17 WS-PGRADE/gUSE in European Projects

Tamás Kiss, Péter Kacsuk, Róbert Lovas, Ákos Balaskó, Alessandro Spinuso, Malcolm Atkinson, Daniele D'Agostino, Emanuele Danovaro, and Michael Schiffers

Abstract. Besides core project partners, the SCI-BUS project also supported several external user communities in developing and setting up customized science gateways. The focus was on large communities typically represented by other European research projects. However, smaller local efforts with the potential of generalizing the solution to wider communities were also supported. This chapter gives an overview of support activities related to user communities external to the SCI-BUS project. A generic overview of such activities is provided, followed by the detailed description of three gateways developed in collaboration with European projects: the agINFRA Science Gateway for Workflows for agricultural research, the VERCE Science Gateway for seismology, and the DRIHM Science Gateway for weather research and forecasting.

17.1 Introduction

Besides developing a core science gateway technology and customization methodology, the other major objective of the SCI-BUS project was to build a large number of science gateways in diverse disciplines. These gateways not only demonstrate the applicability of the core technology, but most importantly support scientists, companies, and citizens in solving complex problems from user friendly environments on distributed computing infrastructures. Based on the relationship between the SCI-BUS project and the targeted user community, the developed gateways fall into three different categories:

1. **Gateways for core SCI-BUS partners**: Representatives and technical experts of these user communities were partners in the SCI-BUS project and were involved in the work from the very beginning. Part II of this book describes several gateways in this category, including the computational neuroscience gateway, the MosGrid gateway, the statistical seismology science gateway, the VisIVO gateway, and the heliophysics gateway. Also, in Sect. 3.4 two commercial gateways in this category are introduced, the eDOX Archiver Gateway and the Build and Test Portal. Four further gateways, the iPortal for the Swiss proteomics community [Kunszt, 2013], the PireGrid Community Commercial Gateway (http://gateway.bifi.unizar.es), the Renderfarm.fi Blender

Community Rendering Gateway (http://www.renderfarm.fi/) and the SimBusPro gateway (http://simbuspro.com:8080/liferay-portal-6.1.0/) complete this list (although their description is not included in this book).

2. **Gateways for SCI-BUS subcontractors**: To extend its user community, the SCI-BUS project ran an open call that resulted in six further communities joining the project as subcontractors. The gateway of one of these communities, the Condensed Matter Physics Community Science gateway, is described in Sect. 2.6. The other subcontractor gateways are the Weather Research and Forecasting science gateway [Blanco, 2013], the Institute of Metal Physics science gateway [Gordienko, 2013], the Academic Grid Malaysia scientific gateway [Pek Lee 2013], the SCI BUS Adria science gateway (http://adria-sci.irb.hr:8080/liferay-portal-6.1.0/), and the ChartEx gateway (http://openml.liacs.nl:8080/liferay-portal-6.1.0/web/liacs/).

3. **Gateways for external user communities**: The SCI-BUS project run an application and user support service that, besides core and subcontracted project partners, was looking for and supporting external user communities who could benefit from the SCI-BUS technology. The main target of this service was large communities, particularly in or associated with other funded European projects. However, SCI-BUS also supported several local activities (for example, teaching and learning gateways) that have the potential to be disseminated to wider communities.

This section concentrates on this third category. Three representative examples of large European projects, agINFRA, VERCE, and DRIHM, that use SCI-BUS technology as the result of intensive support and collaboration with those projects are described below. However, several other projects and communities have also successfully built, or are in the process of building, WS-PGRADE/gUSE based gateways. A summary table of these external community gateways directly supported by the SCI-BUS project is provided in Table 17.1. Some of the gateways are mentioned or described in various levels of detail in other parts of this book as examples or illustration of the utilization and different features of WS-PGRADE/gUSE. Please, note that the majority of these gateways are still under development at the time of writing this book with more or less advanced prototypes or beta releases that are being operated.

Table 17.1 External community gateways supported by the SCI-BUS project

Gateway name	User community and application area	Type of gateway – WS-PGRADE/gUSE technology utilized
agINFRA WS-PGRADE/gUSE gateway	Agricultural research and information management communities, agINFRA FP7 European project (http://aginfra.eu/)	Application-specific gateway using the Remote API from existing Drupal portals
VERCE science gateway	VERCE FP7 European project, Virtual Earthquake and seismology Research Community in Europe e-science environment (http://www.verce.eu/)	Application-specific gateway using the ASM API from newly developed custom user interfaces and heavily relying on the multi-DCI capabilities of the DCI Bridge
DRIHM science gateway	DRIHM FP7 European project, Distributed Research Infrastructure for Hydro-Meteorology (http://www.drihm.eu/)	Application-specific gateway using the ASM API from newly developed custom user interfaces and heavily relying on the multi-DCI capabilities of the DCI Bridge
CloudSME gateways	CloudSME FP7 European project, simulation applications for manufacturing and engineering SMEs (http://cloudsme.eu/)	A set of application-specific gateways using both the ASM and the Remote API based on the actual use-case. Heterogeneous cloud infrastructures are used via the CloudBroker Platform
SHIWA simulation platform	SHIWA (https://www.shiwa-workflow.eu/) and ER-FLOW (http://www.erflow.eu/) FP7 European projects, interoperable workflow solutions from diverse disciplines	Generic WS-PGRADE/gUSE gateway with direct connection to the SHIWA workflow repository and submission service to support the creation and execution of metaworkflows
QPORT gateway	High throughput technologies in life-sciences and bioinformatics analysis, Quantitative Biology Center (QBiC) of the University of Tübingen	Application-specific gateway using the ASM API and modules from the iPortal for the Swiss proteomics community
University of Portsmouth Teaching and Learning gateway	University teaching in the areas of creative studies, e.g. animation rendering or video encoding	Application-specific gateway using the ASM API and local resources of the university (desktop grid, cluster)
University of Westminster Desktop Grid gateway	University teaching in various disciplines, e.g. molecular modelling and animation rendering	Application-specific gateway using the ASM API

