Open science, open government and open data: creating an impact through open online education and Virtual Research Environments

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

Publication date
2018

Document Version
Accepted author manuscript

Published in
Open Education Global Conference 2018: OEGlobal2018

Open Education Global 2018 Conference - *Transforming Education through Open Approaches*

Open science, open government and open data: creating an impact through open online education and Virtual Research Environments

Authors
Anneke Zuiderwijk, Delft University of Technology
Keith Jeffery, the European Research Consortium for Informatics and Mathematics (ERCIM)
Yi Yin, Delft University of Technology

Keywords: open data, open science, open government, open online education, Virtual Research Environments, VRE

1. Introduction

Virtual Research Environments (VREs) are online environments providing access to data, software and resources of various e-research infrastructures (Bornschlegl, Manieri, Walsh, Catarci, & Hemmje, 2016; Candela, Castelli, & Pagano, 2013; Zuiderwijk, Jeffery, Bailo, & Yin, 2016). They offer advanced functionalities and tools to facilitate collaboration between researchers that may be geographically dispersed (Sarwar, Doherty, Watt, & Sinnott, 2013, p. 551). For instance, they support cooperation at intra- and inter-institutional levels and offer mechanisms for data preservation (Carusi & Reimer, 2010). VREs often provide access to open datasets collected or created by researchers both from governmental and non-governmental institutions, usually connected to the related open access publications.

VREs have enormous potential. They can support research excellence by providing innovative services related to metadata, data context and trust generation, data analytics, scientific publications, researchers' collaboration, interoperability, and network, computing and storage, semantics and data preservation (see www.vre4eic.eu). They offer the opportunity to improve the effectiveness and efficiency of researchers. VREs can also span across boundaries. For example, they can potentially support cooperation or collaboration between researchers from different domains and between researchers and policy makers, and they allow different types of data to be combined (e.g. research data from one domain to be combined with research data from other domains or research data to be combined with government data) (Zuiderwijk, 2017). Making new data combinations and setting up new collaborations stimulate innovation (Jeffery et al., 2017) and obtaining novel insights that may provide input for policy makers to solve societal problems. Furthermore, VREs contribute to the open science movement by offering an open environment in which open data and open research are facilitated.

Despite the large potential, creating an impact with VREs is a challenge. Many researchers are already used to working in a certain way and may not be aware of existing VREs. Moreover, even if researchers are aware of VREs, they may not know how to use them for their research purposes and this may require training. To realize the potential of VREs, the user experience is critical (Crosas, 2011) and VREs must have an active and large user community. The use of VREs by a limited number of researchers or the disregard of VREs can lead to their destruction. Solving these challenges is of critical importance for the outreach and impact of VREs. If these challenges are not addressed, the envisioned benefits of VREs and their contribution to open science may not or only partly be attained.

The objective of this paper is twofold. First, we aim to discuss how the elements of one particular VRE stimulate openness in relation to open science, open government, and open data. Second, we aim to explain how the reach and impact of these openness elements can be

strengthened and increased through open online education. The H2020 VRE4EIC project (an abbreviation for: *A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration*) will be used as an example, as this project is already developing a Europe-wide interoperable VRE to empower multidisciplinary research communities and accelerate innovation and collaboration (www.vre4eic.eu). This project is already addressing many of the above-mentioned challenges.

This paper is structured as follows. In the following sections we first elaborate on the VRE elements that stimulate openness. Subsequently, we describe how open online education can be used to strengthen and increase the reach and impact of the openness elements. Then we connect VREs to the domains of Open Government, Open Data and Open Science by using open online education. Thereafter we describe some outcomes from the course evaluations. And, finally, we will draw lessons from our study and provide recommendations for similar VRE-related projects.

2. VRE elements stimulating openness

VREs typically consist of three main components (Zuiderwijk, 2017; Zuiderwijk et al., 2016):

1. "The bottom layer: the e-infrastructures that provide information and communication technology (ICT) facilities (e.g. EUDAT, www.eudat.eu/ and PRACE, www.prace-ri.eu).
2. The middle layer: the e-research infrastructures that provide homogeneous access to heterogeneous data, software and resources of a range of e-research infrastructures for end users (e.g. LifeWatch, www.lifewatch.eu/). They also offer services and applications (Terras, Warwick, & Ross, 2016).
3. The top layer: the VRE itself with its users, who can work together and collaborate through the VRE" (Zuiderwijk, 2017).

