






# Virtual Research Environments for Environmental and Earth Sciences: Approaches and Experiences

Keith Jeffery<sup>1</sup>(✉) , Leonardo Candela<sup>2</sup> , and Helen Graves<sup>3</sup> 

<sup>1</sup> Keith G Jeffery Consultants, Faringdon, UK

keith.jeffery@keithgjefferyconsultants.co.uk

<sup>2</sup> National Research Council of Italy,

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Pisa, Italy

leonardo.candela@isti.cnr.it

<sup>3</sup> British Geological Survey, Nottingham, UK

hmg@bgs.ac.uk

**Abstract.** Virtual Research Environments (VREs) are playing an increasingly important role in data centric sciences. Also, the concept is known as Science Gateways in North America where generally the functionality is portal plus workflow deployment and Virtual Laboratories in Australia where the end-user can compose a complete system from the user interface to use of e-Infrastructures by a 'pick and mix' process from the offered assets. The key aspect is to provide an environment wherein the end-user - researcher, policymaker, commercial enterprise or citizen scientist - has available with an integrating interface all the assets needed to achieve their objectives. These aspects are explored through different approaches related to ENVRI.

**Keywords:** VRE · Science gateway · Virtual laboratory · Workflow management

## 1 Introduction

Research has increasingly become specialised into communities such as oceanography, ecology, geology, materials science. However, many phenomena can only be understood by bringing together the research activities of several communities. Examples include the relationship between shellfish pollution, algal blooms and agricultural use of nitrates or the relationship between ill-health, climate and social conditions. Over the last few years, many communities have developed pan-European research infrastructures (RIs) bringing together several national research teams and assets such as datasets, software, publications, expert staff, sensors and equipment. One way to assist and encourage interdisciplinary research is to bring together the communities and assets of the RIs.

However, this poses a problem. Each community has developed its own standards and practices for research methods, data formats, software to be used, etc. This makes

it difficult for e.g. an ecologist to utilise oceanographic data. The heterogeneity is represented especially in digital representations of data, software, persons, organisations, workflows and equipment. However, many of these assets are represented digitally by metadata providing a succinct description of the asset. The metadata standard chosen varies from community to community. On the other hand, there is a limited set of basic things (entities or objects) that are involved in research (like data, persons and samples) and so the various metadata standards have some commonality in the things they represent – although they do so in different ways.

Thus, the ‘line of attack’ to provide multidisciplinary challenges for researchers is to try to harmonise the metadata and thus gain access to – and (re-)utilisation of – the assets. There are two basic approaches: the software broker approach provides mapping and conversion between pairs of metadata standards. This results in  $n(n-1)$  converter pairs. The alternative approach is to choose a canonical superset metadata standard and convert each metadata standard to/from that. This results in  $n$  converter pairs. This metadata-driven brokering is now regarded as the best approach [1, 8]. However, again we have two choices; the canonical superset may be realised physically – so providing an ‘umbrella’ consistent metadata resource or catalogue over all the participating RIs or the superset metadata may just be a reference syntax (structure) and semantics (meaning) and each RI provides its pair of converters. The latter approach leads to architecture with peer RI to RI communication, requiring quite some software at each RI to interact with the other RIs and generate appropriate workflows. The former leads to a system over the RIs – linked to them via APIs (Application Programming Interfaces) - has the advantage of a ‘helicopter view’ over the participating RIs and so can generate workflows optimally. Either way, the core of a VRE is the superset catalogue (whether conceptual or physical) [12].

In fact, a VRE provides more than access to the assets of RIs; it also provides researcher intercommunication through various means and software to generate workflows to harness the available analytics, visualisation and simulation capabilities of the RIs. Ideally the VRE workflow should be optimised to ensure co-location of data and software which means moving data to the software from the various RIs participating or – especially as datasets become larger – moving the software to the data. This has implications in terms of access rights, privacy and security and in finding an equitable method of ‘payment’ for use of the RI assets. The VRE may also use e-Is (e-Infrastructures) such as external curated storage or supercomputing services with the requirement to manage the deployment of (parts of) the workflow to these e-Is. The VRE should assist the researcher with research management; assisting in finding relevant research, assisting in research proposals, tracking research portfolio and cataloguing research outputs (such as scholarly publications, patents, datasets, software) since increasingly funding organisations utilise such information in planning future research programmes and in evaluating the quality of research proposals.

Recognition of the importance and utility of VREs is increasing. Similar concepts exist in North America (Science Gateways) [2] and in Australia (Virtual Laboratories). The RDA (Research Data Alliance) VRE Interest Group<sup>1</sup> was initiated by the leaders of

<sup>1</sup> <https://www.rd-alliance.org/groups/vre-ig.html>.

VRE4EIC and EVER-EST (this was very much a European initiative) but now includes key experts from Science Gateways and Virtual Laboratories.

This chapter discusses three initiatives dealing with the development of Virtual Research Environments: (i) the D4Science experience, an infrastructure enacting the development of several instances of Virtual Research Environments serving the needs of various communities of practice; (ii) the EVER-EST project, an EU project supporting the development of one Virtual Research Environment for the Earth science community; and (iii) the VRE4EIC project, an EU project proposing a reference architecture for Virtual Research Environments where metadata-based interoperability plays a key role.

## 2 The D4Science Approach and Experiences

D4Science is a hybrid infrastructure specifically conceived to support the development and operation of Virtual Research Environments by the as-a-Service provisioning mode.

The D4Science VRE Manager is a service enacting the definition, deployment and operation of *Virtual Research Environments on demand* (on D4Science infrastructure premises [3]. D4Science-based Virtual Research Environments (VREs) are **web-based, community-oriented, collaborative, user-friendly, open-science-enabler** working environments for scientists and practitioners willing to work together to undertake a certain (research) task [4, 5]. From the end-user perspective, each VRE manifests in a web application (a) comprising several components and (b) running in a plain web browser. Every component is aiming at providing VRE users with facilities implemented by relying on one or more services provisioned by diverse providers. In fact, every VRE is conceived to play the role of a *gateway* giving seamless access to the *datasets* and *services* of interest for the designated community and their tasks while hiding the diversities originating from the multiplicity of resource providers.

