International Journal of Electronic Government Research
Volume 13 • Issue 4 • October-December 2017

Analysing Open Data in Virtual Research Environments:
New Collaboration Opportunities to Improve Policy Making

Anneke Zuiderwijk, Delft University of Technology, Delft, Netherlands

ABSTRACT

This article describes how virtual research environments (VREs) offer new opportunities for researchers to analyse open data and to obtain new insights for policy making. Although various VRE-related initiatives are under development, there is a lack of insight into how VREs support collaborative open data analysis by researchers and how this might be improved, ultimately leading to input for policy making to solve societal issues. This article clarifies in which ways VREs support researchers in open data analysis. Seven cases presenting different modes of researcher support for open data analysis were investigated and compared. Four types of support were identified: 1) 'Figure it out yourself', 2) 'Leading users by the hand', 3) 'Training to provide the basics' and 4) 'Learning from peers'. The author provides recommendations to improve the support of researchers' open data analysis and to subsequently obtain new insights for policy making to solve societal challenges.

KEYWORDS

Collaboration, OGD, Open Data, Open Government Data, Use, User Support, Virtual Research Environment, VREs

INTRODUCTION

Access to data, and particularly to open data, can be beneficial for science, public policy and society, and can inform evidence-based governmental decisions (Sá & Grieco, 2016; Sivarajah et al., 2016). Open data may be obtained from governments, publicly funded research organisations and private organisations (Vercamer, Steurtewagen, Van den Poel, & Vermeulen, 2016), as well as from social media (Kalampokis, Hausenblas, & Tarabanis, 2011; Poel et al., 2015) and sensors (Poel et al., 2015). For instance, open government data (OGD) concerning the design, construction, management and maintenance of the road network can be combined with data from GPS navigation companies and with social media data from individual drivers to identify issues related to traffic congestion. The combination of these data obtained from different sources and disciplines may be used to improve traffic policy making and ultimately to decrease traffic congestion. Valuable insights are expected to be derived from novel combinations of data from different disciplines (Choi & Tausczik, 2017). Open data can be analysed by anybody, and the analysis results can be used to make informed arguments

DOI: 10.4018/IJEGR.2017100105

for embracing, rejecting or proposing new or improved policies (Janssen & Helbig, 2016). Open data can then be used for evidence-based policy analysis and evaluation (Markaki et al., 2014).

Just as new data combinations offer new opportunities (Janssen, Konopnicki, Snowdon, & Ojo, 2017), collaboration between the various actors involved in analysing open data is also expected to facilitate new opportunities to obtain new insights to solve societal problems (Susha, Janssen, & Verhulst, 2017). Collaboration is an important aspect of many open data analysis projects (Choi & Tausczik, 2017). For example, to improve policy making regarding energy sustainability, policy makers need not only data from a range of domains (e.g. energy, industry, pollution, climate, weather, housing, geography), but also the ability to interpret the data. Collaboration to combine the relevant domain knowledge of different parties (e.g. energy saving experts, pollution experts, climate experts) is needed to investigate and interpret the data and to take appropriate measures.

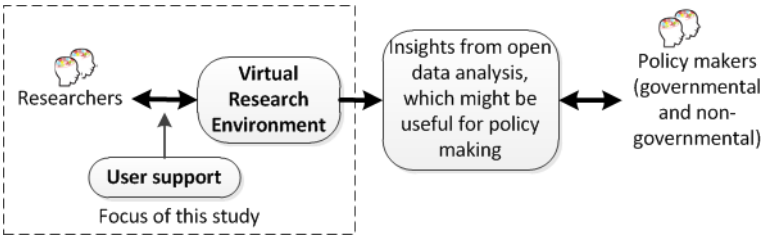
Virtual research environments (VREs) offer new opportunities to collaboratively analyse data and obtain new insights, especially when data from multiple disciplines are combined (Jeffery et al., 2017). Such insights may provide input for public policy making to solve societal issues (Zuiderwijk, Jeffery, Bailo, & Yin, 2016). VREs provide researchers with access to the resources, including data and software, of a multiplicity of e-research infrastructures. Although various VRE-related initiatives are under development, they suffer from user experience issues (Zuiderwijk et al., 2016). There is a lack of insight into how VREs support collaborative open data analysis by researchers and how this might be improved. The potential of VREs and open data is largely unexploited.

This paper clarifies in which ways VREs support researchers in open data analysis. It presents an analysis of VRE cases that employ a variety of support modes for open data analysis in Europe. As such, the contribution of this paper is in the form of an overview of and recommendations for user support for data analysis by researchers using VREs. The user support modes and the recommendations may be used by technical developers in the design and development of VREs, and they may be considered by policy makers of governmental and non-governmental organisations in the creation of open data policies. This paper focuses particularly on VREs that support researchers analysing open data (see Figure 1). Open data includes data from public and publicly-funded organisations (OGD), as well as research data from research organisations and data from private parties, such as data obtained from SMEs and from citizens (social media data).

Policy makers were outside the scope of this study, and the premise was that insights from open data analysis through VREs will support policy making. We focused on revealing functionalities rather than the quality of the data that is provided through the VRE. Moreover, the meeting of requirements, such as reliability and scalability, and how users learn to use VREs were also outside the scope of this study.

This article is structured as follows. First, related literature in the areas of VREs and support for OGD use is discussed. Then the approach used in this study is outlined. This is followed by a cross-case comparison of modes to support OGD use for policy making. Finally, recommendations and conclusions for VRE developers and policy makers are discussed.

Figure 1. Open data analysis in virtual research environments



RESEARCH BACKGROUND

Virtual Research Environments

VREs have become critical to modern research processes (Buddenbohm et al., 2015). VREs are created to "support collaboration, encourage multidisciplinary research, allow the use (and reuse) of data, and facilitate the research environment" (Terras, Warwick, & Ross, 2016, p. 153). A VRE can be defined as "an online environment offering a set of tools aimed at providing a collaborative research environment for researchers that may be geographically dispersed" (Sarwar, Doherty, Watt, & Sinnott, 2013, p. 551). VREs typically consist of three main components (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 et al., 2016);
3. **The top layer:** The VRE itself with its users, who can work together and collaborate through the VRE (Zuiderwijk et al., 2016).

