D-SA3.1 — Initial architecture: first selected components

16/05/2012

Project acronym:	VERCE
Project n°:	283543
Funding Scheme:	Combination of CP & CSA
Call Identifier:	FP7-INFRASTRUCTURES-2011-2
WP:	WP7/SA3, Scientific gateway, user interfaces and knowledge and method sharing
Filename:	D-SA3.1.pdf
Author(s):	A. Spinuso (KNMI), L. Trani (KNMI), L. Fraubert (EMSC)
Location:	http://www.verce.eu/Repository/Deliverables/RP1/
Type of document:	Deliverable
Dissemination level:	Public
Status:	Final
Due date of delivery:	15/5/12
Reviewer:	I. A. Klampanos (UEDIN)
Keywords:	SA3, gateway, portal, components.

Version	*Author*		*Date*	*Comments*
1	A. Spinuso	KNMI	2/04/2012	Initial draft, basic layout.
2	A. Spinuso, L. Trani, L. Fraubert	KNMI-EMSC	01/05/2012	Draft for internal review.
3	A. Spinuso	KNMI	08/05/2012	Edited Executive summary and Introduction, plus corrections provided by I. A. Klampanos

Copyright notice

Contents

List of Figures

Executive Summary

The VERCE scientific gateway is meant to be a community-specific web portal, enabling the use of a well defined and targeted set of tools for data analysis, applications, data collections and services. It will provide access to the underlying set of HPC and Data-Intensive resources available within the VERCE consortium. In this deliverable we'll give an overview of the seismology portal implemented within previous projects, with the joint effort of all the current participants to the SA3 work package. We will use the lessons learned from these experiences in order to evaluate the advantages of adopting similar technologies and interaction paradigms for the VERCE project's gateway developments.

We'll proceed by then describing all of the components that will be part of the VERCE Gateway enabling the execution of pre-defined workflows on a variety of datasets, offering either temporary or persistent storage facilities, depending on the users' needs and the scientific relevance of the products created. We aim to keep users focused on their research objectives, limiting their exposures to the complexities relative to the management of job execution services, security, data shipment and virtualisation. The VERCE gateway also aims to foster collaboration among scientists that will be able to describe, link and share the results obtained through the gateway, within and beyond the VERCE fences.

Below we provide a list of the core gateway's components identified and described in this deliverable.

User Interface Basically, all the applications available through the gateway's web portal, including the portal itself, giving access to all VERCE's resources

Job Management Service A high-level abstraction of the underlaying execution infrastructure, providing feedbacks to the status of the jobs submitted through the gateway

Registry of Data Resources A complete view on all the available data resources with proper descriptions and methods of access.

Registry of Workflows Each selected use case will be implemented and coded using a template and stored in a repository of templates.

Registry of Processing Elements A complete description of the worklflows' components with respect to the inputs and outputs formats, expected behaviours in terms of data consumption and production rate, reference to implementation algorithms and versioning.

The development of the gateway will require frequent interaction with the seismology stakeholders, in order to define the gateway components needed for each use case's implementation with a very high level of detail, in order to steer the technological choices that best match their requirements. We aim at creating a sustainable platform and to achieve this goal we will take into account all of those choices that will make it compatible with the already existing long-term initiatives such as: PRACE[1], EGI[2] and EPOS[3]. Moreover, the VERCE software will provide a framework in which to implement the new functionalities and the services described within the ongoing FP7 seismology project NERA[4].

[1] http://www.prace-project.eu/
[2] http://www.egi.eu/
[3] http://www.epos-eu.org/
[4] http://www.nera-eu.org/

1 Introduction

The need of interfacing the user to the backend infrastructures providing data and services to the seismologists has been debated and tackled within several projects in the past years. The experience gained working with this community allows us to identify two classes of potential users: the users looking for services to connect to and those who need graphical tools to interact with. These differences are often related to the different scientific topics they are involved in. Although in this deliverable we'll be focusing mainly on the description of the components of an interactive gateway, we are not excluding the possibility to access to this components as services. The right balance between the two access paradigms will be driven by the definition and the analysis of the scientific Use Cases. This will be an iterative process that will involve several work packages. It has been already initiated in these first six months and it will be formalised and intensified during the next phase of the project.