The following *key features* characterise the service:

- **Wizard-based VRE characterisation:** the service offers a wizard-based mechanism enabling authorised users (aka VRE designers) to easily select the features (e.g. datasets, facilities and policies) characterising the needed VRE;
- **Dynamic context management:** the service automatically creates the security context needed by the service instances contributing to the VRE to work in a secure and organised manner;
- **Open and extensible resource model:** the service relies on a resource model to know what are the available features to be proposed at *VRE definition time* (i.e. what are the features and the capabilities supportable by the currently available services) and how these features have to be deployed at *VRE creation time* (i.e. what are the services to be configured and how they should be instructed to support the requested features);
- **Per VRE customisable UI:** the service offer facilities enacting authorised users to customise the UI of the VRE, e.g. to define the pages it should be structured in, to allocate the VRE UI components per page and to add web content. Moreover, it provides VRE users with the web app needed to use the working environment and its facilities;

- **Ready to use basic services:** the service equips every VRE with key services enacting the VRE members to cooperate by common facilities, i.e. (a) a *shared workspace* to store and organise items of interest; (b) a *social networking area* to, e.g. post messages, have discussions, express opinions; (c) a *catalogue* to publish artefacts resulting from the VRE activity; (d) a *user management area* to deal with VRE membership (e.g. invite new members), create groups, assign roles.

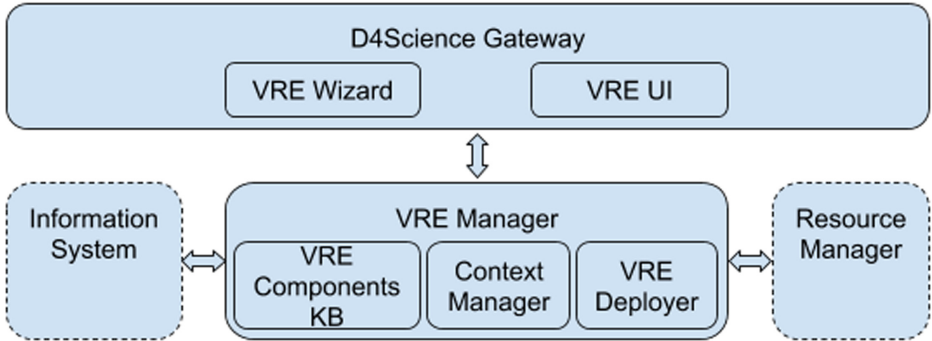


Fig. 1. The overall architecture of the D4Science VRE manager service.

## Design

The D4Science VRE Manager service architecture is depicted (Fig. 1)

The main components are:

- the **VRE Manager service**, i.e. the component implementing the entire business logic related with VRE management. It comprises three subcomponents: (i) the **VRE Components KB**, i.e. the component called to build the knowledge base consisting of potential features (and accompanying services) that can be instantiated at VRE definition time. These features will be built by exploiting the information stored into the *Information System*, namely services and their capabilities, datasets, software components, hosting nodes; (ii) the **VRE Deployer**, i.e. the component called to transform the VRE specification produced by the *VRE Wizard* into a concrete deployment plan consisting of services (and their accompanying configurations) to be deployed to satisfy the specified features. This service is also responsible to implement the deployment plan by either instructing/configuring existing service instances or creating new ones to serve the VRE application; (iii) the **Context Manager**, i.e. the component called to interact with the *Resource Manager* to create the application context needed by the target services to work together and behave as expected;
- the **D4Science Gateway**, i.e. the front-end of the service. It hosts two sets of portlet: (i) the **VRE Wizard**, i.e. the portlet supporting authorised users to specify the features a new VRE should have by selecting them from an ever updated list of possible one resulting from the D4Science offering captured by the *Information System*; and (ii) the

**VRE UI**, i.e. the set of portlets forming the specific VRE working environment. These portlets include those providing access to the basic facilities (e.g. user management and shared workspace) as well as those providing access to specific services deployed in the VRE;

- additional services enacting the VRE Manager to implement the VRE in the overall D4Science infrastructure settings. These services include: (i) the **Information System** providing the VRE Manager with a comprehensive and ever updated list of services and resources currently forming the overall D4Science infrastructure and its operational state; (ii) the **Resource Manager** enacting the VRE Manager to configure existing instances of services or create new ones needed for the VRE operation and to monitor their availability and behaviour.

### Implementation

The VRE Manager service, the Information System and the Resource Managers are all based on the homologous software components of the gCube software system, namely are Java-based Web Services contributing to the gCube system.

The D4Science Gateway is mainly based on the Liferay portal technology. A rich set of portlets (UI components) have been developed to act as access points to the underlying services as well as portal has been equipped with additional software components integrating it with the rest of D4Science services, e.g. components dealing with AuthN and AuthZ, and components interfacing with the Information System.

### Deployment

The components presented above are designed to be allocable on many nodes and to exist in multiple instances.

In particular, the *VRE Manager* service can be deployed on a machine other than that hosting the D4Science Gateway. Moreover, many VRE Managers can be deployed in the infrastructure each serving a specific *virtual organisation*. This deployment option is key for multi-tenancy scenarios where diverse communities are provided with their own features set at VRE definition phase.

The *D4Science Gateway* is conceived to be deployed on a cluster with an instance per node plus a proxy acting as a unifying access point. Every instance can be configured to give access to a number of VREs (e.g. a community gateway contains all the VREs created for the needs of such a community) and to host the VRE Wizard enacting the creation of new VREs. Every VRE consists of a number of portlets organised according to the VRE specification.