Through the VRE, the end user can easily access the data, software and resources of a range of underlying e-research infrastructures without knowing about the underlying e-research infrastructures (Zuiderwijk et al., 2016). Access is provided through a single, uniform web portal that minimises the level of detail and complexity of the back-end process, such as interaction with services and handling heterogeneous data resources (Sarwar et al., 2013). VREs are at a higher level of hierarchy than e-research infrastructures and their underlying e-infrastructures, and they offer more advanced functionalities for their end users than the underlying e-research infrastructures do (Bornschlegl, Manieri, Walsh, Catarci, & Hemmje, 2016; Candela, Castelli, & Pagano, 2013; Zuiderwijk et al., 2016). VREs should assist data users throughout the "scholarly processing cycle", which consists of 1) conditioning and augmentation by metadata, 2) modelling and creating annotated linked sources, 3) applying data analysis and inference tools, and 4) releasing, presenting and publishing the results in various formats (Scholz & Goerz, 2012). Thus, VREs should provide the complete research environment.

VREs are focused mainly on research collaborations, in addition to the support for all aspects of research activities (Sinnott & Stell, 2011). Other research supporting environments, such as science gateways (SGs) and virtual laboratories (VLs), tend to be domain specific and linked with one or a small number of e-research infrastructures (Jeffery et al., 2017). Research collaborations in the context of VREs include 1) access to data, tools and other resources from different research infrastructures, 2) cooperation or collaboration between researchers at the same or different institutions, 3) cooperation at intra- and inter-institutional levels, and/or 4) preserving or taking care of data and other outputs (Carusi & Reimer, 2010).

VREs can potentially support cooperation or collaboration between researchers and policy makers. VREs do not need to be restricted to particular sectors (e.g. governmental organisations, companies, universities or research institutes) or disciplines (e.g. energy, mobility or climate) and can span across them (European Union, 2016). VREs allow researchers to combine OGD with other types of data (e.g. research data and sensor data), to analyse that data and to obtain novel insights that may provide input for policy makers to solve societal issues.

ISSUES OF USING VRES FOR OPEN DATA ANALYSIS

Various VRE-related initiatives are under development, including EPOS for earth/geo-physical sciences (http://www.epos-eu.org/), ENVRIplus for environmental sciences (http://www.envriplus.eu/) and EXCELERATE for biological/biomedical sciences (http://www.elixir-europe.org/about/eu-

projects/excelerate). However, existing VRE projects are restricted to particular domains (Jeffery et al., 2017). For instance, the ESLab VRE is focused on ecosystem services (Holmberg et al., 2015), the ENS@T-CANCER VRE is focused on adrenal cancer research (Sinnott et al., 2016) and the ENROLLER VRE is focused on the humanities (Sarwar et al., 2013).

The literature shows that VRE-related initiatives suffer from user experience issues (Zuiderwijk et al., 2016). Whereas user satisfaction is critical for benefitting from data sharing and reuse in a VRE (Crosas, 2011), "a lack of support emerges as one of the most critical barriers to the uptake of VREs" (Carusi & Reimer, 2010, p. 35), resulting in a lack of community uptake of VREs (Candela et al., 2013). Critical user support issues concern the lack of ongoing technical support for further development and fixing bugs, and the lack of direct support of researchers engaged with VREs (Carusi & Reimer, 2010), and there are many more issues. These user support issues may present a serious barrier to the collaborative analysis of open data in VREs and may thus hinder the creation of new opportunities for public policy making. Most importantly, there is a lack of insight into how VREs can support collaborative open data analysis by researchers.

REQUIREMENTS FOR OPEN DATA ANALYSIS USING VRES

To be successful, VREs must be easy to use and embedded in the scholars' workflow (Connaway & Dickey, 2010). End user requirements for VREs can be categorised as follows:

1. **Intuitive VRE user interface:** It has been stated that "existing VREs lack efficient and effective user interfaces able to satisfy the needs of the different types of people collaborating in performing certain tasks" (Ardito, Costabile, Desolda, Matera, & Buono, 2016, p. 115). VREs have to cater for a broad range of users, including researchers and policy makers, with different backgrounds, reasoning strategies, daily practices, languages and notations (idem). Sarwar et al. (2013) emphasise minimising the level of detail required by end user researchers in the back-end processes in interacting with services and dealing with issues. This is confirmed by Jeffery, Asserson, Houssos, Brasse and Jörg (2014), who posit that the user interface can offer a simplified virtualisation whereby the user does not need to know how or where the computer processing of the data takes place, provided that service level demands are met. Furthermore, user requirements may evolve over time (Terras et al., 2016), and thus the user interface should be easily adaptable to the needs of the end users;
2. **Easy-to-use data analysis tools:** VREs should enable researchers to conduct a variety of complex research activities (Terras et al., 2016). Data processing tools include tools for searching data (Jeffery et al., 2017; Zuiderwijk, 2015), tools for analysing and visualising data, and tools for assessing data quality (Zuiderwijk, 2015). Moreover, the tools should also enable the composing of workflows and the tracking of data publications (Jeffery et al., 2017);
3. **Clear data use and sharing conditions:** To support the use of open data through VREs, it should be clear to VRE users under which conditions they are legally allowed to use the data (with different levels of sensitivity) and with whom and under which conditions they can share the datasets that they have processed and combined. The literature shows that researchers are concerned with privacy and safety aspects when sharing data (Connaway & Dickey, 2010; Rockhold, Nisen, & Freeman, 2016) and that there may be usage restrictions when data is shared (Chen & Zhao, 2012). For example, datasets may contain sensitive information, such as data related to persons and their diseases (Sinnott et al., 2013). Additionally, various trust issues, such as rapid changes in datasets and a lack of clarity regarding how data can be cited, make data sharing and reuse challenging. There may be concerns about ownership and attribution (Candela et al., 2013). To improve the user experience, the data use environment should be user centred, secure, privacy compliant and sustainable (Jeffery et al., 2017);

4. **Support of interaction and collaboration:** An important characteristic of VREs is that they allow for the generation of new results, including scientific results, by facilitating collaboration between scientists (De Roure & Goble, 2007). This is a type of open collaboration, meaning that the collaboration produces a shared artifact and that it is supported by a technological platform that allows for flexible social structures and enables contributors to enter and exit the collaboration easily (Choi & Tausczik, 2017; Forte & Lampe, 2013). VREs can be used to handle the complex tasks that multidisciplinary collaboration demands (Edwards et al., 2014). Within VREs, virtual research communities can be created (Carusi & Reimer, 2010). VREs enable workflows to be made available to scientists and can provide mechanisms to share such workflows within and across communities and disciplines (De Roure, Goble, & Stevens, 2009). When developing VREs, considerable attention should therefore be paid to community building aspects (SURF Foundation, 2010);
5. **Provision of user training and a helpdesk:** Carusi and Reimer (2010) state that in VREs there should be ongoing technical support for further development and fixing bugs. A user helpdesk may support users by answering questions regarding both technical and content-related issues. Furthermore, data users may be unfamiliar with the opportunities afforded by open data and they may not have enough technical knowledge and/or skills to use the data (Welle Donker & Loenen, 2017). Training might improve this situation. Candela et al. (2013) confirm this lack of training, which they term "instructional support". Scholars should be trained to obtain more knowledge of how the system can be used (Connaway & Dickey, 2010).

