

An application may be able to handle specific metadata schemes or, more precisely, specific profiles of metadata schemes with predefined lists of allowed values, mandatory and optional properties etc. The European Commission, for example, has published a set of application profiles of DCAT that it recommends for communication with European data portals. This suggests a need for metadata publishers and consuming applications to be able to specify which metadata schemes are supported in a machine readable way and to validate data against such as scheme. This is orthogonal to whether the data is provided in JSON, RDF or XML.

A further problem in this space is *vocabulary management*. All the metadata vocabularies and profiles cited above are subject to different change management regimes. What is the right balance between being responsive to the community but stable enough to ensure trust in the vocabulary?

### 3.2.1.1 Workshop aims

This workshop aims to clarify the steps needed to improve communication between data repositories and applications that use that data, such as virtual research environments. Applications may simply discover data or visualize it, manipulate it, discuss it, correct it, describe it republish it etc. The outcome may be a new W3C Working Group chartered to extend DCAT and determine how human and machine-readable metadata profiles are defined and made discoverable.

### 3.2.1.2 Workshop topics

Topics for the workshop include, but are not limited to:

- Approaches to dataset descriptions
- Experience of using DCAT and/or other dataset description vocabularies
- Defining and using metadata profiles
- Discovering metadata profiles
- Providing and using metadata in multiple profiles for multiple contexts.

### 3.2.1.3 Expected participation

To ensure productive discussions, the Workshop will include sessions which are primarily technical, but grounded in business needs. The sessions will be conducted in English; we will do our best to accommodate special needs, but signing and continuous translation will not be available. We invite representatives from following communities to submit papers, although this is not intended to be an exhaustive list:

- Data catalogue operators and users especially from Research Infrastructures
- Data managers/curators
- Virtual research environment developers and users

---

[19] http://www.dcc.ac.uk/resources/metadata-standards

- Developers of data-centric applications

### 3.2.1.4 Programme Committee

The programme includes many of the VRE4EIC partners as well as others with an interest in the topic.

Riccardo Albertoni (CNR)

Martin Alvarez (CTIC)

Phil Archer (W3C)

Kevin Ashley (Digital Curation Centre)

Daniele Bailo (INGV)

Tessel Bogaard (CWI)

Valerie Brasse (EuroCris)

Karen Coyle (DCMI)

Thomas D'Haenens (Informatie Vlaanderen)

Bart De Lathouwer (Open Geospatial Consortium)

Makx Dekkers (DCAT-AP editor)

Martin Doerr (FORTH)

John Erickson, (DCAT Editor, RPI)

Markus Freudenberg (InfAI, University of Leipzig)

Sebastian Hellmann (InfAI, University of Leipzig)

Laura Hollink (CWI)

Antoine Isaac (Europeana/VU)

Keith Jeffery (ERCIM)

Dimitris Kontokostas (InfAI, University of Leipzig)

Carlos Laufer, PUC-Rio

Deirdre Lee (Derilinx)

Fadi Maali, (SAP, DCAT Editor)

Carlo Meghini (CNR)

Hans Overbeek (KOOP)

Andrea Perego (European Commission JRC, GeoDCAT-AP editor)

Valerie Pesce (FAO)

Ronald Siebes (VU/OpenPHACTS)

Herbert Van de Sompel (Los Alamos National Laboratory/DANS)

Jacco van Ossenbruggen (CWI)

Zhiming Zhao (UVA)

# 4 Summary

The project has so far engaged very directly with the ENVR$^{IPLUS}$ and EPOS projects with which it has direct overlap of personnel. This has provided the framework in which the architecture of the eVRE platform is being defined and highlighted the wide variety of views about what defines a VRE. It is clear that there is no well-defined boundary between a research infrastructure and a virtual research environment except that the latter should be a single access point for multiple instances of the former.

The project is firmly rooted in the field of scientific research. However, the benefits of sharing data are being realised in other areas, notably government, cultural heritage and publishing. VRE4EIC workshops have therefore been held in the context of those related disciplines. The need for VREs and research infrastructures to be interoperable with each other is a key use case that will feed into potential formal W3C standardisation efforts.