The *Information System* is a conceptually centralised service yet its architecture is highly distributed and scalable thus to be able to serve many communities and cases. The resources are registered per virtual organisation and per virtual research environment (thus implementing the “application context” created by the VRE Manager).

The *Resource Manager* is a conceptually centralised service having actuators on every node hosting a D4Science service. A hierarchy of interoperating instances can be built thus having instances taking care of coordinating the management of services at the level of virtual organisation with instances taking care of resources management at the level of every VRE.

### Use Cases

The D4Science VRE Manager service has been used to deploy and operate hundreds of VREs on D4Science premises. These VREs have been deployed to serve very diverse scenarios stemming from application contexts ranging from agri-food (AGINFRA+) to social sciences and humanities (PARTHENOS), environmental science (ENVRiplus), fisheries and conservation, aquafarming (iMarine and BlueBRIDGE), social mining (SoBigData.eu). A comprehensive list of currently supported VREs is available online<sup>2</sup>.

## 3 The EVER-EST Approach and Experiences

### 3.1 The Challenge

Vast amounts of data about our planet are now available to researchers and it is important that this data is easily discoverable, accessible and properly exploited, preserved and shared in order to provide information for a whole spectrum of stakeholders: from scientists and researchers to decision and policy makers at the highest level.

Virtual Research Environments (VREs) provide the IT infrastructure to enable researchers to collaborate, share, analyse and visualise data over the internet. The development of a number of e-infrastructures within Europe and other areas to support activities such as Data Discovery and access has provided the foundations for the development of VREs.

The EVER-EST project (European Virtual Environment for Research – Earth science Themes: a solution) aimed to create a virtual research environment (VRE) focused on the requirements of the Earth science community. Within the earth sciences there are major challenges such as climate change research and ensuring the secure and sustainable availability of natural resources and understanding natural hazards which require interdisciplinary working and sharing of large amounts of data across diverse geographic locations and science disciplines to work towards a solution.

The project includes a major work stream to develop a virtual research environment, and this builds on a number of e-infrastructures which have been created under European Commission funding in recent years. Other work packages test this emerging infrastructure using appropriate use cases.

### 3.2 Creating a Virtual Research Environment

Scientific research in the Earth Sciences is conducted on many different scales from the local to the global. Much of this research is becoming increasingly multidisciplinary and being conducted by researchers who are not necessarily co-located. To support this increasingly distributed approach to Earth science research there is a demand for virtualised collaborative working environments where researchers can share resources e.g. data, workflows, ideas, knowledge and results.

Key objectives of the EVER-EST project are:

<sup>2</sup> <https://services.d4science.org/explore>.

- Creation of a virtual research environment (VRE) that provides a platform and suite of generic services to support collaborative research in the Earth Sciences;
- Validation of the EVER-EST VRE by the four pre-selected Virtual Research Communities (VRCs) that bring unique use cases in terms of their data, workflows, working practices, and desired outcomes;
- Validation of the novel use of the Research Objects concept for application in the Earth science domain. The concept of Research Objects has previously been validated by other disciplines such as astrophysics;

Engaging with the wider Earth Sciences community to promote adoption of the EVER-EST VRE as a solution for dynamic and potentially cross-disciplinary collaborative research.

### **3.3 Validate the Virtual Research Environment with Four Main Virtual Research Communities**

The VRE was validated and evaluated through these four real-world use cases which are provided by existing communities of practice from the Earth science domain. The EVEREST consortium includes a key representative for each of the four Virtual Research Communities who is responsible for the tailoring and validation phase of the EVER-EST VRE for the specific use case and must also ensure the involvement and engagement of additional members of the community outside the EVER-EST project. The VRCs are:

- Sea Monitoring VRC – led by CNR-ISMAR
- Natural Hazards VRC (floods, geological, weather, wildfires) –led by NERC
- Land Monitoring VRC – led by SatCen
- Supersites VRC (volcanoes and seismic) – led by INGV

### **3.4 Implement and Validate the Use of “Research Objects” in Earth Science**

The EVER-EST project defined, implemented and validated the use of “Research Objects” concepts and technologies in the Earth science domain as a mean to establish more effective collaboration.

Modern scientists are calling for mechanisms that go beyond the publication of datasets. They increasingly need to systematically capture the life cycle of scientific investigations and provide a single-entry point to access the information about the hypothesis investigated, the datasets used, the computations and experiments carried out, their outcomes, the people involved in the research, etc. Research Objects (RO) provide a structured container to encapsulate research data and the associated methodologies along with essential metadata descriptions.

### **3.5 Definition of EVER-EST Building Blocks**

During the initial phase of the EVER-EST project, the technical activities focused on the assessment and definition of the main interface between the EVER-EST building blocks and the integration activities of the core infrastructure.

A study on the novel use of Research Objects in the Earth Sciences was carried out in consultation with the virtual research communities (VRCs). This was combined with an in-depth discussion to identify the requirements for the individual use cases provided by the VRCs including a definition of the data that needs to be integrated into the EVER-EST infrastructure (Fig. 2).