RESEARCH APPROACH

A multiple case study design was used to gain insight into how VREs support collaborative open data analysis by researchers. Multiple-case designs are preferred over single-case designs, as the analytic conclusions that arise independently from multiple cases will be more powerful than those coming from a single case (or single experiment) alone (Yin, 2003). As suggested by Yin (2003), a protocol for the investigation of cases was created.

Theoretical sampling concerns choosing cases because they are expected to replicate previous cases, extend emergent theory, fill theoretical categories or supply examples (Eisenhardt, 1989). Theoretical sampling was used to select the cases, as the aim of this study was to contribute to theory building in the field of open data and VREs, rather than to test theories in this field. Open data is a relatively new field in which many studies neither apply existing theories and models nor develop theory and models from scratch, and no single theory or model dominates (Hossain, Dwivedi, & Rana, 2016). There is a need for theory building, which made theoretical sampling an appropriate approach for the case study selection in this research.

The following criteria were formulated to allow for theoretical sampling and to define which characteristics the cases needed to have:

1. **The cases employ e-research infrastructures that are technically ready to interoperate with VREs or be expandable to VREs:** We selected cases in which the e-research infrastructures underlying the VREs were at least ESFRI Landmarks or ESFRI projects of the European Strategy Forum on Research Infrastructures (ESFRI)[1], which means that the underlying e-research infrastructures were technically ready to interoperate with VREs or be expandable to VREs. Selecting ESFRI Landmarks and projects also ensured that all cases had a European Research Infrastructure Consortium (ERIC[2]) and that the e-research infrastructures had reached the implementation phase before the end of 2015, ensuring at least a certain level of maturity;
2. **The cases involve a relation between data analysis and policy making:** The selection of ESFRI Landmarks and projects ensured that this criterion was met, since ESFRI has an important role in policy making (e.g. concerning health issues; see Calzolari et al. (2014) for an example);

3. **The cases represent open data analysis in Europe:** Many VRE-related initiatives are under development all over the world, but we focused on European initiatives, since the European ESFRI Landmarks and projects, which ensure the technical readiness of the cases to interoperate with VREs or be expandable to VREs, also focus on Europe;
4. **Case study information should be available and accessible:** The cases were explored by performing desk research and interviews. The desk research included studying websites, applications and documents, including policy documents. For each case, at least one interview was conducted, and if this did not result in sufficient information, an additional interview was held. Nine interviews were conducted in total. The interviews were conducted by various members of the VRE4EIC project (www.vre4eic.eu), using an interview protocol (obtainable from the author upon request) that ensured consistency between the interviews. The interviewees were experts who were closely involved in the cases and had knowledge of both the technical and the social aspects.

Based on these criteria, seven cases were selected (see Table 1). Five cases are ESFRI Landmarks, one (CLARIAH) is based on two ESFRI Landmarks, and one (EPOS) is an ESFRI project. A protocol for the characterisation of the VRE-related cases was created in the form of an Excel file that contained all the relevant aspects on which the cases were evaluated and compared. In this paper we focus on the requirements for supporting data analysis by researchers as identified in the 'Research Background'

Table 1. Overview of the selected cases

Case	Domain	Objectives
1. ICOS: Integrate Carbon Observation System	Environmental science / earth sciences	"Integrate atmosphere, ecosystem and ocean greenhouse gas observations to provide timely and reliable data for research, policy making, and the general public. ICOS RI brings together high quality European national research communities and measurement stations and, through coordination and support, constitutes a European-wide research infrastructure that serves both scientists and society." (https://www.icos-ri.eu/icos-research-infrastructure)
2. Euro-Argo: A new European research infrastructure contributing to the international Argo programme	Marine / environmental sciences	"Optimise, sustain and improve the European contributions to Argo and to provide a world-class service to the research (ocean and climate) and operational oceanography (Copernicus Marine Service) communities" (as stated by the interviewee, also see http://www.euro-argo.eu/About-us/The-Research-Infrastructure).
3. EPOS: European Plate Observing System	Earth sciences	"Creating a pan-European infrastructure for solid Earth science to support a safe and sustainable society. EPOS will enable innovative multidisciplinary research for a better understanding of the Earth's physical and chemical processes that control earthquakes, volcanic eruptions, ground instability, tsunami, and all those processes driving tectonics and Earth's surface dynamics." (https://www.epos-ip.org/)
4. ELIXIR: A distributed infrastructure for life-science information	Biological sciences, life sciences	"Unite Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research. It coordinates, integrates and sustains bioinformatics resources across its member states and enables users in academia and industry to access services that are vital for their research" (https://www.elixir-europe.org/).
5. LifeWatchGreece: e-Science European infrastructure for Biodiversity and Ecosystem Research	Biological sciences	LifeWatchGreece Research Infrastructure (LWG RI) is the national effort within LifeWatch, which supports relevant studies related to, for example, tourism development, fisheries, agriculture and maritime transport. LifeWatch aims to "advance biodiversity and ecosystem research and to provide major contributions to addressing the big environmental challenges (e.g. climate change), including knowledge-based solutions to environmental managers for its preservation, by providing access through a pan-European distributed e-infrastructure to a multitude of sets of data, services and tools […] ." (see http://www.lifewatch.eu/)
6. CESSDA: Consortium of European Social Science Data Archives	Social sciences	"Provide large-scale, integrated and sustainable data services to the social sciences. It brings together social science data archives across Europe, with the aim of promoting the results of social science research and supporting national and international research and cooperation" (http://cessda.net/).
7. CLARIAH: Common Lab Research Infrastructure for the Arts and Humanities	Humanities and social sciences	Provide researchers with access to large collections of humanities and social science data and with user-friendly applications for processing this data. "CLARIAH designs, implements and exploits the Dutch part of the European CLARIN and DARIAH infrastructures" (https://www.clariah.nl/en/about)

section. The characterisations of the cases were carried out in 2016 and 2017. The characterised user support aspects were presented to interviewees as a template that they were asked to complete and discuss. After the template had been analysed by the author and the findings had been interpreted, the findings were sent to the interviewees, who were asked to check the findings and to supplement them, if necessary. This resulted in Table 2, in the following section.