Gateway name	User community and application area	Type of gateway – WS-PGRADE/gUSE technology utilized
Fusion Virtual Research Community gateway	Nuclear fusion research community of EGI	Application-specific gateway using the ASM API
AutoDock gateway	Open public gateway for the molecular docking studies	Application-specific gateway using the ASM API
Gateway for massive online course (MooC) for grid Computing	Introductory course for concepts and practices of large-scale computing on grid infrastructure	Generic WS-PGRADE gateway installed and configured on a virtual machine
STARnet Gateway Federation	A network of science gateways to support the astrophysics community	A network of application-specific gateways using the ASM API

17.2 The agINFRA Science Gateway for Workflows

The main objective of the agINFRA project [Pesce, 2013] in the EU 7th Framework Programme is to elaborate and provide an advanced, open, and sustainable data infrastructure with services and tools for agricultural research and information management communities. In the project, the studies on stakeholder needs show clearly that agricultural research communities are located in a wide, multidisciplinary array of research domains, ranging from molecular genetics to geo-sciences, including also social and economic sciences. Since the scope of the project might be considered relatively wide, the presented work focuses on two important stakeholders groups:

- ICT and information managers of shared domain- and subject-based repositories including service providers (e.g., aggregators), and
- developers and providers of IT applications for research and related tasks who may be interested in customizing or developing further agINFRA components and tools.

The current agINFRA infrastructure relies not only on cloud and grid resources but various geographically distributed data sources; registries, pre-existing and new data management, processing, and visualization components are also involved in the infrastructure. Complex workflows have been identified at four levels according to the granularity of individual tasks; starting from the "Workflows of Researchers" including their everyday research and publication activities (Level 1), and ending with the "Workflows of agINFRA Integrated Services" (Level 4). This section deals with the orchestration of "Workflows of data/metadata aggregators and service providers" (Level 3) on the agINFRA infrastructure that form the base of several new integrated services/components to be accessible from user-friendly web portals.

In the project a WS-PGRADE/gUSE-based gateway serves the two above described stakeholder groups with special focus on supporting complex workflows for agINFRA integrated services. As a complementary solution, in order to address directly some end-users (researchers), a different science gateway has also been developed by INFN (http://aginfra-sg.ct.infn.it) for legacy agriculture applications based on the Catania Science Gateway Framework [Barbera/2010].

17.2.1 Required WS-PGRADE/gUSE features

The success of research data infrastructure initiatives depends on understanding in-depth the needs of the intended users, therefore serious efforts were invested in this activity to learn the most relevant technologies and directions, such as CIARD RING. The CIARD - Coherence in Information for Agricultural Research for Development initiative was launched in 2008 by (among others) the Food and Agriculture Organization of the United Nations (FAO). CIARD aims at improving the access to agricultural research and offers the Routemap to Information Nodes and Gateways (RING) [Pesce/2011], where participants world-wide are able to register their information access services, for example repositories, information feeds, etc. Currently more than 1000 services and more than 250 data sources are available from almost 500 partners in the RING. The CIARD RING has been enhanced by agINFRA partners, and become a so-called integrated service that takes on the role of a central broker.

The targeted users and developers are heavily using the CIARD RING and they are mostly familiar with Drupal web portal (https://drupal.org) interfaces. Therefore, as the main concept; the workflow-oriented services of WS-PGRADE/gUSE are to be accessed mostly via its Remote API from Drupal based portals to enhance the CIARD RING provided functionalities.

Moreover, robot certificates are crucial for this community in order to simplify the access by the non-grid aware users. Concerning the workflow support, the conditional branches were proven to be vital (as will be illustrated in the aggregator workflow) when the requirements have been captured. In agINFRA, several data management and processing components have been implemented as REST service. The native support in WS-PGRADE/gUSE enabling the easy invocation of such services at the job level was an important new requirement from the community, too.

17.2.2 Cross-Community Workflow for Agricultural Metadata Aggregation

The presented use-case and example is intended to show how to integrate distributed computing infrastructure (for data processing and storing purposes) together with metadata stores (registered in CIARD RING) containing bibliographical and educational materials to achieve an interoperable cross-community platform for agricultural web portals. One of the major tasks addressed was to enable transparent search and browse in different metadata stores using completely different