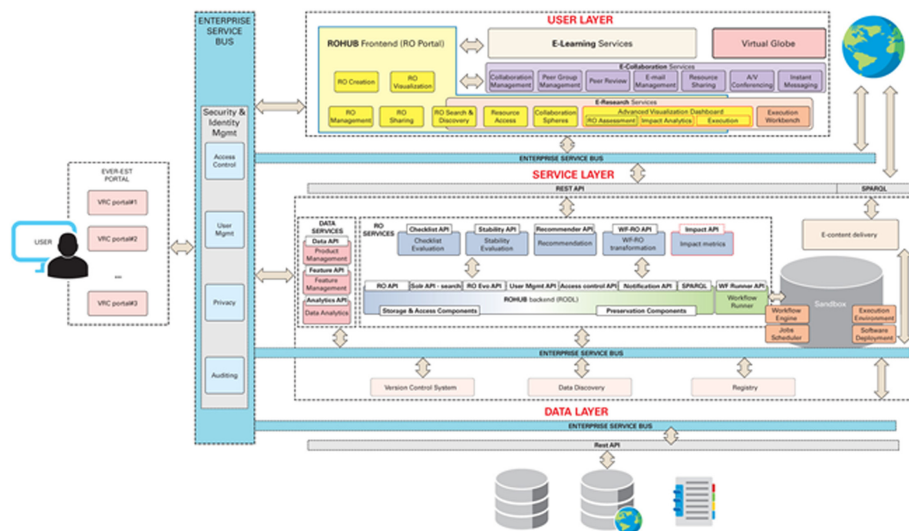


Fig. 2. The overall architecture of the EVER-EST VRE.

## 4 The VRE4EIC Approach and Experiences

### 4.1 Introduction

VRE4EIC aims at providing a model for Virtual Research Environments, which includes requirements, reference architecture and implementation on two use cases to demonstrate its feasibility and innovative impact. VRE4EIC has chosen CERIF<sup>3</sup> (Common European Research Information Format: an EU recommendation to Member States) to denote the superset catalogue.

VRE4EIC has undertaken a considerable amount of requirements collection and analysis, and has characterised many RIs to understand their available interfaces. The architecture has been designed and constructed. The prototype has been evaluated by the RIs that are in the project (ENVRI and EPOS) first, and then other RIs will be invited to evaluate the system.

In parallel, VRE4IC has been cooperating with other VRE projects, notably EVER-EST in Europe but also – via the VRE Interest Group of RDA (Research Data Alliance)<sup>4</sup> -

<sup>3</sup> <http://www.eurocris.org/cerif/main-features-cerif>.

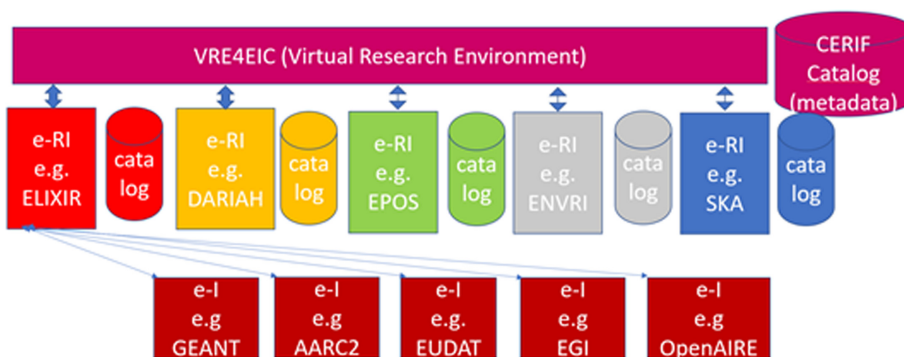
<sup>4</sup> <https://rd-alliance.org/groups/vre-ig.html>.



SGs (Science Gateways) in North America<sup>5</sup> and VLs (Virtual Laboratories)<sup>6</sup> in Australia. In parallel, the various metadata groups in RDA, coordinated by Metadata Interest Group (MIG), are working on a standard set of metadata elements - to be used to describe RI assets in catalogues - which are not simple attributes with values but will have internal syntax and semantics [6].

## 4.2 VRE4EIC in Context

A VRE has to effectively deal with the external resources of data, software services, and infrastructures of computing, storage and network. Figure 3 illustrates how we envision the position of a VRE in the new landscape where e-Infrastructures and Research Infrastructures operate. In particular, e-Infrastructures are seen as providing the basic computational and network resources (like EGI<sup>7</sup>, GEANT<sup>8</sup> and EUDAT<sup>9</sup>) and some fundamental services, such as federated access and authentication and authorisation mechanisms (AARC2<sup>10</sup>) or open access to research publications and data.



**Fig. 3.** Positioning of a Virtual Research Environment in relation to e-Research Infrastructures (e-RIs) and e-Infrastructures (e-Is).

Research-Infrastructures, on the other hand, employ the services and resources of e-Infrastructures to provide resources for their research communities. Each RI is devoted to a specific discipline, or cluster of related disciplines (e.g. DARIAH<sup>11</sup> is for the Humanities, EPOS for the Earth sciences). A VRE, in turn, sits on top of RIs to enable scientific communities to access data, services and tools from and, above all, across RIs. The CERIF-based Catalogue is central to achieve the VRE functionality, as it copes with the interoperability issues implied in that functionality, as described in the rest of the paper.

<sup>5</sup> <https://sciencegateways.org/>.

<sup>6</sup> <https://nectar.org.au/labs/>.

<sup>7</sup> <https://www.egi.eu/>.

<sup>8</sup> <https://www.geant.org/>.

<sup>9</sup> <https://eudat.eu/>.

<sup>10</sup> <https://aarc-project.eu/>.

<sup>11</sup> <https://www.dariah.eu/>.

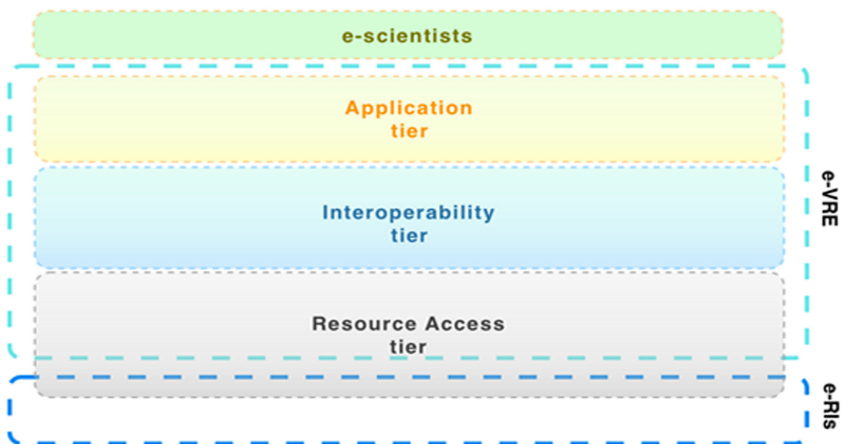
There is an alternative architecture where the e-VRE components are built into each e-RI. However, this means that each e-RI has to maintain in its catalogue the catalogue content of all other e-RIs for interoperability with the usual problems of currency and integrity, especially if the native catalogue of an e-RI uses an insufficiently rich metadata format.