2 Seismology Portal projects

2.1 The Earthquake Data Portal

The Earthquake Data Portal (figure 1) provides a single point of access to diverse, distributed European earthquake datasets provided through a unique joint initiative by observatories and research institutes in and around Europe. Based on Internet-standard Java portlet and web services technologies, it enables the scientists/users to integrate and combine access to different data services.

The Data Portal aggregates portlet-based Data Explorers providing search and access functionalities to datasets from several NERIES and now NERA activities, including event parametric information from the EMSC, broadband waveform data from the European Integrated Data Archive, accelerometric parameter and waveform data from participating EuroMed accelerometric networks, and historical data from the Archive of Historical Earthquake Data.

Data requests and the resulting datasets are managed in user Data Carts. These Data Carts are an important abstraction within the portal and for future developments. They maintain the URIs to references to the downloadable result datasets, and also provide a loosely-coupled mechanism for inter-portlet communication and coordination. Additionally, by exposing the data cart services, they support a mechanism through which access to remote processing systems can be integrated into the portal, as it will be shown in the following section.

The portal is built following the principles of Services-Oriented Architectures[5]. The data explorer and data cart portlets provide interactive User Interface (UI) components that in turn invoke supporting data services. These data web services support both the portal and portlet applications, but are also available as public services. As such, external applications can be — and have been — written to invoke the data services directly. These external client applications are examples of the next generation of research applications that will directly access data stores and invoke distributed remote processing centers. Operated by both EMSC and Orfeus, the Earthquake Data Portal has been installed at the EMSC and is included as part of its default available systems, ensuring its continued operational presence, supporting the broad seismological community and beyond.

2.2 RapidSeis: Enabling User-Defined Seismological Waveform Data Processing

The objective of this JISC-funded pilot project aimed at providing the seismological community with a simplified system that overcame important barriers such as installation and understanding of the analysis package, location and transfer of large amounts of input data and visualisation of results.

[5] http://en.wikipedia.org/wiki/Service-oriented_architecture

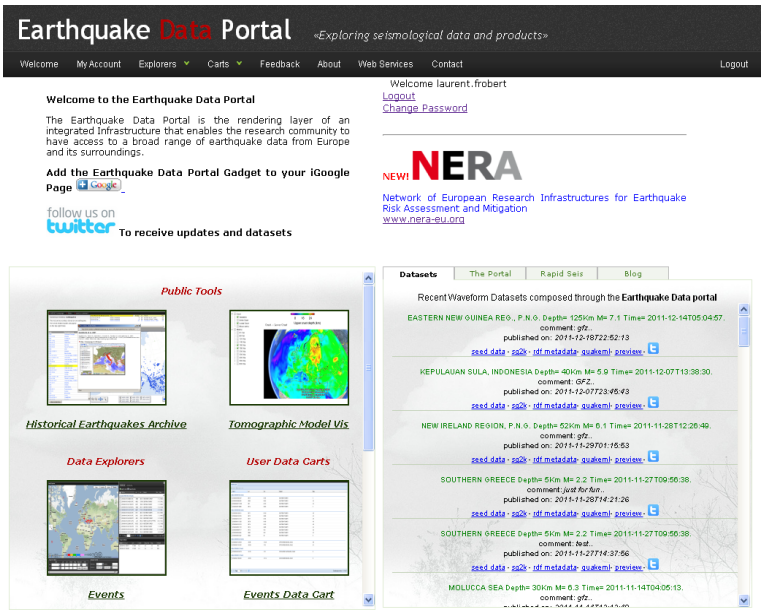

Figure 1: The EarthQuake Data Portal Home Page, http://www.seismicportal.eu.