metadata schemas. One other task concerning the differences among metadata stores was to solve feeding them with new metadata records interactively (involving the users). This latter case is called the harvesting process, and it contains methods (i) to fetch all the records belonging to the repository, (ii) to index them based on vocabularies in order (iii) to facilitate the further searches and filtering methods, and (iv) to eliminate all the duplicated records. This complex process has been implemented as a gUSE workflow, and the involved components during this process are shown in Fig. 17.1.
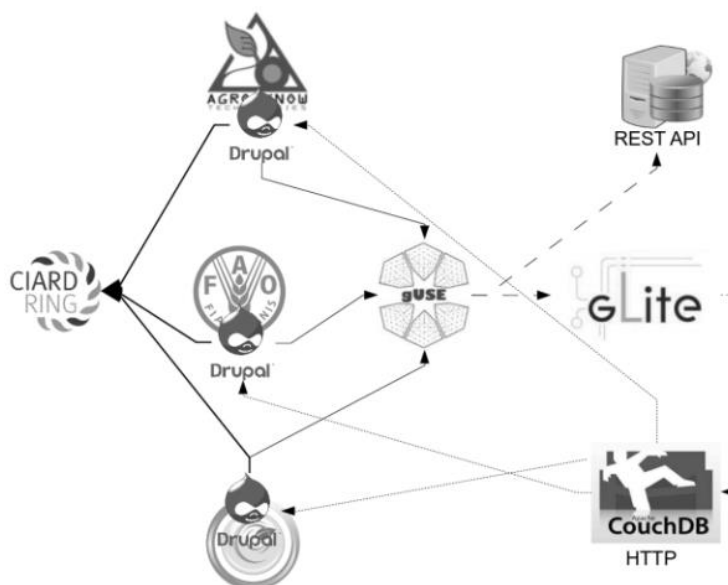
Fig. 17.1 Architecture of remote API-based SG solution in agINFRA

CIARD RING consists of several repositories of metadata that store records in various fields of agriculture. As all the partner companies (e.g., AgroKnow) or public institutes (such as FAO) involved in the agINFRA project offer content management systems (mainly Drupal) as user interfaces, a Drupal module has been developed as additional interface for gUSE to search and to browse in the catalogue of CIARD RING (thick lines), and to be able to submit and interpret the harvesting workflow without using the general gUSE User Interface (see thin lines on the figure). Then, gUSE enacts the workflow by executing harvester and transformator components that are predeployed on the gLite-based distributed computational resources and by invoking REST API web services (denoted as dotted lines on the figure). Records are placed on the distributed computing infrastructure and are registered in the logical file catalogue (LFC). To avoid exposing users to the complicated security issues required by the applied technology, automated processes are triggered by the register process to publish the records for other data

components outside of the grid using CouchDB (couchdb.apache.org). The workflow developed for this scenario is shown in Fig. 17.2.
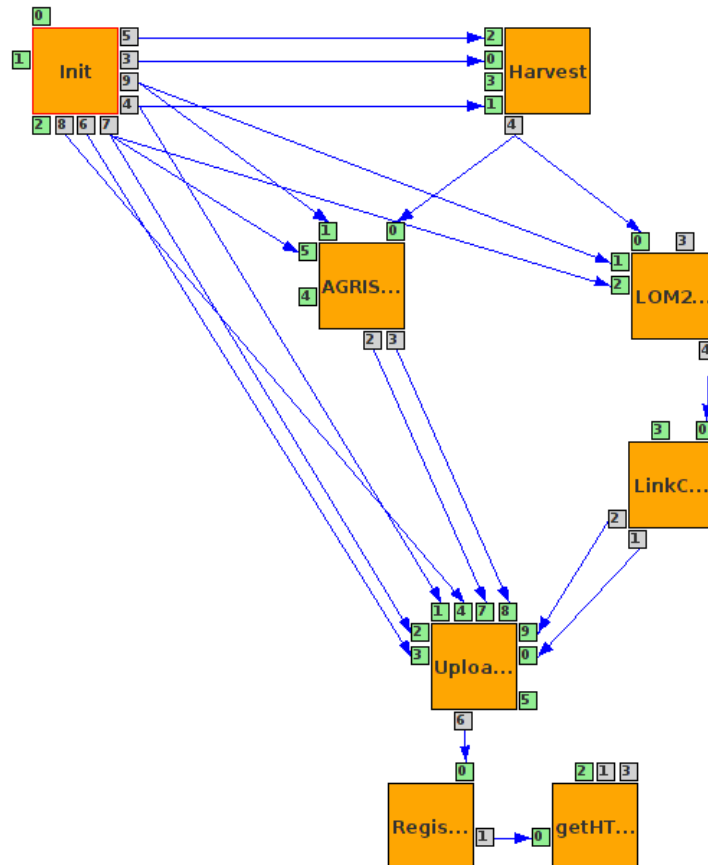

Fig. 17.2 Harvester workflow

At first, job "Init" is executed to set the URL of the selected target to be harvested from CIARD RING, the path of where the records are stored in the grid, the number of records to be harvested, etc. Then job "Harvest" executes and manages the predeployed real harvest process. One of the next parallel conditional branches (see Agris and LOM2+LinkChecker) is executed depending on the metadata format required by the users; it is also responsible for performing the needed transformations. Both branches validate whether a record can be transformed to the metadata format, or not; hence two record-sets will be produced by these jobs with the correct and the wrong records separated. Then, the produced folders will be compressed and uploaded by the job "Upload" into the given LFC

folder. As the next step, the job "Register" registers the files uploaded by REST service invocation to CouchDB so that the files can be accessed from outside the grid. Finally, the last job sends a request to CouchDB to get the metadata information about the registered file-sets in JSON format.

The progress information as well as these output files are sent back (using Remote API) to the users through the Drupal module, providing complete URLs where the harvested records can be found.