### 4.3 The VRE4EIC e-VRE Reference Architecture

At the general level, the Reference Architecture conforms to the multi-tiers view paradigm used in the design of distributed information systems [7]. Following this paradigm, we can individuate three logical tiers in the e-VRE:

- The *Application* tier, which provides functionalities to manage the system, to operate on it, and to *expand* it, by enabling administrators to plug new tools and services into the e-VRE.
- The *Interoperability* tier, which deals with interoperability aspects by providing functionalities for: i) enabling application components to discover, access and use e-VRE resources independently from their location, data model and interaction protocol; ii) publishing e-VRE functionalities via a Web Service API; and iii) enabling e-VRE applications to interact with each other.
- The *Resource Access* tier, which implements functionalities that enable e-VRE components to interact with e-RIs resources. It provides synchronous and asynchronous communication facilities.

Figure 4 depicts the logical tiers of e-VRE and shows their placement in an ideal space between the e-researchers that use the e-VRE and the e-RIs that provide the basic resources to the e-VRE. Based on the analysis of the requirements, a set of basic functionalities have been individuated and grouped into six conceptual components:



**Fig. 4.** Architectural tiers in a Virtual Research Environment.

- The e-VRE management is implemented in the **System Manager** component. The System Manager can be viewed as the component enabling Users to use the *core* functionalities of the e-VRE: access, create and manage resource descriptions, query the e-VRE information space, configure the e-VRE, plug and deploy new tools in the e-VRE and more.
- The **Workflow Manager** enables users to create, execute and store business processes and scientific workflows.
- The **Linked Data (LD) Manager** is the component that uses the LOD (Linked Open Data) paradigm, based on the RDF (Resource Description Framework) data model, to publish the e-VRE information space - i.e. the metadata concerning the e-VRE and the e-RIs in a form suitable for end-user browsing in a SM (Semantic Web)-enabled ecosystem.
- The **Metadata Manager** (MM) is the component responsible for storing and managing resource catalogues, user profiles, provenance information, preservation metadata used by all the components using extended entity-relational conceptual and object-relational logical representation for efficiency.
- The **Interoperability Manager** provides functionalities to implement interactions with e-RIs resources in a transparent way. It can be viewed as the interface of e-VRE towards e-RIs. It implements services and algorithms to enable e-VRE to: communicate synchronously or asynchronously with e-RIs resources, query the e-RIs catalogues and storages, map the data models. The Interoperability Manager is also responsible for efficiently managing the integration of third-party software, enabling the RA to virtually acquire any desired functionality that is not directly offered by any component of the RA. A case in point is the functionality required to assist researchers in communication with peers and in the administrative processes that are implied by research management. In general, this is the strategy chosen by the project to cope with all those aspects that are under standardization and, as such, do not tolerate formalisation at this stage.
- The **Authentication, Authorisation, Accounting Infrastructure (AAAI)** component is responsible for managing the security issues of the e-VRE system. It provides user authentication for the VRE and connected e-RIs, authorisation and accounting services, and data encryption layers for components that are accessible over potentially insecure networks. The AAAI component interfaces with external identity providers to enable single sign-on across the various connected infrastructures. For any authenticated user, it provides authorization services by using attributes provided by the external identity provider (if any). Furthermore, the AAAI component manages security, privacy and trust aspects of the e-VRE and its connections to the e-RIs. This includes user authorisations (role-based access) and accounting and billing of resources for which payment is required, both based on (CERIF) metadata provided by the metadata manager component.

Figure 5 shows how these six components are distributed on the 3-tier space introduced above. The detailed specification of the interfaces of the components of the Reference Architecture.

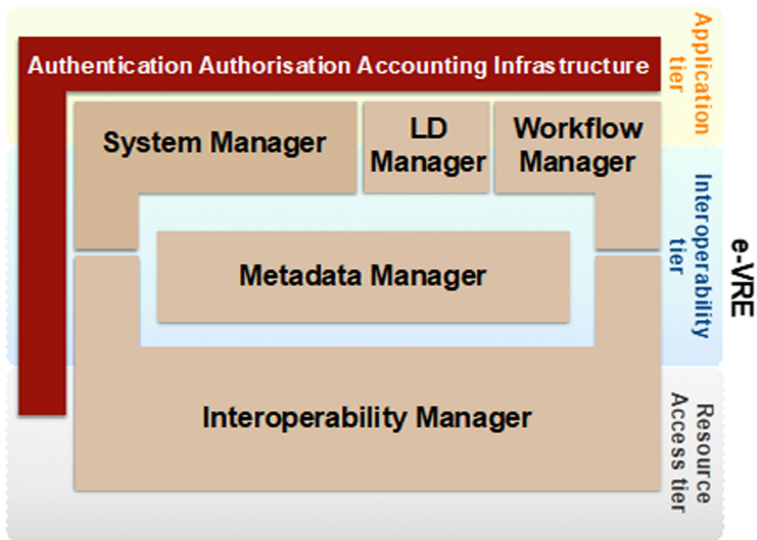


Fig. 5. Conceptual components and logical tiers.