The project combined the expertise of three distinct teams. The Orfeus Data Archive provided discovery services and access management to distributed earthquake waveform data and metadata, supported by technologies such as webservices, SPARQL/RDF and a JSR-168[6] portal framework based on the WSRP portlet architecture. Analysis and visualisation of the waveform data were performed by SDX (Seismic Data eXplorer) developed at the University of Liverpool. SDX formed the core processing engine - this application's functionality can be extended via user-defined algorithms coded as plugins. The processing engine was supported by several webservice-enabled modules used for data and plugin exchange. The user, data and processing infrastructure were connected using the Rapid framework created by the UK National e-Science Centre. Rapid generates intuitive interfaces to the processing core in the form of two self-contained JSR-168 remote portlets. One portlet facilitated plugin creation, whilst the other allowed execution of plugin code within the processing core running on a campus or national Grid infrastructure. The RapidSeis system was deployed for the demonstration of the pilot release within a community gateway: the Earthquake Data Portal [7].

Although RapidSeis originated within the seismological community we suggest it forms a framework which could be exploited in different domains or scientific gateways where users can discover data, store data within a cart and wish to apply specialised processing algorithms remotely through a web browser. Processing performed on Grid infrastructure is completely transparent: the execution portlet handles all the management of the submission and the scheduling of the processing jobs providing to the user the visualisation of the results. The processing algorithms can be created online, stored and shared between users of the community web portal. The data and algorithm metadata are described in RDF and exchanged within the architecture which is commonly geographically distributed.

The concept behind this pilot project match as many aspects of what a scientific gateway for VERCE should aim to provide. The ADMIRE [8] platform is already providing the technology to semantically describe the plugins, that within the VERCE scenario could be seen as new processing elements. The Predefined composite workflows developed within VERCE could require that the behaviour of the single

[6] http://community.java.net/portlet/
[7] http://www.seismicportal.eu
[8] http://admire-project.eu

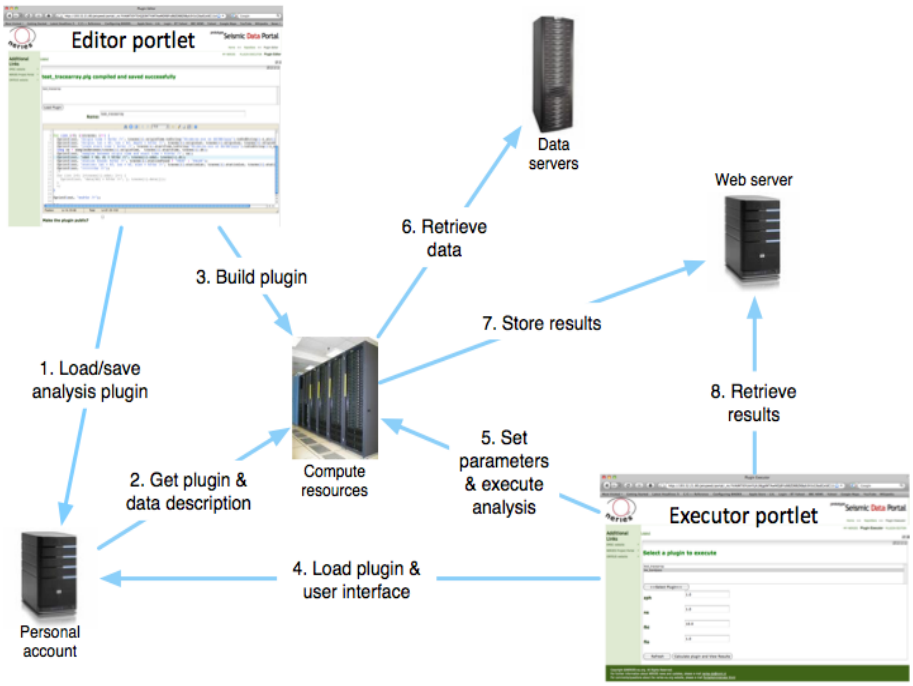

Figure 2: RapidSeis workflow.

analysis components might change and these changes could be delivered by the scientists themselves as new plugins.

3 VERCE Gateway components

3.1 User and Security Management

Each user of the VERCE platform will be registered through the gateway. So far we foresee four classes of users associated to different level of insurance: Administrator, Simple User, Registered User and Certified User.

The Administrator will have the role of to handling certificates requests from the Registered Users and he will be responsible of creating and uploading informative content within the portal. It will be decided later wether we need to split this role in two: User Administrator and Content Manager.