17.2.3 Use of the agINFRA science gateway for workflows

For agINFRA purposes gUSE version 3.6.1 has been currently installed and maintained by SZTAKI. The gateway is configured to the gLite-based agINFRA Virtual Organization including 4 sites from Italy, Serbia, and Hungary all-together equipped with 3500 CPUs and 0.9 PetaByte data storage. As the workflow is in pre-release phase, no user statistics available, however (according to current plans) we expect only a few direct users among information managers but the harvested datasets will serve magnitudes of order more users; researchers, students, librarians, etc.

17.2.4 Further development plans

There is room for improvement at several levels. One on-going work (initiated by the agINFRA partners) is to put the demo version of the cloud-based on-demand agINFRA integrated service deployment into production. Its current implementation is deployed on the OpenNebula-based SZTAKI cloud (i) a gUSE instance with its workflow engine and Remote API to provide external interface, (ii) BIOVEL and agINFRA workflows, (iii) an extensible distributed computing infrastructure based on SZTAKI Desktop Grid [Kacsuk/2009] and 3G Bridge technologies [Kacsuk/2011b]. This approach allows the partners to create the integrated service on-demand, as well as some fine-tailored and temporary micro-portals later that enable the user to handle a subset of harvested data for further in-depth research and studies. The service also has access to external gLite resources by submitting jobs authenticated with robot certificates, or to invoke third-party REST services.

On other hand, more complex workflows with more data processing jobs are expected in the agINFRA project, which would further improve the quality of the harvested datasets.

The elaborated workflows and related solutions will be exploited and developed further in the recently launched AgroDat.hu project between 2014 and 2018. Its main aim is establishing a new Hungarian knowledge center and decision support system in agriculture-based on collected sensor data and international aggregated databases.

17.3 The VERCE Science Gateway: Enabling User-Friendly, Data-Intensive and HPC Applications across European e-Infrastructures

Seismology addresses fundamental problems in understanding earthquake dynamics, seismic wave propagation, and the properties of the Earth's subsurface at a large number of scales. These aim at aiding society in the forecasting and mitigation of natural hazards, energy resources, environmental changes, and national security. The Virtual Earthquake and seismology Research Community in Europe e-science environment (VERCE) is supporting this effort by developing a service-oriented architecture and a data-intensive platform delivering services, workflow tools, and software as a service, integrating access to the distributed European public data and computing e-infrastructures (GRID, HPC, and CLOUD) within the VERCE science gateway.

The development of a scientific gateway is a very complex task. It requires several standard components, which are fundamental to connect and exploit existing computational infrastructures, such as the European Grid Infrastructure (EGI)[1] or PRACE.[2] These components are typically related to job-submission management, user authentication, profiling, credential management, data management, and application registries. An effective science gateway needs to present a consistent integration of data, metadata, application code and methods oriented to the needs of its scientific users. It needs to be easy to use for new users and efficient for experienced users. It should encourage or enforce the governance rules and mores that its community wishes to establish. Therefore, support for standards and customization is necessary in order to produce a dedicated gateway that is tailored for a specific community. In this section we illustrate the rationale behind the adoption of WS-PGRADE/gUSE in relation to those requirements.

17.3.1 The VERCE Use-Cases

The VERCE[3] computational platform aims to provide the facilities to fulfill the requirements of two well-defined seismological applications: cross-correlation analysis and forward modeling/inversion. A short description of these is provided below.

Cross-correlation analysis: This technique, also called passive imaging, allows the detection of relative seismic-wave velocity variations that are associated with the stress-field changes that occurred in the test area during the time of recording. It uses the cross-correlation of noise signatures originating at depth and recorded at surface receivers to retrieve the Green's function between these receivers,

---

[1] www.egi.eu

[2] www.prace-ri.eu

[3] www.verce.eu/

thereby reconstructing the impulse response of the intervening plate [Bensen/2007].

Forward modeling/inversion: Forward modeling generates synthetic seismograms for various Earth models. This is achieved by the execution of high performance computing (HPC) simulation codes, which are called solvers. The synthetic data produced by the solvers may be compared with real observations for earthquakes on a continental scale, in order to foster subsequent model updates and improvement (inversion) [Fichtner/ 2008; Moczo/2006].

17.3.2 WS-PGRADE Integration Strategy

The integration of the VERCE platform within the SCI-BUS scientific gateway framework has been conducted according to the following plan:

Deployment tests: A number of deployment tests had to be performed to verify that the system was flexible enough to run on the VERCE resources. The tests were successful and so far there are two installations available, for test and production respectively.

Investigation of security issues: The authentication and authorization model adopted by VERCE is based on X.509 VOMS-enabled certificates, which reflects the EGI regulations. Indeed, the support for a flexible certificate management system offered by WS-PGRADE was a crucial feature, which immediately confirmed the value of adopting that framework. Moreover, the possibility of producing and uploading users' proxy certificates interactively is offered through the VERCE science gateway. Two applets, MyProxy Tool and a Grid Security Infrastructure - Secure Shell (GSISSH)-Term[4], maintained at LRZ[5], can be used to create and upload both a regular and a Virtual Organization Membership Service (VOMS)-enabled proxy and to access the clusters of the VERCE platform via an interactive GSISSH terminal.

Development of HPC workflows: This activity successfully tested the support from WS-PGRADE for a number of job-submission middleware tools, enabling the communication between the science gateway and the VERCE HPC computational resources. Though, the current setup is mostly based on GT5 resources.