CROSS-CASE COMPARISON: HOW VRES SUPPORT COLLABORATIVE OPEN DATA ANALYSIS BY RESEARCHERS

Table 2 shows how the selected VRE cases support collaborative open data analysis by researchers. The colours of the cells indicate the types of support given in each case. Green indicates that all questions in that cell were answered positively (pointing at user support for open data analysis through the VRE), red that all questions were answered negatively (pointing at a lack of user support) and orange that some questions were answered positively and some negatively. The interpretation of each case requires one to examine the text in the cells; it is not possible to simply count the number of green cells per case, as the colours indicate in which areas more or less support for open data analysis is provided.

The cross-case comparison shows that VREs provide different degrees of researcher support for collaborative open data analysis. For example, in some cases researchers receive much support in terms of training, but limited support in terms of data analysis tools, whereas this is different in other cases. The limited number of red cells in the table shows that all seven cases provide at least some user support for open data analysis for each of the five areas, namely 1) intuitive VRE user interface (UI), 2) easy-to-use data analysis tools, 3) clear data use and sharing conditions, 4) support of interaction and collaboration, and 5) provision of user training and a helpdesk. The limited number of green cells in the table shows that there is also still much room for improvement regarding user support for open data analysis.

We found that the seven cases differ on two main variables, namely 1) interaction and collaboration and 2) training and helpdesk. By combining these two variables, we identified four support modes (see Figure 2):

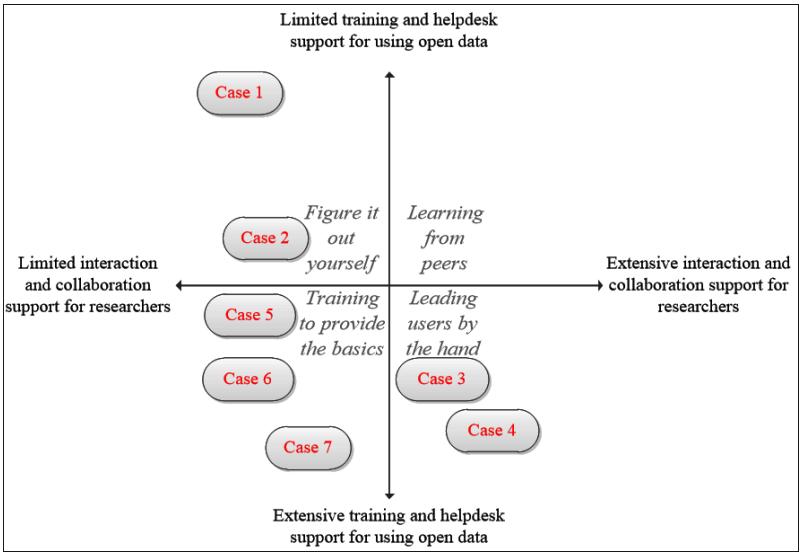
1. 'Figure it out yourself' (limited collaboration, training or other user support) (cases 1 and 2);
2. 'Leading users by the hand' (considerable collaboration, training and other user support) (cases 3 and 4);
3. 'Training to provide the basics' (support in the form of training but only limited collaboration support) (cases 5, 6 and 7);
4. 'Learning from peers' (collaboration support but only limited support in the form of training) (none of the cases belongs to this category).

As far as training and collaboration support are concerned, the first and the second case (ICOS and Euro-Argo) belong to the category 'Figure it out yourself'. In these cases, limited or no training and training-related support in using open data is provided, and there is limited collaboration support for researchers. For instance, these VRE-related initiatives do not offer a group setup, do not allow for group discussions or forums, and do not provide co-editing tools. The first and second cases do not offer an extensive training programme, although the second case does provide user manuals and there are plans to organise training workshops in the future, and it offers a helpdesk. Thus, while collaboration, training and user support are essential if researchers want to collaborate through VREs, these two cases offer limited collaboration and training support to researchers. Researchers who want to use the infrastructures that these cases relate to need a supporting network in case they encounter problems or need to have the knowledge and skills to use this data and solve problems themselves.

Table 2. VRE user support for collaboratively analysing open data in the seven selected cases

