Using Open Research Data for Public Policy Making
Opportunities of Virtual Research Environments

Zuiderwijk-van Eijk, Anneke; Jeffery, Keith; Bailo, Daniele; Yin, Yi

Publication date
2016

Document Version
Accepted author manuscript

Published in
Proceedings of CeDEM16: International Conference for e-Democracy and Open Government 2016

Citation (APA)
Zuiderwijk-van Eijk, A., Jeffery, K., Bailo, D., & Yin, Y. (2016). Using Open Research Data for Public Policy Making: Opportunities of Virtual Research Environments. In *Proceedings of CeDEM16: International Conference for e-Democracy and Open Government 2016*

Using Open Research Data for Public Policy Making: Opportunities of Virtual Research Environments

Anneke Zuiderwijk
Faculty of Technology Policy and Management
Delft University of Technology
Delft, The Netherlands
a.m.g.zuiderwijk-vaneijk@tudelft.nl

Keith Jeffery
GEIE ERCIM
Biot, France
keith.jeffery@ercim.eu

Daniele Bailo
INGV - Istituto Nazionale di Geofisica e Vulcanologia
Roma, Italy
daniele.bailo@ingv.it

Yi Yin
Faculty of Technology Policy and Management
Delft University of Technology
Delft, The Netherlands
y.yin@tudelft.nl

***Abstract*—Governments and publicly-funded research organisations increasingly make research data available openly. Researchers can use this data in Virtual Research Environments (VREs) to conduct multidisciplinary data-driven research and to obtain new insights potentially for governmental policy-making. However, the requirements for such a VRE are not yet clear. The objective of this study is to elicit and define requirements for a multidisciplinary VRE that integrates Open Government Data (OGD) and open research data for public policy making. Based on a VRE case study, we elicit 13 VRE requirements related to data storage, data accessing, data curation and other aspects, and describe a use case of open data for governmental policy-making. Meeting the requirements results in a VRE that 1) overlays the existing e-Research Infrastructures to provide researchers with integrated open data from different domains, 2) offers OGD in combination with data from publicly-funded research, and 3) stimulates innovation and research collaboration.**

Open data; Virtual Research Environment; VRE; policy making; research infrastructure

I. INTRODUCTION

Increasing numbers of governmental datasets are nowadays becoming available for public reuse [1, 2], and also researchers have access to more and more data opened by the government [3]. In addition, researchers generate research data and, when it is from publicly-funded research, there is often an obligation to make it available openly [e.g., 4]. The two types of open data, Open Government Data (OGD) and open research data, can be integrated with a common user interface, as previously demonstrated by the ENGAGE Project (www.engagedata.eu/about/). Open data can be used by researchers to obtain new insights for data-driven research [5] and to generate new datasets, information and knowledge when data from various sources is combined [6]. Especially open research data and OGD may lead to new combinations. Insights that researchers gain through data-driven research might subsequently be used to improve governmental policy-making [7].

Data driven approaches can be used by researchers to define assumptions, discover evidence and validate theories based on a large quantity of observations, measurements, documents and other forms of data collected from all possible sources. A supporting system should be offered that not only integrates necessary tools for searching, accessing and integrating data and software for different workflows in research activities, but that also provides facilities for collaborations among scientists. We refer to such environments with the term Virtual Research Environments (VREs).

Various projects are already producing e-Research Infrastructures and are reaching towards VREs, such as EPOS for earth/geo-physical sciences (www.epos-eu.org), ENVRI+ for environmental sciences (www.envriplus.eu) and EXCELERATE for biological/biomedical sciences (www.elixir-europe.org/about/eu-projects/excelerate). Yet, researchers who use existing – somewhat primitive - VREs to conduct multidisciplinary research in environmental, earth, social and other sciences often face various problems, such as issues related to data heterogeneity, user experience and the composition of heterogeneous software service components. This complicates research on multidisciplinary societal challenges, such as climate change and energy sustainability. Moreover, to the best of our knowledge, there are no studies on how VREs can be used for public policy making with open data. It is not yet clear what the requirements of a VRE could be that 1) overlays the existing e-Research Infrastructures to provide researchers with integrated open data, software and access to resources such as computers and instruments from different domains, 2) that offers OGD in combination with data from publicly-funded research, and 3) that empowers multidisciplinary research communities and accelerates innovation and collaboration.