Integration of the VERCE scientific application: Of great significance for us was the potential to design and develop our own customized applications, delegating the submission of the workflow and the control of the middleware to the WS-PGRADE framework. Effort was focused on the development of the front-end and the implementation of the workflow, instrumenting the control of the middleware via the ASM API. The support of the gUSE team was very helpful to ease our learning, including fostering their implementation of new ASM functionalities, such as being able to programmatically change a workflow's job properties.

---

[4] www.lrz.de/services/compute/grid_en/software_en/gsisshterm_en/

[5] www.lrz.de/

17.3.3 Gateway Services Implementation

In this section we provide a detailed overview on the progress achieved with the development of the VERCE gateway. It particularly concerns the forward modeling tool, which is the target of the first beta release of the gateway. The platform was used for realistic scenarios, demonstrating the feasibility of controlling large-scale computations, producing thousands of data products and related metadata across the EGI and PRACE resources hosted at LRZ.

Forward Modeling

Figure 17.3 illustrates the overall architecture of the application, indicating the components and the user-interaction flow. The GUI will guide a user towards the complete definition of their runs' input data, determining automatically the geographical boundaries of their experiments, as shown in Fig. 17.4. The functionalities of the tool can be grouped under four main categories.

Simulation code selection and configuration: The waveform simulation software can be selected from a list of available codes already deployed within the VERCE computational resources. The GUI allows selection of a solver, configuration of its input parameters and a choice from a set of meshes and velocity models.

Earthquakes and stations selection: Earthquakes and stations can be searched, visualized and selected in two different modes. The users can choose either to upload their own files, or to directly query the International Federation of Digital Seismograph Networks (FDSN) web services. All of the relevant parameters are presented to the user in both tabular and graphical form. An interactive map shows the earthquakes and station data in combination with other geographical layers of interest, such as geological information, hazard levels, seismic faults, etc. They are published by other projects such as OneGeology and European facility for earthquake hazard & risk (EFEHR).

Workflow submission and control: The portal administrators have the rights to enable the forward modeling application to use different workflows, which have been tailored for the specific DCI. For instance, heavy computations can be launched using a PRACE workflow, while short tests might run on workflows tuned for EGI resources.

We have noticed that queues can be relatively long in some clusters, with strict prioritization policies. Therefore iterating over the earthquakes in a single run could guarantee that once the computation has commenced it will eventually complete for all of the earthquakes. On the other hand, we have also included an option which allows the submission of one workflow for each earthquake. This possibility will definitely guarantee speedups in clusters with faster queues for smaller jobs. The status of each run can be monitored from the Control tab, which offers a number of sensible functionalities such as:

1. Download the output and error logs of the jobs in the workflow.
2. Reload the setup of the experiment.
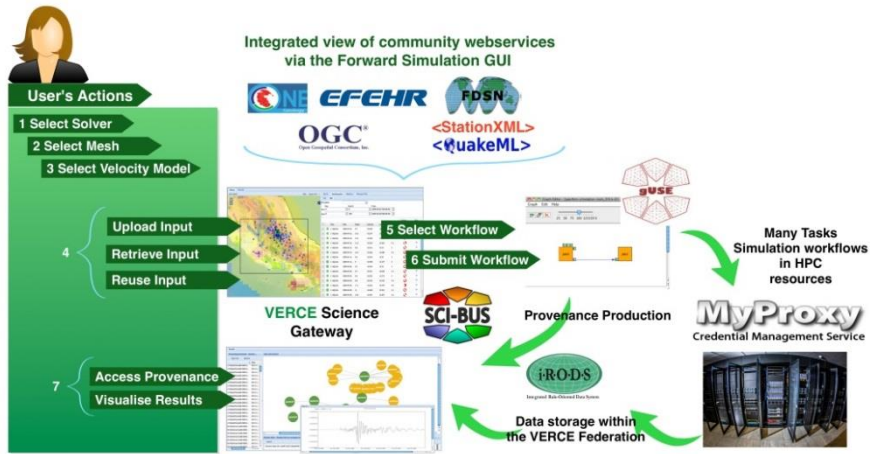3. Abort the run and delete the instance from the user's archive of runs.

Fig. 17.3 Overview of the web application allowing scalable forward-modeling analysis. The image illustrates the user interactive flow with respect to the components and the data services involved in the process
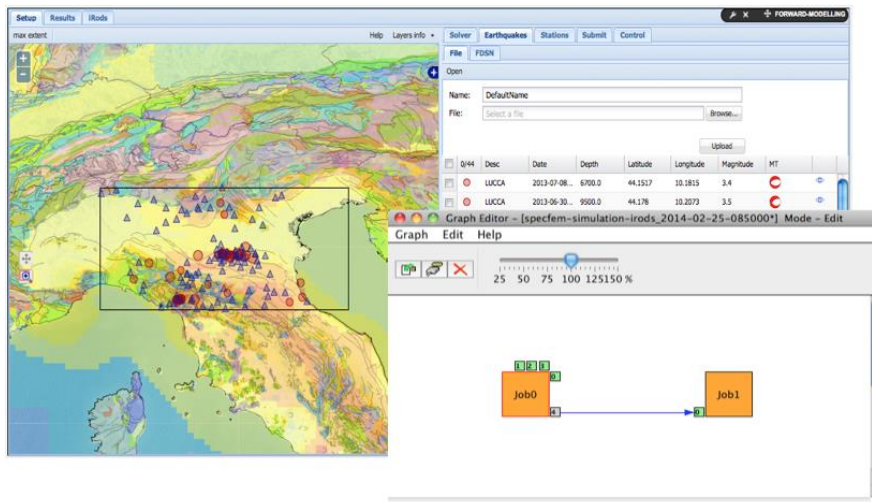