This definition has also been adopted in the VRE4EIC project. Figure 1 depicts the objective of the VRE developed in this project, namely to empower multidisciplinary research communities and to accelerate innovation and collaboration. The project develops a reference architecture and software components for Virtual Research Environments (VREs). The combination of these components bridges across existing e-Research Infrastructures that are used in various disciplines, such as social sciences, physics, environmental sciences and life sciences. Examples of such e-Research Infrastructures are the EPOS Research Infrastructure for Solid Earth Sciences[1] and the ENVRI+ e-Research Infrastructure for Environmental Sciences[2], both represented in the project.

[1] https://epos-ip.org/
[2] http://www.envriplus.eu/

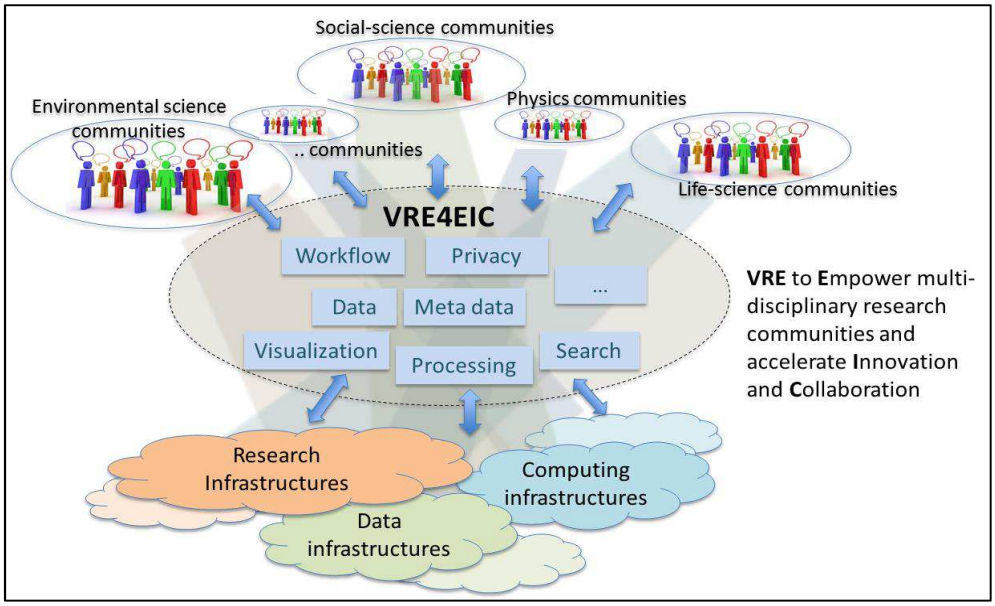

Figure 1: VRE4EIC overview.

The interdisciplinary VRE developed in the project provides a comfortable homogeneous interface for users by virtualizing access to the heterogeneous datasets, software services, resources of the e-Research Infrastructures and also provides collaboration and communication facilities for users to improve research communication. It also provides access to research management and administrative facilities so that the end-user has a complete research environment.

VRE elements under development in the VRE4EIC project that allow for stimulating openness include:

- Mechanisms for enhanced trust:
    a. Keeping track of changes made to datasets. This is to ensure that users of datasets can see whether a certain dataset is the original one or whether changes have been made. If changes have been made, they should be able to see what has been changed to the dataset compared to the original version. This requires interoperable metadata standards to record version information.
    b. Identifiers and citations for different versions of a certain dataset.
    c. Providing provenance information, such as information about the origin of a dataset, who created or collected the data, and which methods have been used to create or collect the data. This requires interoperable metadata standards to record provenance information.
    d. Ensuring that datasets will be available for a long time and that they will not disappear.
- Mechanisms for enhanced security:
    a. Securing access through various identity providers.
    b. Secure storage, backup and the transmission of data at the level of e-Research Infrastructures and the VRE.
    c. Providing metadata about the level of security of a dataset.
- Networking tools integrated with existing and dedicated social media. This allows users of the VRE to collaborate in the analysis of data.
- Interoperable workflows. Workflows refer to a set of activities that are often conducted in combination and that might be reusable in various contexts. An example concerns a statistical test in which the user takes similar steps.
- Interoperation of heterogeneous e-Research Infrastructures leading to new science. The VRE that is under development allows for interoperating between e-Research Infrastructures

from multiple disciplines. For example, combining data from the social sciences with data from environmental sciences could lead to new insights.