The objective of this study is to elicit and define requirements for a multidisciplinary VRE that integrates OGD and open research data for public policy making. In the following sections we provide background information about VREs and existing challenges of using VREs for policy making, the research approach and a case study description. We discuss how open research data and OGD can be integrated for policy-making and draw conclusions of our study.

II. VIRTUAL RESEARCH ENVIRONMENTS (VRES)

A. What are VREs?

VREs have become critical to modern research processes [8]. They have three major components or layers (see Figure 1), namely:

- e-Infrastructures providing Information and Communication Technology (ICT) facilities (e.g. EUDAT, www.eudat.eu/, and PRACE, www.prace-ri.eu),
- 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
- the VRE with its users, who can cooperatively work through the VRE.

VREs support research by interconverting between the multiple underlying e-Research Infrastructures (RIs) supported by e-Infrastructures, while the VRE user neither knows nor cares about the underlying RIs. While VREs depend on e-Research Infrastructures, they are on a higher level of hierarchy than e-(Research) Infrastructures and underlying e-Infrastructures, and provide more advanced functionalities for their end-users. The perspective of users is central to VREs.

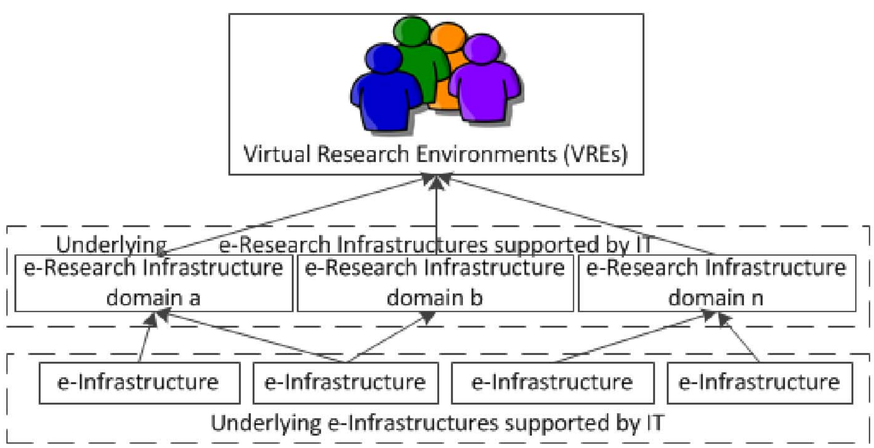

Figure 1. VREs, e-Research Infrastructures and e-Infrastructures

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 [9]. Such research collaborations include a) access to data, tools and 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 [10].

B. *Challenges of using VREs for governmental policy making*

Below we provide an overview of the key VRE challenges related to open data use by researchers and their use for government decision-making. This non-exhaustive overview of challenges is derived from the literature, and has been selected based on VRE experiences of the authors.

- Data context issues. The reuse of research data is challenging outside the borders of an organization [11], and especially beyond scientific disciplines. Accurate metadata associated with documents [12] and datasets [3] 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 [13].
- Data heterogeneity issues. Large amount of data are generated and captured by sensor networks, simulations and instruments from various sources [14]. Open data from different domains can be heterogeneous [15]. This makes large scale integration and interoperability in VREs challenging [16]. VREs typically integrate existing systems and resources [17], 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 [16], and there is a lack of flexibility for the reuse of these components for multidisciplinary research.
- Data quality issues. The quality of information is important for researchers to determine whether they can use a dataset for a particular purpose [3]. A basic set of data quality dimensions includes completeness, accuracy, timeliness and consistency [18]. 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 them 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 interpretation [19]. Privacy sensitivity also depends on combinations of variables and the context in which they are used (idem). Moreover, the combination of data with other sources might make it possible to track the identity of an individual person, especially when it is 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 [19], 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 [20]. Connaway and Dickey [12] 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 [21].
- Technological issues. VREs need to provide access to data, tools, and services [10, 22], through a technical framework that is embedded in a wider research infrastructure [10, 17]. 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 [9], across countries and institutions [22], and across disciplines [21]. 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" [21, p. 70].

The literature has mainly focused on generic challenges for data use in a single discipline. In our case study, we use the generic challenges mentioned above as a framework to identify requirements, and we examine additional challenges for multidisciplinary VREs.