Fig. 17.4 The earthquake selection and visualization interface and the abstract graph for the simulation workflow. Input files consist of earthquake parameters, simulator configuration, station information and library of processing elements (PE). The two jobs take care respectively of, preprocessing solver execution and post-processing (*Job0*), data stage-out and cleanup (*Job1*)
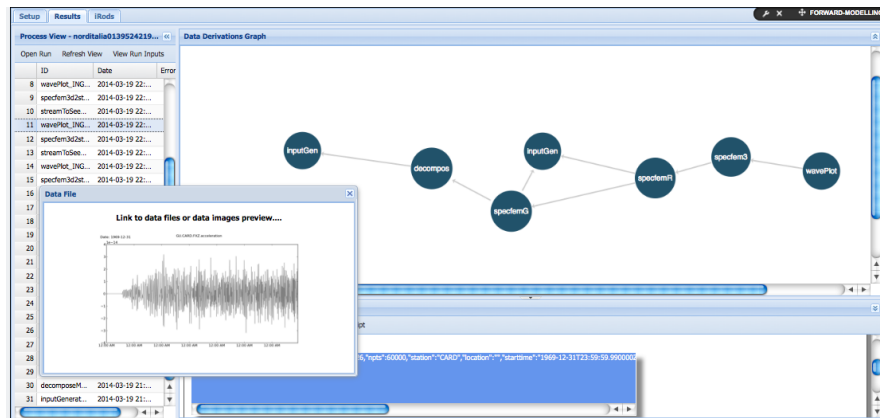
Fig. 17.5 Provenance Explorer GUI: The workflow's provenance information can be explored in a fully interactive fashion, allowing the visualization and the download of the data produced. It provides, moreover, a navigable graphical representation of the data derivation graph. From *right to left*, the dependencies from the wavePlot module to inputGen are made explicit and browsable

Multilayered Workflow Specifications and Provenance

The WS-PGRADE workflows, which have been implemented for the forward-modeling application, consist of two main jobs. One job performs the actual computation, the other takes care of controlling the staging out of the result data from the DCI to a data management system and cleaning up the computational cluster. The last task is extremely important since disk resources within the DCIs are limited and the gateway tries to automate their sustainable exploitation.

As shown in Fig. 17.4, Job0 of the workflow takes as input a number of files: earthquake parameters, station information, simulator configuration, and library of processing elements (PEs). The library of PEs contains all of the user scripts and utilities that will be used by the executable of Job0. These PEs have been developed by the scientists themselves in Python, and they can be scripted and chained into pipelines. They consist of streaming operators that operate on units of data in a stream and avoid moving streams in and out of the cluster. Moreover, the PEs are capable of extracting all of the metadata information and provenance traces related to their execution [Spinuso/2013]. For instance, the MPI applications are also launched from within a PE, in order to capture relevant information. This multilayered strategy to the workflow specification allows the extraction and storage of fine-grained lineage data at runtime. The storage system is based on a document store exposed via a provenance web API [Davidson/2008]. The coverage of the current provenance model can be considered to be conceptually compliant with the W3C-PROV[6] recommendation.

---

[6] http://www.w3.org/TR/2013/REC-prov-dm-20130430

Provenance Explorer

The provenance API is accessible via a user interface, as shown in Fig. 17.5. Users can search for a run of interest by submitting metadata queries on value ranges. The users can examine the processes and the produced data. If any file has been generated this will be linked and available for visualization or download. The interface also offers an interactive navigation across all of the data dependencies for and from a certain product. It is important to notice how the data dependency trace of Figure 5 provides more insight about the actual workflow logic, compared with the high-level WS-PGRADE specification, shown in Figure 4.

VERCE Data Management

User Document Store: The VERCE Science Gateway provides users with a way to upload and manage their own configuration files and input datasets by interfacing with the document store provided by the Liferay portal. The adoption of the Liferay version currently supported by WS-PGRADE, imposed some limitations on the enactment of some of the document store capabilities. The support for more up-to-date versions of Liferay will allow the implementation of new features (i.e., file sharing among users and groups) that will foster new collaborative interactions.

Workflow Data Products: As previously mentioned, all of the data produced by the workflows are shipped at the end of the computation to an external data store. The data store consists of a federation of iRODS instances, internal to the VERCE consortium, which supports authorization and authentication, data replication, and metadata processing services. The VERCE gateway offers to the users the interactive access to the data stored within the nodes of the federation. The GUI is based on the open-source iRODS/web software.[7]

17.4 The DRIHM Science Gateway

Predicting weather and climate and its impacts on the environment, including hazards such as floods and landslides, is still one of the main challenges of the 21st century with significant societal and economic implications. At the heart of this challenge lies the ability to have easy access to hydrometeorological data and models, and to facilitate the collaboration among meteorologists, hydrologists, and Earth science experts for accelerated scientific advances in hydrometeorological research (HMR).

The Distributed Research Infrastructure for Hydro-Meteorology (DRIHM, http://www.drihm.eu) project aims at setting the stage for a new way of doing HMR combining expertise in this field with those in Grid, Cloud, and HPC [Schiffers/2011].