The Simple users will be those visiting the public pages of the VERCE Gateway. They can register to the portal by filling in the registration form online. Once the registration process succeeds they'll become Registered Users and will have access to a limited set of features and information that will not require access to any of the HPC or Data Intensive resource. Accessing to these facilities will require applying for a security certificate.

Registered users should be able to browse the VERCE Registries of Resources and Intermediate Archives, possibly downloading the data and the software libraries available, depending on the policies and licences of individual items.. This will encourage the use of the VERCE products also outside the VERCE computing infrastructure.

Applying for a security certificate will require to comply with the directives determined by SA1 and SA2. After a successful application will succeed the certificate will be provided and the user will be considered as a Certified User. This will allow the user to access to all the functionality that will be specified from the use cases defined by JRA1 and NA2, including data staging and workflow execution scenarios, intermediate results evaluation, description and preservation.

In general the authorisation and identification of users among distributed organisations and infrastructures is an active topic for important initiatives, involving several research communities. Therefore it would be beneficial for the success of VERCE to engage a discussion on a broader scale with other initiatives, such as the one described in [Broeder et al., 2012] on Federated Identity Management.

3.2 User Interface components

One of the main objectives of VERCE is to be a user driven project This means that the whole design and implementation process is being conducted with the close collaboration and involvement of the main stakeholders. This approach is the most appropriate given the ambitious goals set by this Europe-wide collaborative effort, in order to succeed establishing such a complex platform, it is important to try to shape and tailor it to the needs and requirements of the end-users. Therefore one of the main concerns to address is the design of the user interface, so as it becomes straightforward and more user-friendly as possible. The intention is to try to minimize the elements of complexity, which might be intrinsically introduced by new powerful features, and keep a close relationship in terms of look and feel with the tools that are already available and widely assessed in the seismological community. The VERCE scientific gateway will provide an unique access point to a wealth of applications and tools, each one specifically designed to fulfil a precisely-defined purpose. The users will make use of an integrated environment that will allow the definition and execution of complex analysis and simulation tasks, which will be identified within a range of selected use cases. We do foresee though, that a number of functionalities will be common to all the applications. In particular there is a general need to interact with the execution. The scientist would like to have a close control of the execution flow especially at specific points where human intervention is needed to discriminate among partial results and/or retune and adjust the performance. In this perspective it becomes vital to have an effective provenance tracking mechanism that allows storing the information needed to play smart re-runs ([Freire et al., 2006], [Altintas et al., 2006]) or to resume the execution after user-defined checkpoints. The interface will have to accommodate tools to allow a quick visualization and analysis of partial results. Ideally we should also look into the possibility to plug in automatic procedures to discern and discriminate among partial values, allowing the retuning of predefined parameters. In some cases an even deeper user interaction might be needed, in particular allow partial modification of the workflow by changing some components and/or their execution order. In this case the interface should provide the user with a set predefined components to chose from (see 3.7).

3.3 Service components

The VERCE Platform adopts the ADMIRE data-intensive architecture to provide access and execution to a set of predefined scientific workflows. In order to fulfil the requirements of the selected seismological use cases, the platform has to be broad enough and highly adaptive towards the heterogeneous ecosystem of available resources and computational infrastructures. We identified the following components as major features that the scientific gateway, here under definition, should accommodate:

- Job Management Service
- Registry of Data Resources
- Registry of Workflows
- Registry of Processing Elements

3.4 Job Management Service

The ADMIRE architecture provides a high-level abstraction of the underlaying execution infrastructure by separating concerns among targeted users. In principle ADMIRE implicitly contains those concept required by the VERCE Platform such as supporting the transparent enactment on multiple sites. The current implementation, though still relies on a single execution environment based on OGSA-DAI, which appears restrictive given the diverse and heterogeneous computational requirements. One of the first improvements towards a more complete platform should aim at extending the current ADMIRE technology introducing support for multiple execution engines. Crucial activity should be is the identification of technical solutions that better suit the VERCE use-cases. A selection of some technologies is presented in D-JRA2.1, but the actual implementation should be directly derived from specific needs while still keeping a generalisable approach.