III. Research approach

Yin [23] states that case studies can be used to answer questions handling operational links rather than frequencies or incidence. A case study can be defined as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident" [23, p. 13]. We used a case study approach to elicit and define requirements and elements of a VRE that uses open research data for public policy making. This approach was used since case studies are valuable to explorative research to investigate contextual factors over which the researcher has no or little control [23]. We focus on investigating relations between objects and activities (e.g. how the user interacts with the system) rather than on frequencies, and the case study approach is appropriate for this. Moreover, case studies allow for examining the complex and dynamic roles of the variety of stakeholders involved in using open research data for public policy making (e.g. researchers, policy makers and data providers) and the dependencies between them.

Case studies typically combine various data collection methods [24]. Multiple sources are expected to provide more comprehensive results than a single source, and may help to maximise construct validity and reliability [23]. In our case study, we used documents and interviews as the key information sources. In order to analyze the requirements from our case study in the earth science research sector, the VRE user groups were divided into four key types (although additional types are possible), namely: 1) researchers doing research on earth sciences or related scientific sectors; 2) VRE Data and Service Publishers offering different types of research data and related services; 3) VRE System Administrators managing the VRE system; and 4) General Citizens accessing the research data for any purposes. It's worth of remarking that even though the interviewees belonged to the second and third category (being project managers and IT-experts), they a) were domain-experts and representative of huge communities, b) they had also studied the requirements from the other categories and c) they were asked to answer the interview questions from the perspective of each of these four user groups. Therefore the relative small number of interviews (4 documents were collected in 3 interviews with 4 persons from Italy, Germany and the Netherlands - two individual interviews and one group interview), can be still considered representative because the information provided was already a synthesis of the overall community requirements.

An interview protocol was developed, which 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) [25]. The questions cover each of the five ODP viewpoints: enterprise (science), information, computation, engineering and technology. In addition, the questions concerning activities of VRE users addressed user activities in line with those mentioned in the literature [8, 26]. Moreover, questions from the EPOS, ENVRI+ and EUDATA2020 projects and the FURPS+ model have be reused where appropriate.

IV. Case study description

EPOS (European Plate Observing System) is an e-Research Infrastructure. Integrating data from Solid Earth Science and providing one single platform for the access to heterogeneous datasets and services over the whole Europe is a challenge that EPOS is tackling. EPOS will enable innovative multidisciplinary research for a better understanding of the Earth's physical processes that control earthquakes, volcanic eruptions, ground instability and tsunamis, as well as the processes driving tectonics and Earth surface dynamics. To meet this goal, a long-term plan to facilitate integrated use of data and products as well as access to facilities from mainly distributed (existing and new) Research Infrastructures (RIs) has been designed. The EPOS architecture consists of four complementary elements:

1. **The National Research Infrastructures** (NRIs) contribute to EPOS while being owned and managed at a national level and represent the basic EPOS data providers.
2. **The Thematic Core Services** (TCS) enable integration across specific scientific communities. They represent a governance framework where data and services are provided and where each community discusses its implementation and sustainability strategies as well as legal and ethical issues.
3. **The Integrated Core Services** (ICS) represents the e-infrastructure consisting of services that will allow access to multidisciplinary resources provided by the NRIs and TCS. These will include data and data products as well as, synthetic data from simulations, processing, and visualization tools. The ICS will be composed of the ICS-Central Hub (ICS-C) and distributed computational resources including also processing and visualization services (ICS-D). ICS is the place where integration occurs.
4. **The Executive and Coordination Office** (ECO) is the EPOS Headquarters and the legal seat (ERIC) of the distributed infrastructure governing the construction and operation of the ICS and coordinating the implementation of the TCS.

The main concept is that the EPOS TCS data and services are provided to the ICS. The EPOS components (institutions and services) have been organized in three layers: national layer, community layer, integration layer. These three layers are connected by means of a communication layer called the interoperability layer. This layer contains all the technology to integrate data, data products, services and software (DDSS) from many scientific, thematic communities into the single integrated environment of the Integrated Core Services (ICS). More details about the architecture can be found in the ICS-TCS integration document issued by the EPOS IT team [27, 28].