---

[7] code.renci.org/gf/project/irods.php

In particular one of the main goals is the development of a science gateway [Danovaro/2014] able to drive HMR towards new solutions and improved approaches for daily work through:

- the provisioning of integrated HMR services (such as meteorological models, hydrological models, stochastic downscaling tools, decision support systems, observational data) enabled by unified access to and seamless integration of Cloud, Grid, and HPC facilities – the e-infrastructure – in such a way that it is possible to solve substantially larger, and therefore scientifically more interesting, HMR problems;
- the design, development and deployment of user-friendly interfaces aiming to abstract HMR service provision from the underlying infrastructural complexities and specific implementations;
- the support for the user-driven "composition" of virtual facilities in the form of forecasting chains, composed by "Gridified" heterogeneous models [Schiffers/2014], post-processing tools, decision support system models and data, also with the objective of promoting the modeling activities in HMR community and related disciplines.

DRIHM focuses on three suites of experiments devoted to forecast severe hydrometeorological events over complex orography areas as well as to assess their impact, and can be synthetized as:

- Experiment suite 1, incorporates numerical weather prediction (NWP) and downscaling together with stochastic downscaling algorithms to enable the production of more effective quantitative rainfall predictions for severe meteorological events;
- Experiment suite 2, takes the output from NWP executions to produce discharge data from drainage;
- Experiment suite 3, driven by the data produced by experiment suite 2, completes the model chain by adding water level, flow, and impact.

DRIHM is following a "learning by doing" approach: the hydrometeorological scientists not only access the DRIHM e-infrastructure to carry out HMR activities, but they work with ICT researchers to drive the development of all the necessary tools and practices, allowing them to effectively exploit the available resources.

17.4.1 A Brief Description of the DRIHM Science Gateway

The most important feature of the DRIHM science gateway (http://portal.drihm.eu) is the possibility to run a simulation based on the above-mentioned experiment suites using an integrated, user-friendly interface hiding all the low-level details due to the submission of jobs on a heterogeneous infrastructure. A hydrometeorological scientist, in fact, needs only to select the steps to execute, to configure each of them, and then to submit the experiment. To this extent, there are two important features of WS-PGRADE/gUSE that have been exploited in DRIHM: the DCI Bridge and the ASM.

State-of-the-art software tools (called model engines or just models) are available for each experiment suite. These models differ in terms of platform requirements and model coupling [Clematis/2012]: some models are designed for POSIX systems, and thus are easy to port onto different platforms. Others are designed for Windows or place strong requirements on libraries and other software tools a hosting system has to provide. Moreover, some models are based on the Ensemble forecasting, which is a kind of Monte Carlo analysis, and therefore they require high throughput computing, while others require HPC to provide meaningful results for large scale simulations. This is the reason why in the project there is the need to consider all the resources provided by the European e-Infrastructure ecosystem. This means that, beside the Grid resources provided to the project by the EGI, there is the need to consider other "components" that grant services that did not have a place (for performance and/or functionality reasons) in the core Grid platform. In particular, the DRIHM project considers resources from the Partnership for Advanced Computing in Europe (PRACE) [Fiori/2014], data repositories, (Linux) dedicated nodes, specialized hardware, Windows nodes, Cloud (in particular, the EGI FedCloud infrastructure), and Web services. Consequently, the necessity for a common interface layer like the DCI Bridge, which provides a standard access to all of these components, is obvious.
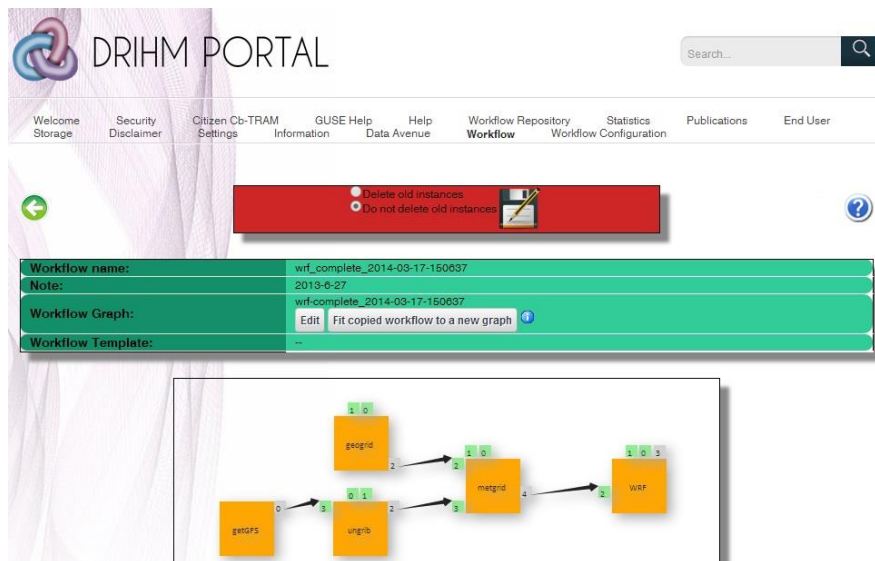
Fig. 17.6 The graph of a workflow for the execution of the WRF-ARW weather forecast model, downloaded from the SHIWA repository

As regards the ASM, this is a key feature for providing end users with high-level interfaces. For example, with WS-PGRADE it is possible to define the execution of the WRF meteorological model with the workflow presented in Fig. 17.6

that has been downloaded from the SHIWA repository [Plankensteiner/2013]. But this approach is not completely satisfactory considering the project goals for two reasons.