From the point of view of the scientific gateway, the job management is an almost transparent service to the final user. This component will choose the appropriate middleware, among those available in the platform, to interact with the specific computational infrastructure. In other words the user will not be able to choose where to run a specific job, but this choice will be accomplished during the workflow definition. A use case scenario could in principle contain computational elements to be executed on different HPC infrastructures, potentially adopting different scheduling and monitoring strategies and softwares.

At the moment we don't anticipate a fully-fledge user-interface will be required (e.g.: complete workflow composition, graphic tools for workflow design etc), but instead a more "guided" approach, where a possible solution could be realised through QoS parameters and/or checkpoints combined to intermediate storage steps. But given the heterogeneity of the execution environments, identifying a minimum set of functionalities to be exposed to the users is not a trivial task.

Another aspect that could be considered account is the runtime monitoring of the overall platform and the dynamic provisioning, for more details on this issue we refer to D-SA1.1 and D-SA2.1. Concerning to this problem the gateway will have to accommodate a mechanism to assign priorities and policies to a single or groups of users.

3.5 Registry of Data Resources

The VERCE Platform will contain a complete view on all the available data resources with proper descriptions and methods of access. The mechanism to store and present this information has been identified in a central registry, already present in ADMIRE. The VERCE registry will be an extension of the ADMIRE one. Particular attention will be given on the administration interface that will allow to properly register new resources and make them available to the whole community. A specialization or subset of these resources is represented by the Intermediate Data Archive.

3.5.1 Intermediate Data Archives and Preservation.

We expect that the workflow execution results (final or intermediate) should be preserved under predefined policies end the technologies to describe and maintain these data are currently investigated.

The processing needed to perform the analysis required by the use cases, will bring along all sorts of information about the creation process that will contribute to the results. It is useful for validation and reproducibility to understand what happened during the execution of the workflow, where several distributed components are used to accomplish complex computational tasks. Most of the scientific workflow systems (VisTrail, Taverna) are aware that preserving execution data is extremely relevant when validating the obtained results, which is a fundamental requirement of any scientific computation.

Information such as provenance traces and the the description of the workflow configuration that generated the results, are part of a package that is considered to be informative as a whole. This whole package has has to be considered as single digital unit that will make the the research object reproducible. Each VERCE user should be able to use these partial results within the execution of a workflow (full or partial), to create other products that might eventually turn into new research objects as well.

To achieve these objectives, VERCE has to find solutions to answer problems related to information management infrastructures, workflow preservation, provenance, database integration and persistent storage have to be tackled realistically. For workflow preservation and provenance we will be looking at solutions that converge to a single and interoperable framework (eg. The Open Provenance Model)[9] [Moreau et al., 2008, Ellqvist, 2010] for recording provenance data. We need the proper abstraction to improve interoperability among workflows systems running within VERCE, towards a better understanding of the execution of a scientific computation that connects different tools, processing services, databases and computational infrastructures.

Database integration and persistence storage will be achieved by investigating products such as Fedora Commons[10], dCache[11], iRods and also no-sql databases, in order to reach a flexible and extensible solution to describe, reference and generate relations among the produced data, possibly stored on different locations. The storage will most likely hosted outside the fences of the HPC clusters, as those will not provide data preservation services.

Several studies on workflow preservation are currently on-going within the Wf4ever project[12]. VERCE will look into the proposed solutions to realise what could realistically help in achieving these goals.

3.6 Registry of Workflows

Each selected use case will be implemented and coded using the DISPEL language as a template and stored in a repository of templates. The workflow will then be coupled with the proper user interface and/or tool, providing users with features useful to the specific computation. Some tools could provide the options to customise aspects of the processing elements, choosing from a fixed selection (3.7), while others might allow deeper modifications to the workflow itself.

3.7 Registry of Processing Elements

VERCE aims to offer predefined workflows implementing specific use cases, but this should not exclude the possibility to adjust and perform slight modification of some of the processing steps. This functionality will be implemented so that it allows the selection of some of the processing elements from a dedicated registry. This registry will contain a complete description of the components with respect to the inputs and outputs formats, expected behaviors in terms of data consumption and production rate, reference to implementation algorithms, versioning.