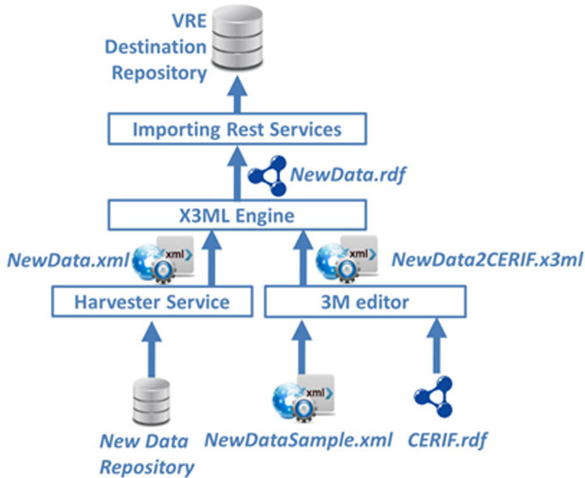
### Catalogue and Mapping

The metadata catalogue describes, provides access to and records actions on the assets of the RIs addressed by e-VRE. Mapping is required to represent the inhomogeneities [8, 12] of each RI in a homogeneous way to permit interoperability using the catalogue and thus is core to the reference model. The 3 M web application<sup>12</sup> is an open source application suite which supports schema mapping, Unique Resource Identifier (URI) definition and generation, (meta)data transformation, provision and aggregation. 3 M is based on the X3 ML mapping definition language for describing the schema mappings. 3 M is used to define mappings between various metadata formats used in existing VREs/RIs and the e-VRE. 3 M allows data experts to transform their internal structured data and other associated contextual knowledge to other formats. Fields or elements from a source database are aligned with one or more entities described in the target format. The purpose of this is typically for integration with other (meta)data also transformed to the same target format.

The process of mapping (meta)data using the 3 M tool is shown (Fig. 6). The first step is to define the mapping between two formats using the 3 M tool. This step needs at least two resources: the source schema (or an XML sample) and the target schema, in this case CERIF expressed in RDF. This step produces an X3 ML document describing the mapping that has been realised in 3M.

This result is used by the X3ML engine to apply the transformation defined in the mapping to a set of data. This data is harvested from a source repository through a harvester to get a set of data that has exactly the same format as the source schema (or XML sample). The X3ML engine is then able to transform the data to the target schema using the rules defined in the X3ML file resulting from step 1. The result of this second

<sup>12</sup> [https://www.ics.forth.gr/isl/index\\_main.php?l=e&c=721](https://www.ics.forth.gr/isl/index_main.php?l=e&c=721).



**Fig. 6.** Mapping of metadata to a common format.

step is an RDF file containing the data harvested in the target schema. This last result can finally be imported in the destination repository using a REST service.

### Development

During the VRE4EIC project, we have conducted a Gap Analysis to identify the most needed components in existing e-RIs and VREs, with a special attention to the EPOS and ENVRIplus Research Infrastructures [11]. The analysis highlighted the heterogeneity of approaches and technologies adopted by current VRE and e-RI systems, especially in relation to the management of resource catalogues; additionally, the security infrastructure technologies adopted by most of VREs/e-RIs have limitations when executing operations on a distributed workspace. The components that have been selected by the Gap Analysis are the Metadata Manager, the AAI Manager and the Node Manager. In order to implement these components and fit them into the EPOS and the ENVRIplus architectures, the VRE4EIC Consortium has made a plan that is illustrated next [10].

Figure 7 shows an overview of the Reference Architecture, including the sub-components in which every component has been structured for modularity reasons. For instance, the Workflow Manager has three sub-components: the WF Configurator that implements workflows definition functionalities, the WF executor implementing execution functionalities and the WF repository component implementing storage management for workflows. For the same modularity reasons, the Query Manager has been elevated to the role of independent component. Thus, overall the Reference Architecture includes seven main components, each corresponding to a functional area:

- Identity, access and logging Management (AAAI Manager)
- System management (System Manager)
- Catalogues management including interoperability conversions (Metadata Manager)
- Metadata publishing using Linked Data Paradigm (LD Manager)
- Workflows management (Workflow Manager)

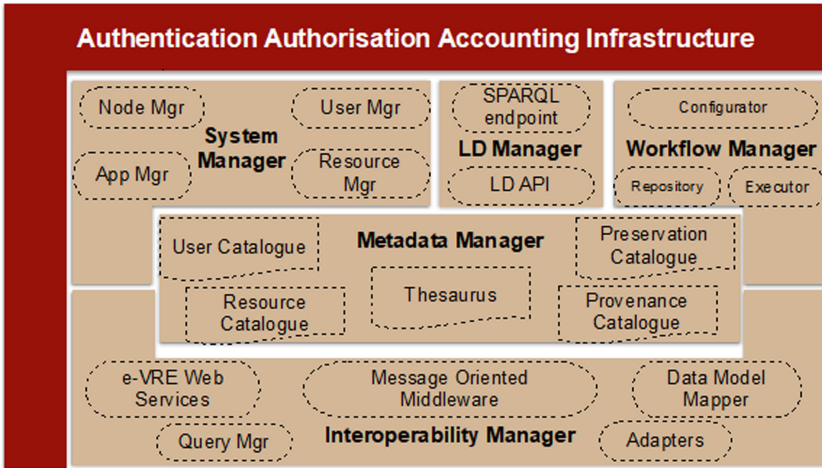


Fig. 7. The developed reference architecture.

- External application integration management (Interoperability Manager)
- Distributed query management (Query Manager).

Virtual Research Environments are dynamic systems; when new tools or technologies emerge a VRE should be able to integrate them. This means that the e-VRE architecture should be easily expandable by adding new software modules or replacing existing software components. Additionally, a component should be replaced or evolved (for instance using new software libraries) without affecting other components. The e-VRE should be potentially used in every research domain; for every domain it should be able to adopt the right technology to implement its functionalities. Deep integration (i.e. integration via Adapters) should be exposed as services in a standardised way to enable users to build clients not depending from the particular integration technology. The e-VRE system must be scalable to meet dynamic changes in the load of research computing processes at component level and independently deployable since they can be reused in other VREs.