V. VRE requirements and a use case integrating open data

A. VRE requirements for earth science research

Table 1 depicts the thirteen requirements that appeared to be most important in our case study. These requirements cover the key VRE challenges as described in section 2B. It appears that researchers are the most frequent end-users of a VRE system. Req1, Req5 and Req11 in Table 1 indicate that researchers want to store data such as earth displacement or real-time waveform data of the earthquake from the installed sensors or satellites in an appropriate way. This means the data need to be stored in a known format, described in widely accepted metadata model. Req2, Req3, Req6, Req8 and Req10 show that researchers from different research background, e.g. scientists from earth science communities or environmental science communities also want to get all the recent and historic data about the earth displacement and make analysis on the basis of this data to generate new scientific findings. As a data sharing provision, a VRE encourages researchers to make their research data public. However, researchers still want to control their data and restrict the use or access of their data. Req7 indicates researchers want to receive the research data for their own research activities from trustful organizations or individual researchers. Req4 shows that when researchers share their own data, they want to manage the data, e.g. making changes on the existing data, and tracking the data flow to see who used it and what kind of changes were made.

VRE Data and Services providers (in EPOS the so called TCS) interact with EPOS by providing research data from multi-disciplinary domains and related services. Req5 and Req6 show that VRE Data and Services Publishers want to connect their own VRE to EPOS by providing metadata rich information and by taking advantage of the access to the EPOS ecosystem. Req4 indicates that this user type also wants to monitor the status and changes on his owned data. Req10 shows that VRE data and services publishers want to provide the researchers with specific services according to their requests, for instance, earth science researchers want to use interactive 3D graphics simulating the earth displacement. Also, Req10 points out the need of standardized machine to machine interfaces to allow interoperability of systems. Another example is a researcher needing a computational application to compare synthetic seismograms generated from simulations results with real observations regarding the earth plates activities. Therefore, VRE data and service publishers need a certain service interface such as software APIs to develop their service(s) for potential users, including researchers and other data and service publishers. By offering specific services regarding the research data, e.g. data analytics, simulation, visualization, the providers can charge users. In this circumstance, finical services like accounting (Req12) are needed by the researchers, VRE Data and Services Publisher and VRE administrators.

TABLE I. USER REQUIREMENTS OF A VRE FROM THE EPOS USE CASE.

Number	***Requirements (req.)***	***Researcher***	***VRE Data and Services Publisher***	***VRE system Admini-strator***	***General citizen***
Req1	Quick data storage of big research data from multiple domains (e.g. enormous types and large volume of data covering documents, audio, video, project files, experiments raw data, sensor data)	x	x		
Req2	Access to multidisciplinary data (through browsing, querying datasets or visualisations)	x	x	x	x
Req3	Data computational services (e.g. processing, analysis, visualization)	x	x		
Req4	Data curation (data citation, and data status monitoring)	x	x		
Req5	Data cataloguing (maintain a catalogue of all accessible data)	x	x	x	
Req6	Linkage between RIs from multiple disciplines (e.g. linking existing RIs to EPOS for data uploading and downloading)	x	x	x	
Req7	User identification (user account registration, authen-tication and authorization and identity screening)	x	x	x	x
Req8	Researcher or community collaboration support (e.g. multidisciplinary collaboration support, user communication tools and project management tool)	x	x		
Req9	User communities training and support services	x	x	x	x
Req10	Service interface (e.g. API)	x	x	x	
Req11	Simplicity and ease of use	x	x	x	x
Req12	Accounting service (preferably free of use)	x	x	x	
Req13	Sustainable business model for the long-term operation of a VRE	x	x		

Besides the main functionalities around the data itself for the researchers and VRE Data and Service Provider in a VRE, some additional services are needed. The interviews showed that collaboration support, such as communication tools, social media interaction support or even collaborative project management tools, are required (Req8). Req9 shows that all types of users need training and support service for using VREs, because user satisfaction is essential for gaining all the benefits of using a VRE for research activities. In most cases of the development and the operation of a VRE and the RI is being driven by research project funding [29], some VREs cannot attract enough active users because of lacking research data or useful services. Req13 indicates that an earth science VRE combining data from multiple disciplines needs to have a sustainable business model for long-term operation. Offering data and services can be considered as one potential business model to create tangible financial benefits. Meanwhile, accounting services related to a sustainable business model is very important for a VRE, especially for a VRE like that planned to be associated with EPOS integrating within EPOS different ICS and TCS but also across e-Research Infrastructures such as across EPOS and ENVRI+.