- An improved metadata model. Comprehensive information about the resources in the VRE will be provided. The CERIF metadata model is used to provide rich metadata (euroCRIS, 2018; Jeffery, Houssos, Jörg, & Asserson, 2014).
- Mechanisms to improve the user experience, such as the possibility to create groups that can collaborate in data analysis and interpretation, the possibility to search for and find researchers in other scientific domains and to contact them through the VRE.
- Services to apply data mining and text mining to datasets through the VRE.
- Tools to model and visualize datasets available through the VRE.
- Models for the Description of data and service resources according to existing standards, such as Semantic Web standards and recommendations developed by working groups of the World Wide Web Consortium (W3C).
- Application Programming Interfaces (APIs) for developers to include new e-Research Infrastructures to the research environment and to connect them to other e-Research Infrastructures, so that multidisciplinary research is facilitated.

The VRE developed in the VRE4EIC projects offers its end users with access to resources, such as data and software, from a variety of e-Research Infrastructures. Often this data is a type of *open* data, meaning that it is structured, machine-readable data that individuals and organizations actively publish on the internet for public reuse and that can be accessed without restrictions and used without payment (European Commission, 2011, 2013; Geiger & von Lucke, 2012; Gurin, 2014; Open Knowledge Foundation, 2015). Open data can be provided both by researchers (as studied in the area of Open Science) and by governments (as studied in the areas of Open Government and Open Data). Thus, VREs are related to multiple research areas including Open Science, Open Government and Open Data. There is much potential for connecting VRE elements to these other research areas, as will be described in the next section.

3. Online courses for open connections

To strengthen and increase the reach and impact of the above-mentioned openness elements we developed two online courses:

1. A Massive Open Online Course (MOOC) concerning Open Government (Delft University of Technology, 2018), offered through the EdX platform (www.edx.org). This MOOC was aimed at a large audience, ranging from students to professionals, researchers and policy advisors. It covered the topics of an introduction to Open Government, an introduction to Open Government Data, and public values in Open Government. In the developed MOOC participants collaborated, for example, by sharing cases of openness in governments on a forum and by analyzing these cases through a structured survey. They also evaluated the cases and peer-reviewed each other's cases.
2. An online course about Open Data (Delft University of Technology, 2017), offered through the TU Delft Professional Education platform (www.profed.tudelft.nl). This course was aimed at professionals who had to pay for most elements of the course and hence it attracted a smaller audience. However, the videos were available to anyone and were shared with a broader audience. This course covered the topics of governance issues related to open data use and open data policy making, including data quality, privacy, sustainability, metadata and other topics. The course participants could choose one out of two tracks (policy making or development), or they could choose to follow both.

Interaction was highly stimulated in both courses. Each course consisted of videos, papers, quizzes (to check whether participants had understood the content of the videos and papers), assignments

and a final exam. Each of these course elements was graded and all the grades together contributed to a final grade. If the final grade was at least six out of ten the learner had passed the course and could get a certificate. Both courses ran in 2016 and 2017. More than 8,000 learners enrolled for these courses in 2016 and 2017 in total (see Table 1).

Online course	# learners	# countries
Open Government MOOC 2016	4,928	159
Open Data course 2016	17	13
Open Government MOOC 2017	3,102	151
Open Data course 2017	36	5
Total	8,083	NA

Table 1: Participants of the two online courses: Open Government and Open Data.