The first one is that with this approach the user has to deal explicitly with the selection of the computational resource to use, and she/he has to create and provide all the input files containing the parameters and, possibly, also a specific or "Gridified" version of the executable to run. The second reason is that this workflow represents only a component of the project's experiment suites, therefore it has to be combined with further components. Considering experiment suites 1 and 2, three models for the weather forecast (i.e., WRF-ARW, WRF-NMM, and Meso-NH), and three models for the hydrological forecast (DRIFT, RIBS, and HBV) were selected, resulting in the need for the user to explicitly manage nine different workflows.

On the contrary the design principle adopted for the DRIHM science gateway is to provide the scientists with a high-level interface represented by a set of portlets that allows users to define the experiment parameters in an integrated way. This means, for example, the possibility to check if the domain of the basin considered for the hydrological forecast is included in the domain selected for the weather forecast, as shown in Fig. 17.7 and that the time interval for both the simulations is coherent. This is in addition to the possibility to select, in an automatic way, the computational resources for each step of the workflow, considering the requirements of the executables, the size and the location of the data to transfer, and possible users' grants on specific resources, as, for example, those belonging to PRACE. This concept of hiding workflows by user-oriented portlets was easily implementable thanks to the ASM concept of WS-PGRADE/gUSE.

17.4.3 Further Development of the DRIHM Gateway

The DRIHM science gateway is under development at the time of writing this book. Some model-specific portlets have to be completed, and several aspects have to be properly addressed due to the complex and heterogeneous nature of the considered e-infrastructure. Regarding future development, the most important directions are related to some planned features of gUSE/WS-PGRADE. The first one is related to providing data provenance information for the results of the experiments. The possibility to specify in detail how a result was achieved is of particular importance. In a future versions gUSE/WS-PGRADE will be interfaced with the Provenance manager (PROV-man) framework in order to make these data available [Benabdelkader/2011].
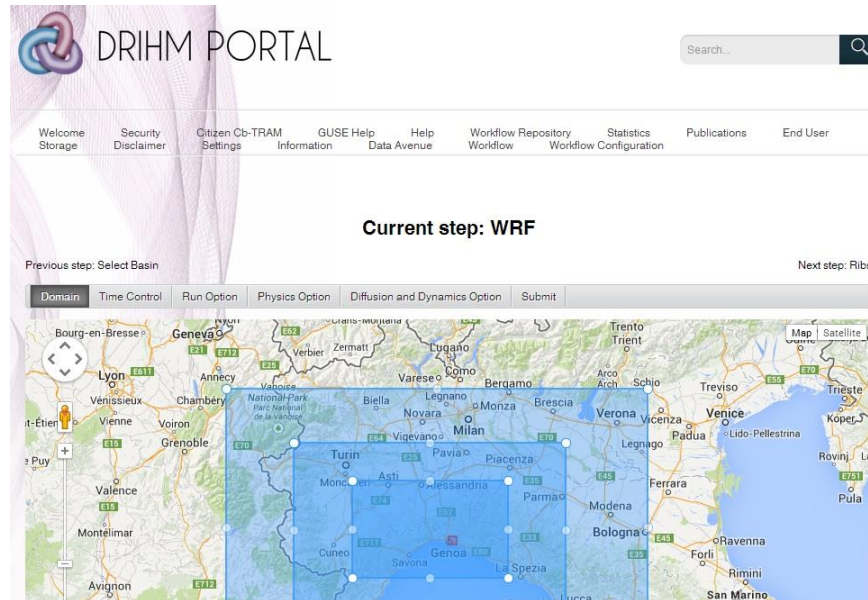
Fig. 17.7 The portlet for the definition of an experiment involving the execution of WRF followed by a hydrological forecast. The *small box* near Genoa represents the Bisagno river basin that has to be included in all the nested domains for WRF simulation

The second one is related to the activities carried out within a related project, Distributed Research Infrastructure for Hydro-Meteorology to United States of America (DRIHM2US, http://www.drihm2us.eu) is an international cooperation between Europe and the USA with the aim to promote the development of an interoperable HMR e-infrastructure across the Atlantic (storms and climate change effects do not respect country boundaries) [Parodi/2012]. In a future version, the DCI-BRIDGE will provide access also to XSEDE resources, and this will result in the possibility to run HMR simulation in Europe using data repositories on XSEDE resources, and vice versa, in a nearly straightforward way.

17.5 Conclusions

This chapter provided an overview of science gateways that were developed by projects and user communities external to the SCI-BUS project. Three representative examples of such gateways, the agINFRA, the VERCE, and the DRIHM gateways, were presented in detail.

Generic feedback from these user communities indicates that using the WS-PGRADE/gUSE framework speeded up the implementation of these gateways significantly. Although using the framework requires a learning period and several configuration and integration exercises, its generic low and higher level services make integration to various DCIs and the development of user-friendly gateway

solutions much more efficient. Gateway developers clearly appreciated the documentation and support that was available, making troubleshooting and problem-solving activities more effective. On the other hand, feedback from these communities provided valuable input for the core WS-PGRADE/gUSE developer team, resulting in important enhancements of the platform. The aim is that such activities will contribute and eventually result in building a strong extended community behind the gateway framework technology that guarantees sustainability and constant evolution.