In order to meet these requirements, an approach based on Microservices<sup>13</sup> has been chosen. As a result, the building blocks of the TA are autonomous services cooperating with each other to implement the above functional areas. The interaction between the TA services is mainly implemented using an asynchronous paradigm, based on the concept of event. The result is an event driven architecture [9]. Figure 8 shows the resulting micro-services, highlighting the components included, each characterised by the colour relative to the tier where the component belongs.

A repository has been created on GitHub to host the codebases of e-VRE services (VRE4EIC project, 2018). The e-VRE Services will be developed independently and the integration will be done using the APIs published by the Node service. A server has

<sup>13</sup> <https://martinfowler.com/articles/microservices.html>.

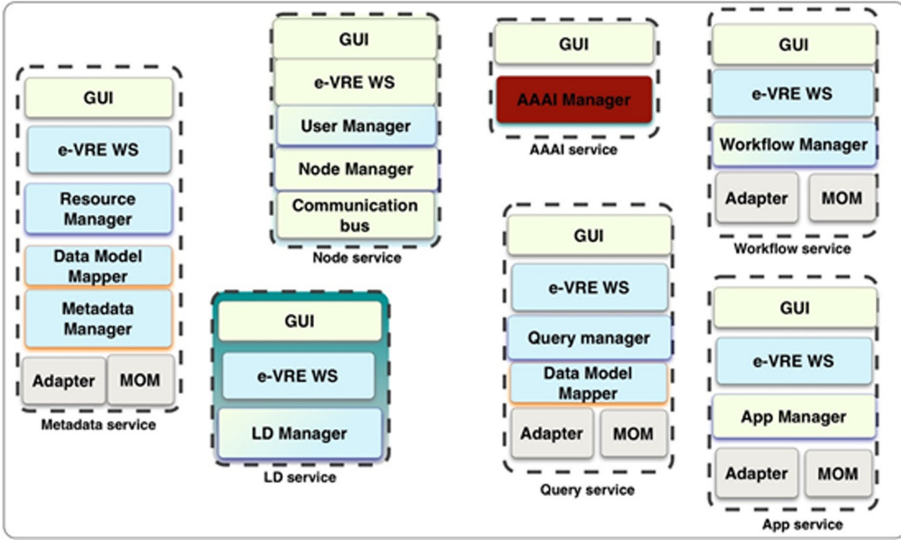


Fig. 8. The services in the technical architecture of e-VRE.

been created on the CNR ISTI cloud (v4e-lab.isti.cnr.it), hosting a continuous integration framework and a number of services used in the development of integration tests.

A repository has been created on GitHub to host the codebases of e-VRE services (VRE4EIC project, 2018). The e-VRE Services will be developed independently and the integration will be done using the APIs published by the Node service. A server has been created on the CNR ISTI cloud (v4e-lab.isti.cnr.it), hosting a continuous integration framework and a number of services used in the development of integration tests.

**Use of e-VRE**

Novel elements of the proposed reference architecture for an enhanced Virtual Research Environment include the metadata mapping, the microservice architecture and the co-development (i.e. evaluation on the architecture via workshops, and keeping developers and end-users in the feedback loop). In this section we will briefly demonstrate these novel elements by presenting two scenarios.

The first scenario demonstrates the integration between an external application and the e-VRE system (see Fig. 9). The proposed use case is to use the e-VRE Taverna plugin to enable users to create workflows, using resources from, for example, the European Plate Observing System.

The three boxes represent the EPOS, the e-VRE and the “user system” (e.g. laptop). Initially an EPOS user launches the TAVERNA workbench application in order to execute some scientific workflow (step 1 in the figure) on his/her own laptop. In order to access to workflows provided by the e-VRE system, the user installs a plugin that automatically connects to the workflow configurator component (in the e-VRE system) and fetches web services descriptions managed by the e-VRE metadata manager (step 2 and step 3 in Fig. 9). The metadata manager, in turn, accesses web services descriptions in the EPOS workflows catalogue (step 4, which can be executed at runtime or off-line by

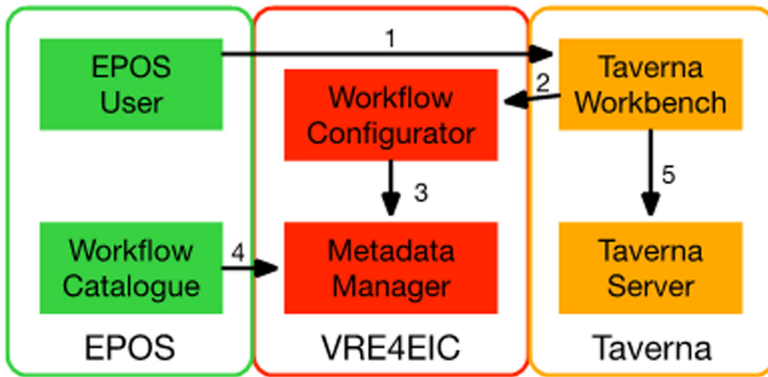


Fig. 9. Illustration of the first scenario.

ingesting information in advance). The so created workflow is then saved into the Taverna repository and executed on the Taverna Server (step 5). The description of the Workflow is also saved in the storage of the Metadata Manager, so that it can be launched later or re-used in the context of another workflow. This enables any non-skilled user to take advantage of workflows and web services from EPOS domain (potentially, from any domain) just by installing a plugin on its workflow application (in this case the Taverna Workbench).