4. Connecting VREs to open data, open government and open science through open online education

The two online courses allowed for connecting several themes related to openness, including open government, open data and open science. Together these courses offered the participants videos and assignments that provided the necessary background to understand elements of openness in relation to VREs. For instance, videos concerning basic technologies for Linked Open Data (see Figure 1) were offered, as well as videos data about W3C recommendations and standards for data portals in Europe (see Figure 2). Such technologies and standards are important for the interdisciplinary VRE that is being developed in the VRE4EIC project, since this VRE should give access to open datasets provided by many different research infrastructures in different research disciplines that each use their own standards. The online courses made it possible to strengthen the reach of the VRE4EIC project and to provide more researchers and citizens with information and training material concerning open data, open government and open science.

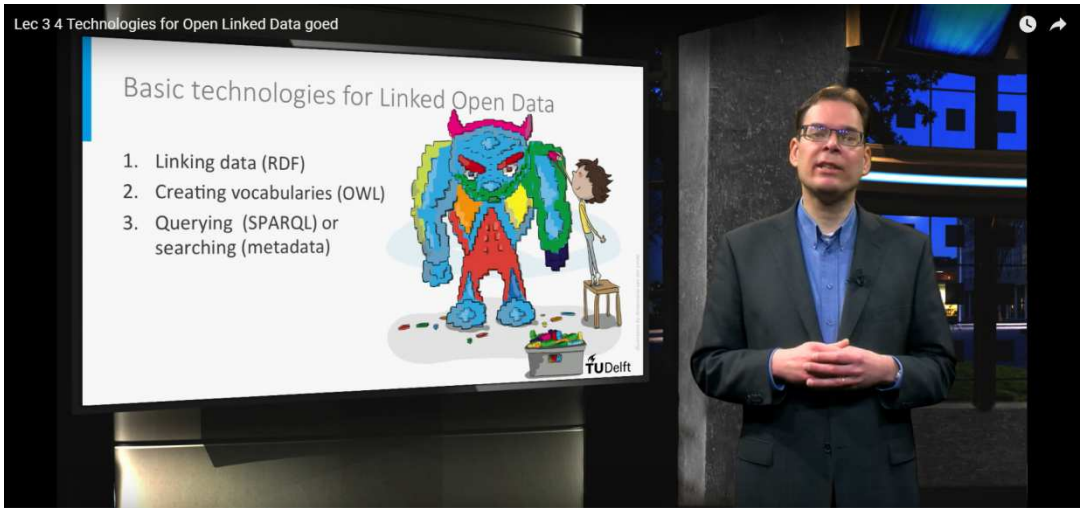

Figure 2: Video about Linked Open Data

In many of the videos the presenters made the connection between the topics they discussed and how the VRE4EIC project was implementing these openness elements in their VRE. For example, in a video about linked open data technologies the presenter stated “Do choose a standard rather than making up your own way of doing things. One of the things we’re doing in the VRE4EIC project is to help bridge between these standards.” (see Figure 3)

Figure 3: Video about Linked Open Data technologies

Open connections were also stimulated through interactive assignments. As part of one of their assignments, participants were encouraged to search for an Open Government initiative themselves, to describe it according to a structured template, and to share their case descriptions on a forum. Subsequently, the MOOC participants were asked to read each other's case descriptions and to vote for the top three descriptions. This resulted in interactive discussions between the participants about how open government cases should be understood and how they differed between countries. Figure 4 shows an example of such a discussion (anonymized).

Discussion topic: Digital government
Posted in May 2017 by [Participant A]

I agree that the approach should be to eradicate the distinction between "government" and "digital government", so that eventually it's all simply "government". However, I hope that we will be conscious of the digital divide even as we progress so that those who are not technologically savvy, lack access to technology, or prefer not to conduct so much of their life online (for privacy issues and otherwise) will not be ignored.
Post related to Week 1 / Week 1 Questions
This post is visible to everyone

> Response by [Participant B] - Posted in May 2017
> I think public libraries could help here by supplying internet access and offering support with electronic government processes such as applying for a passport, etc.
>
> > Response by [Participant C] - Posted in May 2017
> > In Scotland all libraries must provide public access to the internet however most government agencies are not providing vast amounts of open data.
> >
> > > Response by [Participant D] – Posted in May 2017
> > > In Turkey all libraries do not provide public access to the internet. However, we do have e-Government (e-devlet) here in Turkey!