In addition, a VRE can be used as a public service for general citizens who have the right to access the research data. The general citizen can be non-governmental organizations, enterprises or individual citizens. For instance, multi-level governmental authorities need to make land administration policies in terms of more safe and environmental-friendly development of the local economy. The drafting, issuing, implementation and evaluation of the public policies always need to be on the basis of the earth displacement data from the VRE. The most important requirements for general citizens are to search, browse and download research data (Req1, Req2 and Req7).

B. A use case using earth science data for policy making

The EPOS case study provided insight in VRE requirements. In this section we describe a use case to clarify how open research data can be used for government decision and policy making, focusing on researchers as VRE users. In the use case, a governmental agency, namely the Ministry of the Environment in Italy, would like to have insight in all the research conducted on earth displacements. This may, for instance, facilitate the decision about funding allocation in specific high-risk areas to prevent damages caused by natural hazards (e.g. earthquakes). Indeed the funds will likely be spent in earthquakes engineering initiatives (improve building structures resistency to earthquakes), thus allowing also a growth for such regions. A civil servant of the Italian government starts then an in-depth analytical study of safety related to earth displacements and provides information and knowledge to support policy making.

Figure 2 shows a UML use case diagram for using a VRE for policy making, describing the actions and relationship among civil servant, VRE data and service publishers and earth science researchers. In order to perform the study, the civil servant registers in the role of researcher at the VRE. The VRE integrates Research Infrastructures concerning earth sciences, as well as multidisciplinary data from seismic stations, GPS stations, satellite earth observations, geography, climate and other data, since this data may all influence earth displacements. Some of the data is provided by governmental organizations (OGD) as a VRE data service publisher, while other data is provided by publicly funded research organizations (open research data) as a researcher. In this use case the civil servant needs to combine multiple datasets (including OGD and open research data) and perform a multidisciplinary analysis using data-driven services, such as computational tools and software (preferably online) provided by the expert communities (e.g. seismologists and geodesists) as a VRE service provider. The civil servant should then be able to recalculate previous results, produce new results, and to compare them in an integrated way.

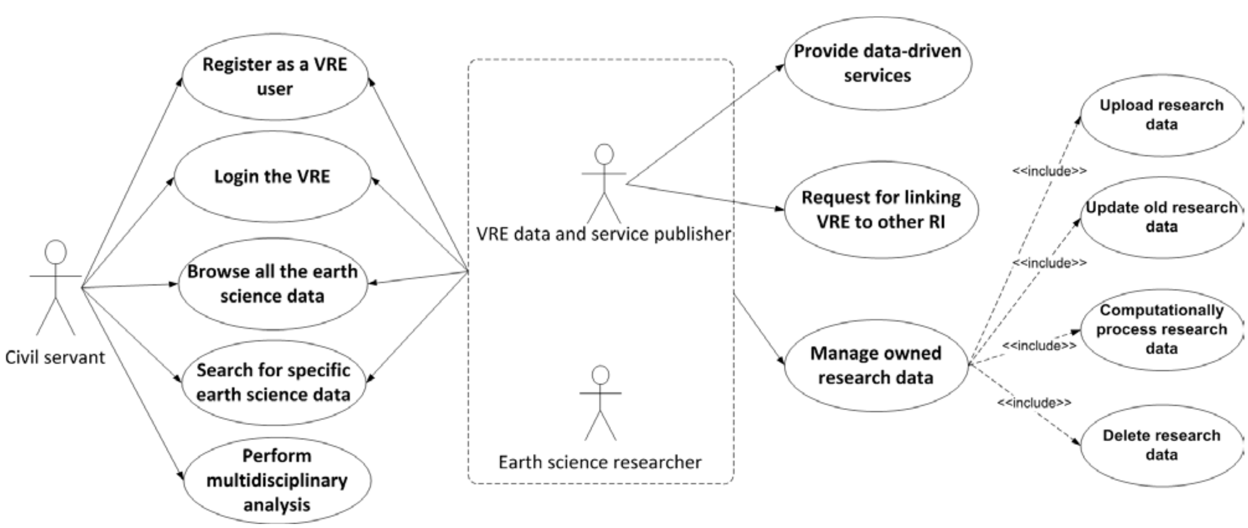

Figure 2. Use case for using earth science data for policy making.