The second scenario demonstrates the use of the e-VRE metadata catalogue to discover assets across RIs (see Fig. 10). Once the descriptions have been acquired and transformed into the CERIF format via the 3M technology described in Sect. 6.2, the user authenticates (step 1), implying his credentials being verified (2) and passed on the involved services (3). He then executes a catalogue search (4) which returns metadata records relative to resources belonging to multiple domains. The described assets can then be accessed (5) to be viewed on the appropriate viewer (e.g. for geological maps) or to be given as input to some simple local processing engine, such as waveform plotting, matlab, and the like (6).

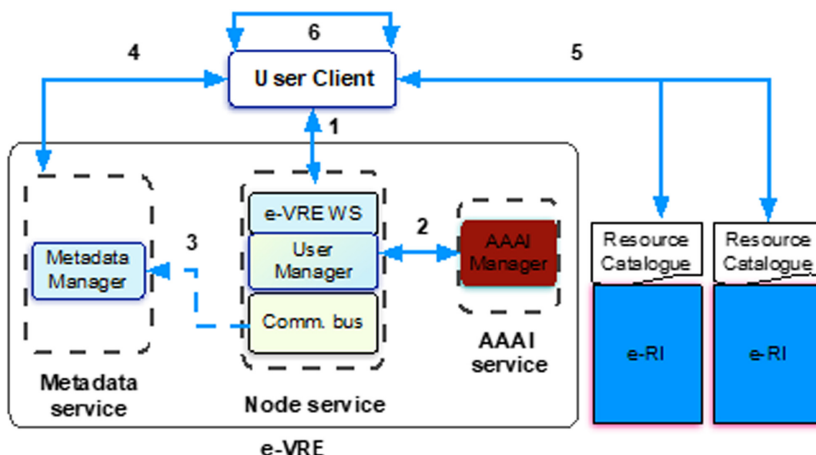


Fig. 10. Discovery of and access to assets.



## 5 Summary

This chapter reviewed briefly the origins of the VRE concept and then covered three recent EC-funded VRE research projects with relevance to ENVRI. It is clear that all the approaches share the same objective of enabling users to discover, access and re-use assets for their own purposes. All systems provide capabilities for accessing assets and composing into a workflow for deployment. D4Science concentrates on a vertically-integrated architecture but this approach has provided many domain-specific VREs. EVER-EST has also concentrated on domain-specific examples but achieves a more general architecture by the use of research objects, encapsulating the ‘working set’ of assets into one object which can then be managed and utilised. VRE4EIC provides a reference architecture and component services to achieve this, but also goes further.

Other support is required to approach a full researcher workbench including access to a communications system, office system and systems related to the management of research. The reference architecture of VRE4EIC has appropriate interfaces to achieve this.

The global recognition of the need for VREs (and similar SGs in North America and VEs in Australia) promises a vibrant future research and development activity in this area leading to better offerings for the researchers (and other user) community.

**Acknowledgements.** This work was supported by the European Union’s Horizon 2020 research and innovation programme via the ENVRIplus project under grant agreement No 654182.

## References

1. Nativi, S., Jeffery, K., Koskela, R.: Brokering with Metadata. ERCIM News 100, Special theme: Scientific Data Sharing and Re-use (2015). <http://ercim-news.ercim.eu/en100/special/rda-brokering-with-metadata>
2. Barker, M., et al.: The global impact of science gateways, virtual research environments and virtual laboratories. *Fut. Gener. Comput. Syst.* **95**, 240–248 (2019). <https://doi.org/10.1016/j.future.2018.12.026>
3. Assante, M., et al.: The gCube system: delivering virtual research environments as-a-service. *Fut. Gener. Comput. Syst.* **95**, 445–453 (2019). <https://doi.org/10.1016/j.future.2018.10.035>
4. Candela, L., Castelli, D., Pagano, P.: Virtual research environments: an overview and a research agenda. *Data Sci. J.* **12**, GRDI75–GRDI81 (2013). <https://doi.org/10.2481/dsj.GRDI-013>
5. Assante, M., et al.: Enacting open science by d4science. *Fut. Gener. Comput. Syst.* **101**, 555–563 (2019). <https://doi.org/10.1016/j.future.2019.05.063>
6. Jeffery, K., Koskela, R.: RDA: The Importance of Metadata. ERCIM News 100, Special theme: Scientific Data Sharing and Re-use (2015). <http://ercim-news.ercim.eu/en100/special/rda-the-importance-of-metadata>
7. Schuldt, H.: Multi-tier architecture. In: Liu, L., Ozsu, M.T. (eds.) *Encyclopedia of Database Systems*. Springer (2009). [https://doi.org/10.1007/978-0-387-39940-9\\_652](https://doi.org/10.1007/978-0-387-39940-9_652)
8. Martin, P., Remy, L., Theodoridou, M., Jeffery, K., Sbarra, M., Zhao, Z.: Mapping heterogeneous research infrastructure metadata into a unified catalogue for use in a generic virtual research environment. *Fut. Gener. Comput. Syst.* **101**, 1–13 (2019). <https://doi.org/10.1016/j.future.2019.05.076>

9. Michelson, B.M.: Event-driven architecture overview - event-driven SOA is just part of the EDA story: Patricia Seybold. Group (2006). <https://doi.org/10.1571/bda2-2-06cc>
10. Bailo, D., Jeffery, K.G., Spinuso, A., Fiameni, G.: Interoperability oriented architecture: the approach of EPOS for solid earth e-Infrastructures. In: 2015 IEEE 11th International Conference on e-Science, pp. 529–534. IEEE, Munich (2015). <https://doi.org/10.1109/eScience.2015.22>
11. Zhao, Z., et al.: Reference model guided system design and implementation for interoperable environmental research infrastructures. In: 2015 IEEE 11th International Conference on e-Science, pp. 551–556. IEEE, Munich (2015). <https://doi.org/10.1109/eScience.2015.41>
12. Remy, L., et al.: Building an integrated enhanced virtual research environment metadata catalogue. J. Electron. Libr. (2019). <https://zenodo.org/record/3497056>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