3.8 Collaborative components

An environment that aims to foster collaboration among people that are working or interested in the same research objective, needs tools such as forums, dataset annotation interfaces, data-sharing functionalities and potentially even personal profile pages. The question is, how to use these services within the VERCE gateway? To which extent should these collaborative interactions be limited within VERCE? Users might want to publish preliminary results on their digital profile outside the VERCE portal, or they might just

[9] http://openprovenance.org/
[10] http://fedora-commons.org/
[11] http://www.dcache.org/
[12] http://www.wf4ever-project.org

Figure 3: Seismic waveform and event parameters available through the Earthquake Data Portal can be published and shared on Twitter.

prefer to post the link to the descriptive page of an interesting dataset via email. All this should be easily achieved assuring the full control on the VERCE products lifecycle.

For example, if reusing a dataset generates additional metadata, we might want to keep the audit-trail related to the history of these metadata variations, besides allowing to comment and share the updated results. Moreover, under well defined policies, VERCE might provide the opportunity to the users to deploy and run their own algorithms as workflow components, like the RapidSeis concept of section 2.2. This might lead to other interesting scenarios where users could deploy and share algorithms to be tested and executed within multiple workflow setups (or Use Case Implementations). The users' interaction on these items has to be recorded within VERCE and possibly shared with the outside world.

Many features are currently supported by most of the modern social networks, from the more generic ones (Twitter, Facebook, G+) to the more specific, such as MyExperiment[13]. All the collaborative activities such as sharing dataset URIs, like in Figure 3, or discussions on results could possibly make use of third party platforms that already provide very advanced functionalities, (Figure 4). This is possible thanks to many available widgets and APIs, provided by most of these social networks, that embed their collaborative elements within other websites, making those somewhat bound to the users' digital identity. We will look for a collaborative paradigm that exploit and integrates these platform within the VERCE gateway, also keeping an eye on privacy matters concerning the VERCE Users.

4 Gateway Metrics

4.1 Usage related measurements (access)

To be able to measure the usage of the different services provided by the portal we have to make use of logs. Accordingly with the technology provided by JRA2, we will integrate for each service a log system to keep relevant information. Based on these logs we could then generate responses. Our log system should log the following informations:

- number of user

[13] http://www.myexperiment.org

Figure 4: Facebook Comments Plug-In in Wordpress.

- period of activity
- number of job created
- number of job terminated
- number of job interrupted
- size of data generated
- most used services

4.2 Performance related measurements (access)

The gateway will provide a dedicated page to visualise and access the information collected about the performance indicators proposed in D-SA2.1. This sort of dashboard will allow a clear overview of the current status of the running jobs, with the possibility for the user to define and display aggregates and time averages of the measurements of interest. Visual monitors will give an immediate indication of the progress of the entire workflow helping to identify critical parts in order to provide the user with immediate feedback which should aid rerunning and improving the overall process.

5 Development Strategy in cooperation with other WP

5.1 Use Cases Development

The definition of the Use Case specification provided by WP2 will steer the developments of the scientific application provided through the VERCE Portal. SA3 has the task of developing the user interface components that allow the setup of the workflow's parameters and input data for the specific application through the Gateway. As suggested from WP3 within the D-SA2.1, the release management process will follow a PDCA (Plan, Do, Check, Act) approach, that should help in producing a sound platform within

VERCE's time limits. WP7/SA3 will align the developments to these overlapping cycles of Figure 5 in close relationship with JRA1, NA2 and JRA2, since the first two work packages will provide the information needed to properly design the user interface, while the latter will provide guidelines and technological support for the implementation of the needed functionalities and components mentioned in section 3

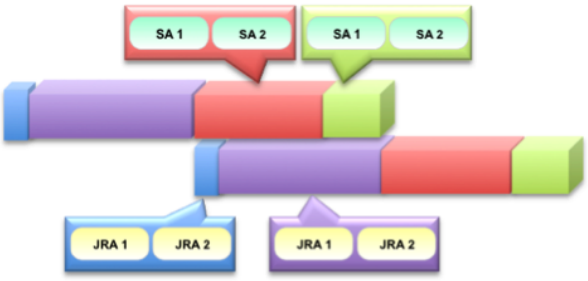

Figure 5: PDCA and overlapping cycles within VERCE. Plan=red, Do=violet, Check=red, Act=green.