	Intuitive VRE user interface (UI)	Easy-to-use data analysis tools	Clear data use and sharing conditions	Support of interaction and collaboration	Provision of user training and a helpdesk
1. ICOS	Language? ***UI in English only*** Information on the usability of the user interface? ***No***	Data analysis tools provided? ***Yes, on statistics and visualisation*** User assisted in recognising the appropriateness of certain data analyses? ***Yes*** User protected against errors (e.g. errors in doing certain analyses)? ***Generally not***	Data license: ***CC BY 4.0***[3] Software license: ***Open source solutions wherever possible***	Group setup? ***No*** Group discussions or forums? ***No*** Co-editing tools? ***No*** Data request tools? ***Yes, users can request data*** System improved based on user feedback? ***Yes*** Information available on how well the system operates (e.g. statistics on quality and availability of data)? ***Yes***	Training provided? ***No*** User helpdesk provided? ***No*** (ICOS is not yet in full operation; more services will be available soon)
2. EURO-ARGO	Language? ***UI in English only*** Information on the usability of the user interface? ***Yes*** A series of SLAs is managed, including the user interface availability	Data analysis tools provided? ***Yes*** User assisted in recognising the appropriateness of certain data analyses? ***Yes, for certain data analyses*** (a quality control for each observation) User protected against errors (e.g. errors in doing certain analyses)? ***Yes, but only if user selects the 'delayed mode data'*** (data that have been checked and adjusted by a dedicated scientist)	Data license: ***CC BY 4.0*** [5] Software license: ***CC BY 4.0*** [5]	Group setup? ***No*** Group discussions or forums? ***Yes, but not used by community*** Co-editing tools? ***No*** Data request tools? ***Yes, users can request data*** System improved based on user feedback? ***Yes*** Information available on how well the system operates (e.g. statistics on quality and availability of data)? ***Yes***	Training provided? ***No, but detailed user manuals are available***. Euro-Argo-ERIC plan to organise training workshops both at technical/hardware level and at data use level User helpdesk provided? ***Yes***
3. EPOS	Language? ***Plans for multilingual UI*** Information on the usability of the user interface? ***Feedback from users available***	Data analysis tools provided? ***Yes, planned*** User assisted in recognising the appropriateness of certain data analyses? ***No*** User protected against errors (e.g. errors in doing certain analyses)? ***Yes, under development***	Data license: ***Generally some form of CC*** Software license: ***Elaborating license agreement for software***	Group setup? ***No*** Group discussions or forums? ***Yes*** Co-editing tools? ***No*** Data request tools? ***Yes, users can request data*** System improved based on user feedback? ***Yes*** Information available on how well the system operates (e.g. statistics on quality and availability of data)? ***May be provided in the future***	Training provided? ***Yes*** User helpdesk provided? ***No***
4. ELIXIR	Language? ***UI in English only*** Information on the usability of the user interface? ***Yes***	Data analysis tools provided? ***Yes*** User assisted in recognising the appropriateness of certain data analyses? ***Yes, through online guidance*** User protected against errors (e.g. errors in doing certain analyses)? ***No***	Data license: ***Under development*** Software license: ***Under development***	Group setup? ***No*** Group discussions or forums? ***Yes*** Co-editing tools? ***Yes*** Data request tools? ***Occasionally allows users to request data*** (for sensitive and personal data) System improved based on user feedback? ***Yes*** Information available on how well the system operates (e.g. statistics on quality and availability of data)? ***Yes***	Training provided? ***Yes, extensive training*** (e.g. Training eSupport System (TeSS) portal, e-learning and face-to-face training) User helpdesk provided? ***No, aimed at expert users***
5. LIFEWATCH - GR	Language? ***UI in English only*** Information on the usability of the user interface? ***No***	Data analysis tools provided? ***Yes*** User assisted in recognising the appropriateness of certain data analyses? ***No*** User protected against errors (e.g. errors in doing certain analyses)? ***In some cases***	Data license: ***Multiple. Providers indicate the data license and embargo*** during data uploading Software license: ***MIT license (Open Source)***	Group setup? ***No*** Group discussions or forums? ***Yes*** Co-editing tools? ***No*** Data request tools? ***No*** System improved based on user feedback? ***Yes*** Information available on how well the system operates (e.g. statistics on quality and availability of data)? ***No***	Training provided? ***Yes*** User helpdesk provided? ***No***
6. CESSDA	Language? ***Multilingual UI under development*** Information on the usability of the user interface? ***No*** (catalogue is still under development)	Data analysis tools provided? ***Yes (e.g. for data publishing and online analysis)*** User assisted in recognising the appropriateness of certain data analyses? ***No*** User protected against errors (e.g. errors in doing certain analyses)? ***No***	Data license: ***Commonly CC-BY 4.0***[3]***, but depends on the data providers*** Software license: ***Commonly CC-BY 4.0***[3]	Group setup? ***No*** Group discussions or forums? ***No*** Co-editing tools? ***Not*** Data request tools? ***Only in certain cases*** (where this is the access mechanism) System improved based on user feedback? ***No*** Information available on how well the system operates (e.g. statistics on quality and availability of data)? ***No***	Training provided? ***Yes,*** see https://cessda.net/CESSDA-Training) User helpdesk provided? ***No***
7. CLARIAH	Language? ***UI available in English, some parts also in Dutch*** Information on the usability of the user interface? ***No*** (UI is new)	Data analysis tools provided? ***Yes, large number of tools and software components*** User assisted in recognising the appropriateness of certain data analyses? ***No*** User protected against errors (e.g. errors in doing certain analyses)? ***No***	Data license: ***CC BY-NC, CC BY-NC-SA***[4] Software license: ***CC BY-NC, CC BY-NC-SA***[4]	Group setup? ***Yes, provided through OpenConext teams (both for internal and external collaboration)*** Group discussions or forums? ***Yes*** Co-editing tools? ***No*** Data request tools? ***No*** System improved based on user feedback? ***No*** Information available on how well the system operates (e.g. statistics on quality and availability of data)? ***No***	Training provided? ***Yes, LOD Lab tutorial at the ESWC conference*** User helpdesk provided? ***Yes***, through e-mail and a chat room

Figure 2. Overview of the four user support models concerning interaction, collaboration, training and helpdesk identified in the seven VRE-related cases



However, the fact that the first and second cases do not provide much training and collaboration support does not mean that they do not support open data use. For instance, in the ICOS case, data processing is supported as all analysis are documented and accompanied by uncertainty estimates. An example from the Euro-Argo case is that the user is assisted in recognising the appropriateness of certain data analyses, as each observation in Euro-Argo has an associated quality control flag, which is assigned by automated quality controls and, when needed, by a visual inspection performed by a specialist in Argo data. This shows that data use is supported in a different way, but not that much via training and collaboration support. Moreover, both ICOS and Euro-Argo offer publicly accessible data free of charge and allow users to request data, and open source solutions are used for ICOS.

A second type of user support model is 'Leading users by the hand'. Compared to the 'Figure it out yourself' model, this is the other side of the coin: much collaboration support for the use of open data through VREs is offered, and extensive training in and training-related support for using open research data are given. In this model, less support from the researcher's own network is needed compared to the other models. Instead, data users can ask other researchers for support through the VRE or they can get support through the offered training programme. The third and the fourth case (EPOS and ELIXIR) belong to the 'Leading users by the hand' model. Both EPOS and ELIXIR allow for group discussions and/or forums and for data requests, and they offer a programme to train end users. For instance, ELIXIR provides a Training eSupport System (TeSS) portal, e-learning and face-to-face training. In the case of both EPOS and ELIXIR, the system is improved based on feedback from users and information is (or will be) available on how well the system operates. In both cases, information on the usability of the user interface is collected. However, also in these two cases user support can be improved. For example, neither case has a group setup or a user helpdesk. EPOS does not offer co-editing tools and does not assist the user in recognising the appropriateness of certain analyses (whereas ELIXIR does), and ELIXIR does not protect the user against errors and does not offer a multilingual user interface (whereas EPOS does).