The analysis and comparison of multidisciplinary data is not easy, as it requires much information about the context in which the data was created. Moreover, the integration of the data is often complicated, for example when datasets are provided in heterogeneous formats, when they are on different aggregation levels, or when the semantics are unclear or differ among datasets. Data interpretation is critical, since this is needed to derive information that can be used in government policy making, yet mistakes can be made easily. Allowing for the integration of the heterogeneous datasets and software and maintaining good user experience is challenging, and the integration of large amounts of data raises questions about privacy issues and data sensitivity. Such considerations lead to the conclusion that in order to get valuable and interpretable information, a civil servant needs to take advantage not only of the opportunities created by the VRE, but in addition he may need the support of (a pool of) scientist able to interpret the data and to access to higher level data products (hazards maps, bulletins describing already aggregated data, high level overview of the geological scenario in some regions) more understandable and simple to interpret.

VI. Discussion and conclusions

This study aims to elicit and define requirements for a multidisciplinary VRE that integrates OGD and open research data for public policy making. We conducted a case study, including interviews and document analysis, in the area of earth and environmental sciences. The following 13 VRE requirements were elicited from the case: 1) data storage, 2) data accessing, 3) data computational services, 4) data curation, 5) data cataloguing, 6) linkage between VREs, 7) user identification, 8) researcher or community collaboration support, 9) user communities training and support services, 10) service interface, 11) simplicity and ease of use, 12) accounting service and 13) sustainable business model for the long-term operation of a VRE. Subsequently, a use case for using earth science data for governmental policy making was derived.

Our study provided requirements for a VRE that 1) overlays the existing e-Research Infrastructures to provide researchers with integrated open data from different domains, 2) offers OGD in combination with data from publicly-funded research, and 3) empowers multidisciplinary research communities and accelerates innovation and collaboration. Our study shows that the combination of OGD and open research data might be used for public policy-making, yet using this data to improve public policy-making is accompanied by many barriers. All the challenges found in the VRE literature were also found in our case study. In addition, we found that the VRE needs to integrate multidisciplinary data and provide access to OGD and open research data in an integrated and direct way. The user should be able to fetch the data immediately from the e-Research Infrastructures, having functions to filter and select it. While multidisciplinary research is required for public policy making, most existing so-called VREs do not combine data from multiple disciplines and many of them just harvest and redirect the user to the data source, without providing any services and software to analyse the data and without dealing with heterogeneity issues. The interviews also showed that VREs need to have a sustainable business model for long-term operation. Moreover, the use case showed that policy makers are not interested in the data itself, but in the information that can be derived from the data. They need either data products (analysis of data, maps, bulletins describing the results of a complex scientific analysis) or simple tools for data analysis. Collaborations with other researchers and community support plays an essential role in deriving useful information for governmental policy making.

A limitation of our study is that it focuses on requirements for earth science VREs. Further research is needed to examine whether the VRE requirements are more widely applicable to other sciences. Our case study suggested that this is at least partly possible, since various requirements also seem to be important in other research domains where OGD and open research data play a role. For example, the VRE requirements related to access to data and computational services, and to providing a sustainable business model may also be relevant in other disciplines in which VREs are developed. The requirements need to be detailed further for specific domains and for each of the different VRE user groups. Furthermore, additional user groups may be identified, based on additional interviews. We recommend future research to collect a variety of use cases to detail and evaluate the user requirements. In addition, developing a reference architecture and prototypes to be used for future VREs, including building blocks that can be used to improve existing VREs, could stimulate the use of OGD and open research data to support governmental policy making. This challenge is also addressed by the H2020 VRE4EIC project, which focuses on the key data and software challenges in supporting multidisciplinary data driven sciences. This type of research can stimulate the integration of OGD and open research data to support governmental policy making.

Acknowledgment

This paper is related to the H2020 VRE4EIC project (www.vre4eic.eu) and the H2020 EPOS-IP project (https://www.epos-ip.org/). The authors would like to thank their colleagues of these projects for their input for this paper, although the views expressed are the views of the authors and not necessarily of the projects.

References

[1] S. Kulk and B. van Loenen, "Brave new open data world?," *International Journal of Spatial Data Infrastructures Research,* vol. 7, pp. 196-206, 2012.