In coordination with SA2 we'll consider the execution of acceptance tests, defined after a deep analysis of the requirements gained during the definition of the VERCE Use Cases. This will be relevant to identify the minimal conditions under which the developed interface will be considered to fulfil the expectations. At the current stage the use case that we foresee, as already mentioned within the *User Interface Components* 3.2 section, consist in enabling the user to get updates on the status of the execution of the workflow and to visualise the results, describing and storing those when needed, within the intermediate archives. Depending on the requirements, the user interface might give the possibility to change the workflow components, by either inverting their execution order or by selecting a completely different one from a registry of suitable alternatives.

5.2 Gateway's components impact - current status

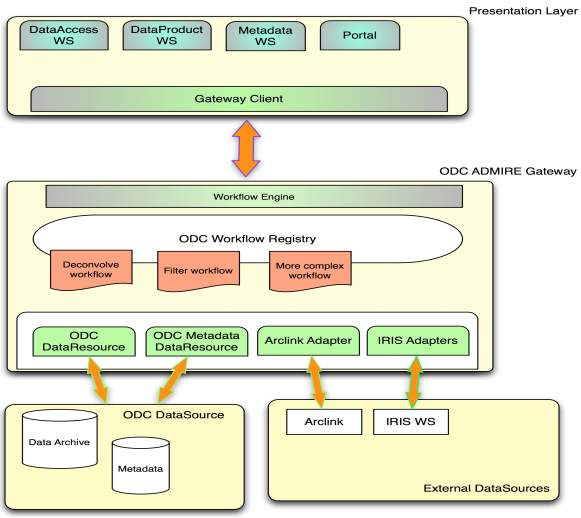

Figure 6: Orfeus prototype installation of the Admire Gateway.

During this first iteration one of the first activity we addressed was to apply the ADMIRE technology in a real use case scenario. In particular at ODC we focused on the possibility to deliver new data products to the users of data archives using the ADMIRE software stack expressly adapted for the task. This is also related to the new functionalities described within the NERA project. Therefore we installed an ADMIRE gateway at the datacenter and developed some adapters to our data and metadata archives. In

addition we implemented a few simple workflows with DISPEL embedding the procedures to integrate and process the data needed to assemble the final product. We implemented also a RESTful[14] service layer on top as an access point to the infrastructure. With this scenario in mind, we emulated a simplified version of what we envisage the VERCE Platform would be in the future. Although at this stage we did not interface with external high performance computational resources and we did not deal with any security issue, we concluded that ADMIRE is a flexible framework which allows, with relatively little effort, developing and executing scientific workflows to create seismological products.

References

[Altintas et al., 2006] Altintas, I., Barney, O., and Jaeger-Frank, E. (2006). Provenance collection support in the kepler scientific workflow system. In Moreau, L. and Foster, I., editors, *Provenance and Annotation of Data*, volume 4145 of *Lecture Notes in Computer Science*, pages 118–132. Springer Berlin / Heidelberg.

[Broeder et al., 2012] Broeder, D., Jones, B., Kelsey, D., Kershaw, P., L§ders, S., Lyall, A., NyrŽnen, T., Wartel, R., and Weyer, H. J. (2012). Federated identity management for research collaborations. oai:cds.cern.ch:1442597. Technical Report CERN-OPEN-2012-006, CERN, Geneva.

[Ellqvist, 2010] Ellqvist, T. (2010). *Supporting Scientific Collaboration Through Workflows and Provenance*. Linköping studies in science and technology: Thesis. Department of Computer and Information Science, Linköpings universitet.

[Freire et al., 2006] Freire, J., Silva, C., Callahan, S., Santos, E., Scheidegger, C., and Vo, H. (2006). Managing rapidly-evolving scientific workflows. In Moreau, L. and Foster, I., editors, *Provenance and Annotation of Data*, volume 4145 of *Lecture Notes in Computer Science*, pages 10–18. Springer Berlin / Heidelberg.