Figure 4: Example of a discussion by Open Government MOOC participants.

It should be noted that the discussions on the course forum did not yet focus on Virtual Research Environments in particular, since these were not part of the assignments. Nevertheless, there is potential for making a stronger connection between the assignments, VREs and the VRE4EIC project, so that the impact of VREs can be increased further.

5. Evaluation of the courses

The two online courses were evaluated in various ways. First, we received feedback through the forum. An example of such feedback can be seen in Figure 5.

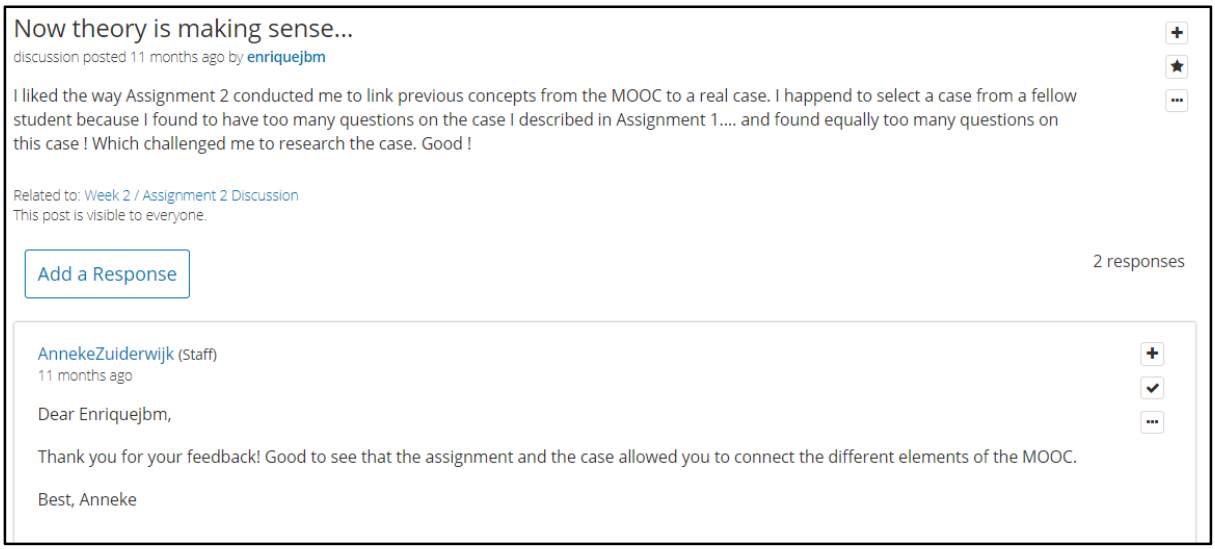

Figure 5: Example of feedback provided on the forum used for the Open Government MOOC.

We also conducted more structured evaluation. All participants who had enrolled for the course at the beginning were sent a survey before the course started, in the middle of the course and after the course had ended. By doing so we could obtain information about the background of the course participants and their motivations for participating (pre-course survey), if they liked the course when it was just running for two weeks (mid-course survey), and whether the course met their expectations and what could be improved (post-course survey). The pre-course and post-course surveys were relatively comprehensive (about 30 questions), whereas the mid-course survey online consisted of three questions.

Both courses were evaluated positively. For example, the Open Government MOOC received an overall grade of 8,08 in 2016 and 8,25 in 2017 (out of 10), although this number is difficult to interpret. Course participants particularly liked the assignments in which they analysed real-life cases, that we had a combination of flexible and strict deadlines and that we gave the participants access to academic papers that usually were behind a paywall. These positive evaluations are confirmed by the following quotes that we took from the evaluation surveys:

> *"It was really great to look outside my own country and see the open initiatives and portals from other places. Metadata and the various standards – the potential of APIs and data visualization."*
>
> *"Papers, assignments, exams and video classes together result in an easy and motivating learning experience."*
>
> *"It was well structured and interesting. Breaking everything down into bite sized chunks made it very manageable."*
>
> *"It was more interactive than I'd anticipated and had a great deal more assessment than I was expecting."*
>
> *"I saw that the instructors and professors were engaged and answering all questions in the forums."*

The participants also mentioned several aspects that could be improved, namely the navigation on the forums, the assignments in which peer reviewed each other's work (there was an imbalance

since some participants received a very thorough review whereas others were very brief) and little attention for legal aspects in the course. Some participants thought the course was too easy, others thought it was too difficult. The later also shows that it is important to be very clear at the beginning of the course about what participants can expect and what they should not expect to learn. Although we provided much information about the learning objectives and topics covered in the courses, this should be extremely clear.