The third and fourth models are variations of the 'Figure it out yourself' model and the 'Leading users by the hand' model. In the third model, the VRE provides 'Training to provide the basics' (training but limited collaboration). In this model, the VRE enables users to learn how to use data and basic data use skills. However, there is limited collaboration support. The fifth, sixth and seventh cases

(LifeWatchGreece, CESSDA and CLARIAH) belong to this model. LifeWatchGreece and CESSDA do not offer a group setup, provide co-editing tools or offer a user helpdesk. CLARIAH does offer a group setup and a user helpdesk. Moreover, in all three cases the user is not assisted in recognising the appropriateness of data analyses. Regarding the provision of tools to support data analysis, LifeWatchGreece and CLARIAH do not allow users to request data, whereas CESSDA allows this only in certain cases. CESSDA does not enable group discussions or forums (yet LifeWatchGreece does). For all three cases there is no information available on how well the system operates or how usable the user interface is. On the other hand, training is provided. For instance, CESSDA participates in the FOSTER–CESSDA training project, which "supports European researchers in implementing sustainable data management and sharing concepts in their projects by offering training and resources on these topics" (https://cessda.net/CESSDA-Training). Furthermore, in the three cases, open source software is provided. CESSDA data are also licensed for reuse, yet for LifeWatchGreece this depends on the data license that the data providers indicate. CLARIAH data cannot be reused commercially. The CESSDA user interface will be multilingual. Thus, whereas users are trained in such skills as data management and preservation, there are various opportunities to improve user collaboration support.

The fourth user support model for VREs is 'Learning from peers'. In this model, there is collaboration support but limited training for end users. Since none of the cases that we analysed belongs to this fourth category, the question is whether this model actually exists in practice.

RECOMMENDATIONS FOR IMPROVING VRE OPEN DATA ANALYSIS AND SUBSEQUENT POLICY MAKING

Even though there is much potential for using open data to make more informed arguments for embracing, rejecting or proposing new or improved policies (Janssen & Helbig, 2016; Markaki et al., 2014), our case studies showed that this is not as simple as it may seem. Based on the findings presented in the previous sections, we drew up the following recommendations for VRE developers and open data policy makers to support open data analysis in VREs.

Learn From Other VRE Initiatives and Further Develop User Support

First, the cases showed that user support is often limited and should be improved. Table 2 revealed that there are still areas for improvement. This confirms previous findings. For instance, Candela et al. (2013) state that there is a lack of community uptake of VREs, and among the reasons for this are user support factors. According to these scholars, "Virtual Research Environments should be designed, since the beginning, to promote uptake, ensure usability, and guarantee sustainability" (idem, p. GRDI79). We found that users are supported in many ways, which suggests that the stakeholders involved in the VRE design can learn from each other. Some user support aspects may be similar for all VRE initiatives (e.g. how to use a particular tool that can be used for co-editing datasets, or how to use a tool to ask for feedback from other researchers). We recommend that VREs encourage the reuse of user support modules and functionalities that have already been developed, including the following:

- A multilingual user interface;
- Feedback mechanisms (e.g. for users to provide feedback on the usability of the user interface or on how well the system operates);
- Functionalities for automatic user protection against errors in data analysis;
- Functionalities to offer users a group setup, group discussions and forums;
- Functionalities to allow users to co-edit and request data;
- Training modules (e.g. online courses and joint workshops with end users);
- A user helpdesk.

The modules and functionalities that are used in one case may also be useful in another case, so that the usability of the VRE-related initiatives can be improved. In addition, user support should be developed further, since several user support types were not found in the cases. For example, massive open online courses (MOOCs) can be created to train users and improve their data analysis skills.

Develop a VRE for Multidisciplinary Data to Obtain New Insights for Policy Making and Encourage New Collaborations

We found that most of the investigated VRE-related initiatives are focused on a single domain, whereas it is stated in the literature that the added value comes from new combinations of data (Janssen et al., 2017), especially when the data come from different domains (Jeffery et al., 2017). Collaboration between researchers who analyse open data from different domains and policy makers who use the outcomes of this open data analysis, requires a different type of support than a collaboration involving the analysis of open data from a single domain. Although some of the tools supporting the collaboration are generic and may be similar across domains (e.g. Skype and Google Documents), the actual data analysis may be different and some data analysis tools may be domain specific (e.g. certain modelling and simulation tools). This is because each domain has its own standards and semantics to describe datasets. For instance, standards used to describe geographical data are often different from standards used to describe social science data, and terms may have different meanings. Analysing open data from different domains requires support for providing homogeneous access (e.g. data analysis tools that can be used across domains, a single point of access, a similar user interface) hiding the complexity of integrating heterogeneous data sources, standards and domain-specific tools. We recommend the development of a VRE that offers access to data from multiple disciplines combining both domain-specific data analysis tools and generic data analysis tools. This can lead to new insights when, for instance, new data combinations are made or researchers and policy makers from different disciplines start collaborating:

1. VREs should offer clear and effective communication mechanisms for researchers (as open data users) and policy makers (using the outcomes of open data analysis).

Using the outcomes of open data analysis as input for policy making is complex. Public policy making can be considered a wicked problem (Rittel & Webber, 1973). Wicked problems are related to cognitive, strategic and institutional uncertainty and different stakeholders are dependent on each other (Bueren, Klijn, & Koppenjan, 2003). In the context of this study, policy makers are dependent on researchers who analyse open datasets, which means that the communication between these two stakeholders should be well supported. Moreover, for wicked problems, there are no criteria that tell when the or a solution has been found (Rittel & Webber, 1973) and the problem definition of wicked problems changes over time. Furthermore, different types of societal stakeholders are involved in value conflicts about the wicked problem (Batie, 2008; Dentoni & Bitzer, 2013, 2015). Various analyses of open data sets could provide different or even contrasting results and stakeholders may disagree about their value. It may not be clear which outcomes can be trusted and which may be disregarded, and policy makers may not be able to assess this. Based on the foregoing, we formulated the recommendation for VREs to offer clear and effective communication mechanisms for researchers working with open data on the one hand, and policy makers using the outcomes of open data analysis on the other hand. Furthermore, the quality of the data and the data analysis should be very clear, so that the value of the data analysis results can be assessed by policy makers.

2. Develop multilingual VREs.

Many of the societal challenges studied by researchers (e.g. climate change, health in relation to aging) do not respect national borders. Open data analysis to study these challenges often requires

data from multiple countries as well as collaboration between researchers in multiple countries. These data are often in different languages, which makes it difficult to combine the data. A multilingual user interface and multilingual metadata help researchers to study datasets from different countries in different languages. This also enables policy makers to compare data from certain countries (e.g. regarding the level of pollution in the UK) to data on the same topic in other countries (e.g. the level of pollution in the US and China) and to work on joint solutions and programmes to solve international societal problems.