[2] G. Magalhaes, C. Roseira, and L. Manley, "Business models for open government data," presented at the International Conference on Theory and Practice of Electronic Governance, Guimarães, Portugal, 2014.

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

[4] European Commission. (2015, December 15). *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

[5] A. K. Krotoski, "Data-driven research: open data opportunities for growing knowledge, and ethical issues that arise," *Insights: the UKSG journal,* vol. 5, pp. 28-32, 2012.

[6] P. F. Uhlir and P. Schröder, "Open data for global science," *Data Science Journal,* vol. 6, pp. 36-53, 2007.

[7] D. C. Esty and R. Rushing, "Governing by the numbers. The promise of data-driven policymaking in the information age," United States2007.

[8] S. Buddenbohm, H. Enke, M. Hofmann, J. Klar, H. Neuroth, and U. Schwiegelshohn, "Success Criteria for the Development and Sustainable Operation of Virtual Research Environments," *D - Lib Magazine,* vol. 21, 2015.

[9] 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.

[10] 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

[11] D. De Roure and C. Goble. (2007, December 1). *myExperiment–a web 2.0 virtual research environment*. Available: http://eprints.soton.ac.uk/263961/1/myExptVRE31.pdf

[12] L. S. Connaway and T. J. Dickey. (2010, January 5). *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* Available: http://repository.jisc.ac.uk/418/2/VirtualScholar_themesFromProjects_revised.pdf

[13] K. Jeffery, A. Asserson, N. Houssos, V. Brasse, and B. Jörg, "From open data to data-intensive science through CERIF," presented at the 12th International Conference on Current Research Information Systems, Rome, Italy, 2014.

[14] A. J. G. Hey, S. Tansley, and K. M. Tolle, *The fourth paradigm: Data-intensive scientific discovery*. WA: Microsoft Research Redmond, 2009.

[15] O. J. Reichman, M. B. Jones, and M. P. Schildhauer, "Challenges and opportunities of open data in ecology," *331,* vol. 11, pp. 703-705 2011.

[16] L. Candela. (n.d., January 5). *Virtual Research Environments*. Available: http://www.grdi2020.eu/Repository/FileScaricati/eb0e8fea-c496-45b7-a0c5-831b90fe0045.pdf

[17] K. Jeffery and A. Asserson, "e-Science, Cyberinfrastructure and CRIS," in *Grey Literature in Library and Information Studies*, D. J. Farace and J. Schöpfel, Eds., ed: De Gruyter, 2010.

[18] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," *ACM Computing Surveys,* vol. 41, pp. 1-52, 2009.

[19] A. Zuiderwijk and M. Janssen, "The negative effects of open government data - investigating the dark side of open data," presented at the Proceedings of the 15th Annual International Conference on Digital Government Research, Aguascalientes, Mexico, 2014.

[20] M. Crosas, "The dataverse network®: an open-source application for sharing, discovering and preserving data," *D-lib Magazine,* vol. 17, p. 2, 2011.

[21] P. Edwards, E. Pignotti, C. Mellish, A. Eckhardt, K. Ponnamperuma, T. Bouttaz*, et al.*, "Lessons learnt from the deployment of a semantic virtual research environment," *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 27–28, pp. 70-77, 2014.

[22] Joint Information Systems Committee. (2011, December 1). *Virtual Research Environment Programme*. Available: http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx

[23] R. K. Yin, *Case study research. Design and methods*. Thoasand Oaks: SAGE publications, 2003.

[24] K. M. Eisenhardt, "Building theories from case study research," *Academy of Management Review,* vol. 14, pp. 532-550, 1989.

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

[26] D. De Roure, C. Goble, and R. Stevens, "The design and realisation of the Virtual Research Environment for social sharing of workflows," *Future Generation Computer Systems,* vol. 25, pp. 561-567, 2009.

[27] D. Bailo, K. Jeffery, A. Spinuso, and G. Fiameni, "Interoperability Oriented Architecture: The Approach of EPOS for Solid Earth e-Infrastructures," presented at the IEEE 11th International Conference on e-Science, Munich, Germany, 2015.

[28] EPOS-IP, "ICS-TCS Integration Guidelines - Handbook for TCS integration: Level-2," 2015.

[29] M. Fraser. (2005, January 5). *Virtual research environments: overview and activity*. Available: http://www.ariadne.ac.uk/issue44/fraser