6. Recommendations to other VRE-related projects

Recommendations for how other VRE-related projects in the area of Open Government, Open Data and Open Science may also connect different openness elements are as follows.

- *Record short videos (6-10 minutes) explaining the basics of VRE-related concepts*, such as data infrastructures, data quality, factors influencing trust, privacy enhancement mechanisms and metadata standards. This helps MOOC participants to understand the idea and potential of VREs. Participants appreciated the short videos provided in the course. For instance one respondent stated: "It was well structured and interesting. Breaking everything down into bite sized chunks made it very manageable."
- *Record short videos (6-10 minutes) explaining how the VRE can be used.* These videos should be tutorials that provide participants with real examples of the VRE advantages. We have not yet created such videos in our courses, but we are planning to do so in the following year.
- *Connect the video content to engaging assignments.* As stated by one of the respondents, "papers, assignments, exams and video classes together result in an easy and motivating learning experience." This combination that is typical for open online education may engage large numbers of participants.
- *Connect the assignments to the VRE.* The assignments required participants to work with existing initiatives and data sharing platforms. The online courses contained examples of how data sharing and use can be improved, with specific reference to the VRE developed by the VRE4EIC project. This can enhance the user base of VREs.
- *Let course participants share their findings with each other on a forum and learn from each other*. The first assignment of our MOOC requested participants to search for cases of open government and open data in their countries. Since the MOOC participants can from more than 150 different countries, this resulted in a comprehensive and interesting overview of how openness is taking place world-wide and which barriers are faced in certain countries and for certain openness initiatives. This was a positive experience for both the MOOC organizers and the participants. For example, one respondent stated: "It was really great to look outside my own country and see the open initiatives and portals from other places. Metadata and the various standards – the potential of APIs and data visualization."
- *Disseminate the recorded videos and tutorials widely.* The recorded videos and tutorials are published on the project website and are also disseminated through social media, to reach an audience as large as possible.
- *Engage with the course participants*. The evaluations showed that the course participants liked that the instructors were very much involved in the course. For instance, they actively responded to questions and discussion posts on the discussion forum and they provided general feedback to all the course participants through feedback videos that were recorded just after participants had submitted their assignments. In this way, the course remained interactive and participants were also assessed in non-graded ways.

Connecting the openness elements of the VRE to open education strengthens the reach and increases the impact of our VRE project. Leveraging VRE open elements with online education provides open access affecting the use of open research data and open government data.

7. Future work

We recently decided to split the learning material concerning Open Government, Open Data and Open Science into different parts. We decided to focus the Open Government MOOC more and to develop two other MOOCs that would specifically pay attention to Open Data and Open Science. These two new MOOCs are currently under development and the Open Science MOOC is expected to start by the end of 2018. We expect that splitting up the MOOC into three shorter but more focused MOOCs has several advantages:

- Learners can obtain more in-depth insight concerning each of the three topics (Open Government, Open Data and Open Science) and collaborate with peers with similar interests.
- Learners may at first be interested in one of the three topics but by connecting the three topics their interest for these other topics may be raised.
- Practically it is easier for learners to participate in three short MOOCs spread throughout the year than to participate in one long MOOC, since the latter is more difficult to combine with their busy daily life.