[Moreau et al., 2008] Moreau, L., Freire, J., Futrelle, J., McGrath, R. E., Myers, J., and Paulson, P. (2008). The open provenance model.

[14] http://en.wikipedia.org/wiki/Representational_state_transfer

Glossary

API Application Program Interface, an inter-software communication specification used for accessing functionality or services from programs.

component One of the computational elements involved in a data-intensive or computational process, such as: application codes, scripts, workflows, services, catalogues, registries, data collections, data resources, functions, gateways, libraries, PEs, PE instances, format definitions and types.

data archive The long-term storage of scientific data and methods.

data-intensive An adjectival phrase that denotes that the item to which it is applied requires attention to the properties of data and to the ways in which data are handled.

gateway A software subsystem, typically at the middleware level, that accepts requests for computational and data-handling tasks. It vets those requests to establish whether they are valid, e.g. are syntactically and semantically consistent, and are authorised. Requests that are not validated are rejected. Requests that are accepted are passed to other software systems, at the same or other locations, for execution. The gateway may partition and translate requests in order to combine heterogeneous services.

grid A system that is concerned with the integration, virtualisation, and management of services and resources in a distributed, heterogeneous environment that supports collections of users and resources (virtual organisations) across traditional administrative and organisational domains (real organisations).

high-performance computing (HPC) Use of powerful processors, high-speed networks and parallel supercomputers for running computationally intensive applications.

metadata Data that describes data. Metadata may include references to schemas, provenance, and information quality. In Seismology, metadata may also refer to data required in order to sanitise a seismograph's response.

myExperiment Collaborative virtual research environment for sharing scientific workflows.

OGSA-DAI Open Grid Service Architecture Data Access and Integration, an open source product for distributed data access and management.

ORFEUS Observatories and Research Facilities for European Seismology.

portal In the context of knowledge discovery, a tool designed for a particular group of domain experts that can be used via their browsers; it enables them to enables them to establish their identity and rights, and to pursue conveniently a set of research tasks for which the portal is designed.

processing element – PE A software component that encapsulates a particular functionality and can be used to construct a workflow.

registry A persistent store of definitions and descriptions of data or software components and their relationships accessed by tools and other elements of a distributed research environment. It is intended to facilitate discovery and use of the components.

repository A store holding software definitions, other shared code and data, that supports distributed concurrent access, update and version management.

research object A research item which some researcher wishes to identify. It may be a collection of primary or derived data, code, a workflow, a service, an ontology, a set of metadata, etc. It may be a paper or a talk. Often it is a composition of such elements.

science gateway A consistently presented set of facilities designed to be a convenient working environment for researchers in a particular domain, in this case seismology. It should bring together access to all of the capabilities and resources such a researcher needs: including catalogues of available data and tools, established methods and arrangements for applying them with specified parameters to specified data.

Taverna Open source scientific workflow management system.

VERCE architecture A high-level and coherent design for the VERCE e-Infrastructure; it evolves as the seismological goals and digital environment evolve and become better understood. It should guide the development of successive VERCE Platforms.

VERCE e-Infrastructure An envisaged result of VERCE, as an integrated computational and data environment that presents a coherent virtual research environment in which to conduct seismology research and eventually research in other Earth sciences.

VERCE Platform The current realisation of the VERCE e-Infrastructure at any time in the VERCE project. Initially this is not fully integrated and may only constitute a partial implementation. Nevertheless, it is sufficient both to pursue research identified as priority seismology use cases and to develop and test the design of the VERCE e-Infrastructure. The VERCE Platform is an approximation to the VERCE e-Infrastructure. These approximations should converge on the VERCE e-Infrastructure by the end of the VERCE project. virtual research environment (VRE) A presentation of (ideally all of) the resources a researcher may need in a consistent and easily used form. These resources include catalogues, data, metadata, libraries, tools, workflows, programs, services, visualisation systems and research methods.

W3C World Wide Web Consortium, an international community of member organisations and the public that works to define and promote standards for web technologies.

web service A software system designed to support interoperable machine- or application-oriented interaction over a network.

workflow A process of composed data-handling tasks, computational tasks and human interactions intended to implement a research method or established working practice.