Title

A Reference Architecture for enhanced Virtual Research Environments

Authors

Cesare Concordia and Carlo Meghini

Teaser

Virtual Research Environments (VREs) are rapidly becoming a common technology for supporting scientists during their research work. In order to overcome the current heterogeneity, a Reference Architecture for VREs is being developed by the VRE4EIC Project.

Body

The goal of a Virtual Research Environment **(**VRE) system is to decouple Science from ICT complexity, by providing researchers with a facility that takes care of ICT so allowing them to focus on their work. In this sense, a VRE is a fundamental component of an e-RI (e-Research Infrastructure) as it makes the resources of the e-RIs easily accessible and reusable to the community of researchers that owns the e-RI. Here, by e-RI we mean "facilities, resources and related services used by the scientific community to conduct top-level research in their respective fields"[1] while *resource* indicates any ICT entity that is of interest in an e-science community. Typically, a resource is owned by an e-RI that provides an identity for the resource and manages it, making it accessible and re-usable. Examples of resources are: datasets, workflows, algorithms, Web Services, computational or storage facilities, cloud endpoints etc. In general, a VRE is expected to:

- allow researchers to communicate with each other and to share and use the resources available in the community's e-RI
- allow researchers to advance the state of the art by building new resources as the result of processing existing resources with the available tools. Such processing may be the application of an individual piece of software to a dataset, such as the extraction of certain knowledge from a single file; or, it may result from the execution of a complex workflow obtained by combining available services, including other workflows
- allow research managers to apply economy of scale models to access and manage resources that researchers or single organizations alone could not afford.

The most advanced e-RIs have developed their own VRE, showing awareness of the crucial role that a VRE can play for their researchers. Others are currently designing their VRE. However, the number of currently existing or designed VREs is very limited; more importantly, these VREs show a great heterogeneity in scope, features, underlying protocols and technologies, partially defeating the interoperability goal that lies at the very heart of a VRE. One of the major goals of the VRE4EIC project is to overcome this issue by proposing a Reference Architecture for an e*nhanced VRE* (e-VRE). Based on a thorough analysis of the requirements of a VRE, and on the characterization of an ample range of existing Research Infrastructures, the project has individuated three logical tiers in 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)

[1] http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=what

publishing e-VRE functionalities via a Web Service API; and iii) enabling e-VRE applications to interact each others.
- The *Resource Access* tier, which implements functionalities that enable e-VRE components to interact with eRIs resources. It provides synchronous and asynchronous communication facilities.

In each tier, a set of basic functionalities has been grouped into six *conceptual components*:

- 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 **Authentication, Authorization, Accounting Infrastructure (AAAI**) component is the 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.

Each conceptual component is further structured into one or more actual software components and possibly sub-components. The complete list of these components can be found in [1], which also provides the interfaces implemented by each component, and the usage relationships between such interfaces. Each interface is in turn articulated into a set of methods, whose signature is also provided.

For instance, the Metadata Manager component consists of four sub-components, each devoted to a specific metadata type: the User, Resource, Preservation and Provenance Catalogue. The project is now entering into its development phase, during which it will provide implementation for some components, retrofitting them to existing VREs, in order to enhance these VREs.

References

1. Carlo Meghini (ed.) Architecture Design. Deliverable D3.1 of the VRE3EIC Project. Available on demand by the project coordinator (see http://www.vre4eic.eu/).