8. Conclusions

This papers aims to: 1) discuss how the elements of one particular VRE stimulate openness in relation to open science, open government, and open data, and 2) explain how the reach and impact of these openness elements can be strengthened and increased through open online education. Openness elements that were described are related to metadata, data context and trust generation, data analytics, scientific publications, researchers collaboration, interoperability, network, computing and storage, semantics and data preservation. We described how the reach and impact of VRE openness can be improved by 1) recording short videos explaining the basics of VRE-related concepts, 2) recording short videos explaining how the VRE can be used, 3) connecting the video content to engaging assignments, 4) connecting the assignments to the VRE, 5) letting MOOC participants share their findings with each other on a forum and learn from each other, 6) disseminating the recorded videos and tutorials widely and 7) engaging with the course participants.

Using open online education in relation to our VRE project allowed us to reach thousands of participants that might otherwise not have been reached through traditional means, such as journal papers, conference presentations and workshops. We are working on connecting the Open Government MOOC to two other MOOCs, namely an Open Data MOOC and an Open Science MOOC. This should make it possible for learners to learn more about the different elements of openness. We recommend VRE-related projects in the area of Open Government, Open Data and Open Science to also use open education as a means to connect the different openness elements. Connecting openness elements of VREs to open education allows for strengthening the reach and increasing the impact of VRE projects. Leveraging VRE open elements with online education provides open access affecting the use of open research data and open government data.

Acknowledgement

This paper is related to the H2020 VRE4EIC project (www.vre4eic.eu). The authors would like to thank their colleagues of this project for their input for this paper, although the views expressed are the views of the authors and not necessarily of the project.

Reference list

Bornschlegl, M. X., Manieri, A., Walsh, P., Catarci, T., & Hemmje, M. L. (2016). *Road mapping infrastructures for advanced visual interfaces supporting big data applications in virtual research environments*. Paper presented at the Workshop on Advanced Visual Interfaces AVI.

Candela, L., Castelli, D., & Pagano, P. (2013). Virtual Research Environments: An Overview and a Research Agenda. *Data Science Journal, 12*, GRDI75-GRDI81. doi:http://dx.doi.org/10.2481/dsj.GRDI-013

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

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

Delft University of Technology. (2017). Open Data Governance: from Policy to Use. Retrieved from https://online-learning.tudelft.nl/courses/open-data-governance-from-policy-to-use/

Delft University of Technology. (2018). Open Government. Retrieved from https://online-learning.tudelft.nl/courses/open-government/

euroCRIS. (2018). Main features of CERIF. Retrieved from https://www.eurocris.org/cerif/main-features-cerif

European Commission. (2011). Digital agenda: Commission's open data strategy, questions & answers. Retrieved from http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/891&format=HTML&aged=1&language=EN&guiLanguage=en

European Commission. (2013). Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the Re-use of Public Sector Information Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF

Geiger, C. P., & von Lucke, J. (2012). Open government and (linked) (open) (government) (data). *Journal of e-Democracy and Open Government, 4*(2), 265-278.

Gurin, J. (2014). *Open data now. The secret to hot startups, Smart investing, savvy marketing, and fast innovation*. New York: Mc Graw Hill Education.

Jeffery, K., Houssos, N., Jörg, B., & Asserson, A. (2014). Research Information Management: The CERIF Approach. *International Journal of Metadata, Semantics and Ontologies, 9*(1), 5-14.

Jeffery, K., Meghini, C., Concordia, C., Patkos, T., Brasse, V., Ossenbruck, J. v., . . . Marchetti, E. (2017). *A Reference Architecture for Virtual Research Environments*. Paper presented at the 15th International Symposium of Information Science, Humboldt-Universität zu Berlin, Germany.

Open Knowledge Foundation. (2015). Open Definition version 2.0. Retrieved from http://opendefinition.org/od/

Sarwar, M. S., Doherty, T., Watt, J., & Sinnott, R. O. (2013). Towards a virtual research environment for language and literature researchers. *Future Generation Computer Systems, 29*(2), 549-559.

Terras, M., Warwick, C., & Ross, C. (2016). Building Useful Virtual Research Environments: The Need for User-led Design. In P. Dale, J. Beard, & M. Holland (Eds.), *University Libraries and Digital Learning Environments* (pp. 151). London: Routledge.

Zuiderwijk, A. (2017). Analysing open data in virtual research environments: New collaboration opportunities to improve policy making. *International Journal of Electronic Government Research, 13*(4), 76-92. doi:10.4018/IJEGR.2017100105

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.