CONCLUSION

This paper contributes to the literature by providing insight into how VREs support researchers in open data analysis. Seven cases presenting different modes of VRE user support for open data analysis in Europe were investigated and compared. Four types of support were identified: 1) 'Figure it out yourself' (limited collaboration, training and other user support), 2) 'Leading users by the hand' (considerable collaboration, training and other user support), 3) 'Training to provide the basics' (training but limited collaboration support) and 4) 'Learning from peers' (collaboration but limited training). While collaboration, training and other user support are essential characteristics of VREs, some cases offer limited collaboration and training support to end users. This is remarkable, since we focused on ESFRI Landmarks and ESFRI projects, which are considered relatively mature cases. Although various initiatives are developing VREs, most of them provide limited user support and can still be improved. Furthermore, we found that most of the investigated VRE initiatives are focused on a single discipline and are not multilingual, whereas societal challenges often require data from multiple disciplines and countries. Recommendations for improving policy making using the results of open data analysis in VREs include:

1. Learn from other VRE initiatives and further develop user support;
2. Develop a VRE for multidisciplinary data to obtain new insights for policy making and encourage new collaborations;
3. VREs should offer clear and effective communication mechanisms for researchers (as open data users) and policy makers (using the outcomes of open data analysis);
4. Develop multilingual VREs.

This paper provides an overview of and recommendations for user support for data analysis by researchers using VREs. The user support modes and the recommendations may be used by technical developers in the design and development of VREs, and they may be considered by policy makers of governmental and non-governmental organisations in the creation of open data policies.

This study has several limitations that need to be taken into account when interpreting the findings. First, a limited number of user support elements were selected, whereas there are other elements that may directly or indirectly influence user uptake of VREs and they have not been assessed. Furthermore, we focused on seven relatively mature VRE initiatives and do not claim that the findings can be generalised to all types of VREs. In addition, the assessment of the VRE initiatives was based on desk research and interviews with one or more persons per case, and therefore reflects the understanding of the author and the selection of the interviewee(s). And finally, the cases evolve over time, so their evaluation and comparison may become outdated relatively quickly. Some interviewees indicated that certain user support elements were under development. Future research should investigate additional VRE-related initiatives and additional elements that influence user support for analysing open data through VREs, such as sustainability and types of tools that are provided to data users. Furthermore, the understanding of VREs should be enhanced by examining the evolution of VREs and the support that they provide to open data users over time.

ACKNOWLEDGMENT

This work was carried out within the VRE4EIC project and received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676247. The author should like to thank her colleagues in this project for their input for this paper (particularly Yi Yin, Zhiming Zhao, Sarah Jones, Theodore Patkos, Daniele Bailo and Jacco van Ossenbruggen for coordinating the interviews). The views expressed in this paper are the views of the author and not necessarily of the VRE4EIC project.

REFERENCES

Ardito, C., Costabile, M. F., Desolda, G., Matera, M., & Buono, P. (2016). A Meta-design Approach to Support Information Access and Manipulation in Virtual Research Environments. In M. X. Bornschlegl, F. C. Engel, R. Bond et al. (Eds.), *Advanced Visual Interfaces. Supporting Big Data Applications: AVI 2016 Workshop, AVI-BDA 2016*, Bari, Italy, June 7–10 (pp. 115-126). Cham: Springer International Publishing. doi:10.1007/978-3-319-50070-6_9

Batie, S. S. (2008). Wicked problems and applied economics. *American Journal of Agricultural Economics*, *90*(5), 1176–1191. doi:10.1111/j.1467-8276.2008.01202.x

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*. doi:10.1145/2909132.2927471

Buddenbohm, S., Enke, H., Hofmann, M., Klar, J., Neuroth, H., & Schwiegelshohn, U. (2015). Success Criteria for the Development and Sustainable Operation of Virtual Research Environments. *D-Lib Magazine*, *21*(9/10). doi:10.1045/september2015-buddenbohm

Bueren, E. M., Klijn, E. H., & Koppenjan, J. F. (2003). Dealing with wicked problems in networks: Analyzing an environmental debate from a network perspective. *Journal of Public Administration: Research and Theory*, *13*(2), 193–212. doi:10.1093/jopart/mug017

Calzolari, A., Valerio, A., Capone, F., Napolitano, M., Villa, M., Pricci, F., & Belardelli, F. et al. (2014). The European research infrastructures of the ESFRI Roadmap in biological and medical sciences: Status and perspectives. *Annali dell'Istituto Superiore di Sanita*, *50*(2), 178–185. PMID:24968918

Candela, L., Castelli, D. & Pagano, P. (2013). Virtual Research Environments: An Overview and a Research Agenda. *Data Science Journal, 12*, GRDI75-GRDI81. doi: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

Chen, D., & Zhao, H. (2012, March 23-25). Data security and privacy protection issues in cloud computing. *Paper presented at the International Conference on Computer Science and Electronics Engineering (ICCSEE)*, Hangzhou, China.

Choi, J., & Tausczik, Y. (2017). Characteristics of Collaboration in the Emerging Practice of Open Data Analysis. *Paper presented at the ACM Conference on Computer Supported Cooperative Work and Social Computing*, Portland, OR. doi:10.1145/2998181.2998265

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

Crosas, M. (2011). The dataverse network: An open-source application for sharing, discovering and preserving data. *D-Lib Magazine*, *17*(1), 2.

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

De Roure, D., Goble, C., & Stevens, R. (2009). The design and realisation of the Virtual Research Environment for social sharing of workflows. *Future Generation Computer Systems*, *25*(5), 561–567. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167739X08000939. doi:10.1016/j.future.2008.06.010

Dentoni, D., & Bitzer, V. (2013). Dealing with wicked problems: managing corporate social responsibility through multi-stakeholder initiatives. Paper presented at the Journal of Management Studies Workshop "Managing for Corporate Social Responsibility", Copenhagen, Denmark.

Dentoni, D., & Bitzer, V. (2015). The role(s) of universities in dealing with global wicked problems through multi-stakeholder initiatives. *Journal of Cleaner Production*, *106*, 68–78. doi:10.1016/j.jclepro.2014.09.050

Edwards, P., Pignotti, E., Mellish, C., Eckhardt, A., Ponnamperuma, K., Bouttaz, T., & Gotts, N. et al. (2014). Lessons learnt from the deployment of a semantic virtual research environment. *Journal of Web Semantics*, *27–28*(0), 70–77. Retrieved from http://www.sciencedirect.com/science/article/pii/S1570826814000560. doi:10.1016/j.websem.2014.07.008

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550.

European Union. (2016). E-infrastructures: making Europe the best place for research and innovation. Retrieved from https://ec.europa.eu/digital-single-market/en/news/e-infrastructures-making-europe-best-place-research-and-innovation

Forte, A., & Lampe, C. (2013). Defining, understanding and supporting open collaboration: Lessons from the literature. *The American Behavioral Scientist*, *57*(5), 535–547. doi:10.1177/0002764212469362

Holmberg, M., Akujärvi, A., Anttila, S., Arvola, L., Bergström, I., Böttcher, K., & Huttunen, M. et al. (2015). ESLab application to a boreal watershed in southern Finland: Preparing for a virtual research environment of ecosystem services. *Landscape Ecology*, *30*(3), 561–577. doi:10.1007/s10980-014-0122-z

Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, *26*(1-2), 14–40. doi:10.1080/10919392.2015.1124007

Janssen, M., & Helbig, N. (2016). Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly*. doi:10.1016/j.giq.2015.11.009

Janssen, M., Konopnicki, D., Snowdon, J. L., & Ojo, A. (2017). Driving public sector innovation using big and open linked data (BOLD). *Information Systems Frontiers*, 19(2), 189–195. doi:10.1007/s10796-017-9746-2

Jeffery, K., Asserson, A., Houssos, N., Brasse, V., & Jörg, B. (2014). From open data to data-intensive science through CERIF. *Paper presented at the 12th International Conference on Current Research Information Systems*, Rome, Italy. doi:10.1016/j.procs.2014.06.032

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.

Kalampokis, E., Hausenblas, M., & Tarabanis, K. (2011). Combining social and government open data for participatory decision-making. *Paper presented at the International Conference on Electronic Participation*, Delft, The Netherlands. doi:10.1007/978-3-642-23333-3_4

Markaki, O., Kokkinakos, P., Koussouris, S., Psarras, J., Lee, H., Löhe, M., & Glikman, Y. (2014). Infusing Innovation in the Policy Analysis and Evaluation Phases of the Policy Cycle: The Policy Compass Approach (pages 19-36). *International Journal of Electronic Government Research*, *10*(3), 19–36. doi:10.4018/ijegr.2014070102

Poel, M., Schroeder, R., Treperman, J., Rubinstein, M., Meyer, E., & Mahieu, B. (2015). Data for Policy: A study of big data and other innovative data-driven approaches for evidence-informed policymaking. Retrieved from https://ofti.org/wp-content/uploads/2015/05/dataforpolicy.pdf

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences*, *4*(2), 155–169. doi:10.1007/BF01405730

Rockhold, F., Nisen, P., & Freeman, A. (2016). Data sharing at a crossroads. *The New England Journal of Medicine*, *375*(12), 1115–1117. doi:10.1056/NEJMp1608086 PMID:27653563

Sá, C. & Grieco, J. (2016). Open Data for Science, Policy, and the Public Good. *review of Policy Research, 33*(5), 526-543.

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. doi:10.1016/j.future.2012.03.015

Scholz, M., & Goerz, G. (2012). WissKI: a virtual research environment for cultural heritage. *Paper presented at the 20th European Conference on Artificial Intelligence*, Montpellier, France.

Sinnott, R., Beuschlein, F., Effendy, J., Eisenhofer, G., Gloeckner, S., & Stell, A. (2016). Beyond a Disease Registry: An Integrated Virtual Environment for Adrenal Cancer Research. *Journal of Grid Computing*, *14*(4), 515–532. doi:10.1007/s10723-016-9375-x

Sinnott, R., Bruns, L., Duran, C., Hu, W., Jayaputera, G., & Stell, A. (2013). Development of an endocrine genomics virtual research environment for Australia: building on success. *Paper presented at the International Conference on Computational Science and Its Applications*, Ho Chi Minh City, Vietnam. doi:10.1007/978-3-642-39640-3_27

Sinnott, R., & Stell, A. J. (2011). Towards a Virtual Research Environment for International Adrenal Cancer Research. *Procedia Computer Science*, *4*, 1109–1118. Retrieved from http://www.sciencedirect.com/science/article/pii/S1877050911001761 doi:10.1016/j.procs.2011.04.118

Sivarajah, U., Weerakkody, V., Waller, P., Lee, H., Irani, Z., Choi, Y., & Glikman, Y. et al. (2016). The role of e-participation and open data in evidence-based policy decision making in local government. *Journal of Organizational Computing and Electronic Commerce*, *26*(1-2), 64–79. doi:10.1080/10919392.2015.1125171

SURF Foundation. (2010). Collaboratories: Connecting Researchers How to facilitate choice, design and uptake of online research collaborations. Retrieved from https://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2010/Collaboratories+Connecting+Researchers9april.pdf

Susha, I., Janssen, M., & Verhulst, S. (2017). Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: taxonomy development. *Paper presented at the 50th Hawaii International Conference on System Sciences*, Hawaii. doi:10.24251/HICSS.2017.325

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* (p. 151). London: Routledge.

Vercamer, D., Steurtewagen, B., Van den Poel, D., & Vermeulen, F. (2016). Predicting consumer load profiles using commercial and open data. *IEEE Transactions on Power Systems*, *31*(5), 3693–3701. doi:10.1109/TPWRS.2015.2493083

Welle Donker, F., & Loenen, B. (2017). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, *10*(3), 284–306. doi:10.1080/17538947.2016.1224938

Yin, R. K. (2003). *Case study research. Design and methods. Thousand Oaks*. SAGE publications.

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

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. doi:10.1109/CeDEM.2016.20

ENDNOTES

1 See http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

2 See https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric

3 CC-BY 4.0 means that the data/software user is free to share and adapt the data/software, but must give appropriate credit, provide a link to the license and indicate whether changes were made, see https://creativecommons.org/licenses/by/4.0/.

4 CC-BY-NC means that the data/software user is free to share and adapt the data/software, but must give appropriate credit, provide a link to the license and indicate whether changes were made. It is not allowed to use the data/software for commercial purposes. See https://creativecommons.org/licenses/by-nc/2.0/

5 CC-BY-NC-SA means that the data/software user is free to share and adapt the data/software, but must give appropriate credit, provide a link to the license and indicate whether changes were made. It is not allowed to use the data/software for commercial purposes. If the user remixes, transforms or builds on the data/software, the user must distribute his/her contributions under the same license as the original. See https://creativecommons.org/licenses/by-nc-sa/3.0/

Anneke Zuiderwijk is a researcher at Delft University of Technology. She holds a PhD (with honor) in open data infrastructures and was ranked as one of the most prolific open data researchers (Hossain et. al., 2015). In 2016, Anneke won the Digital Governance Junior Scholar Award of the ASPA and DGS. See http://www.tbm.tudelft.nl/AZuiderwijkvanEijk