CMST 15(1) 119-138 (2009)
DOI:10.12921/cmst.2009.15.01.119-138

Contents

RINGrid

1 EXECUTIVE SUMMARY

This document presents the final outcome of the RINGrid [RINGRID] project's work, whose principal aim was to provide recommendations concerning the usage of scientific instrumentation in a Grid environment and make it available to the e-Infrastructure. It is useful to preface this Executive Summary with the most important reasons why Remote Instrumentation on the Grid is an advance from plain, standalone Remote Instrumentation. We adopted the term Remote Instrumentation Services (RIS) to indicate instrumentation that is made accessible by exposing it in terms of Grid Services, as the Grid is a Service Oriented Infrastructure (SOI).

We start the document off by presenting advantages of integrating remote instrumentation with the Grid. This list of RIS advantages neither is nor is meant to be exhaustive. One could elaborate further on the reasons why this integration is a good idea. Among other aspects, we point out scalability, dynamic nature of workflow-based applications, resiliency through replication of data, and standardized representation of information on resources.

Next, the results of interaction with end users are presented. Results were collected in two complementary ways: by performing experiments on prototype installations, and by questionnaires specifically designed in the project and used to solicit input from various user groups regarding the experiments that they typically execute as part of their work. The interpretation of results presents real users' needs in different aspects of experiment execution. Based on this analysis, a "model use case" is proposed. It expresses the steps that a remote experiment will typically go through. The requirements that are implied or directly suggested by users, like security and policies, instrument virtualization, monitoring and accounting, workflow and execution management - but even the presence of a local collaborator - are elaborated on at this point.

Following, user requirements are translated into recommendations for RIS. The recommendations are build on the Open Grid Services Architecture (OGSA), to enhance and adapt it to the requirements of Remote Instrumentation. The proposed "conceptual design" includes not only recommendations that result from direct and obvious technical requirements of the instruments, but also those that are based on user comments and should be considered as development guidelines for interfaces. Finally, requirements, recommendations and the list of gaps that should be further investigated are presented as a coherent and self-contained table.

2 INTRODUCTION

This White Paper is intended to be a final, complete, and coherent set of requirements and recommendations for instrumentation Grid infrastructures based on results obtained in the RINGrid project.

There are numerous areas of science, industry, and commerce that require broad international cooperation for their success. A number of problems may be addressed by using sophisticated equipment and top-level expertise, which is often locally unavailable. Therefore, the development and dissemination of techniques and technologies that allow virtualized, remote and shared access to industrial or scientific instruments are essential for the progress of society. The possibility of using scientific or industrial equipment independently of their physical location helps towards equality of opportunity for and unification of communities, and subsequently provides new opportunities for industry, science, and business. Furthermore, it has a very important political and strategic impact as a step towards unified Europe.

In the RINGrid project we conducted a systematic identification of instruments and corresponding user communities, a definition of their requirements, as well as a careful analysis of the remote instrumentation synergy with next-generation high-speed communications networks and

Grid infrastructure. These were the bases for the definition of recommendations for designing next-generation Remote Instrumentation Services.

3 BENEFITS OF REMOTE INSTRUMENTATION

Remote instrumentation has a significant impact in the area of Grid-embedded experimental measurements and scientific cooperation. Instruments, sensors, and processing equipment play a fundamental role in science and engineering. Today's instrumentation is based on computational technology, and data collection interacts dynamically and in depth with phenomena analysis, modelling and interpretation. Instrumentation's role as a Grid component is stressed by the dynamic increase of international scientific collaboration, by the requirements of real-time remote access to specialized instruments, and by the deployment of large-scale sensor networks. The evaluation of the synergy of remote instrumentation with next-generation high-speed communications networks and Grid infrastructures will allow for better usage of the existing, often under-utilized research instrumentation, as well as of the networking and Grid infrastructure itself.

RIS creates new, previously unnoticed, possibilities for many research projects that would be impossible to perform because of the lack of specific and sophisticated research equipment. This is especially true for scientific communities from countries that cannot provide enough funding for their research infrastructure and for the required specialized devices. The possibility of performing research through access to Grid-embedded remote instrumentation, with the utilisation of next-generation high-speed networks, will even out their chances in the scientific world. Even in rich and developed countries, scientists are not always aware of the possibilities that remote instrumentation is likely to offer in the near future. RIS focuses on the creation of **new possibilities** for remote-instrumentation-based cooperation among scientists from different countries, working in the same or closely related scientific domains.

In most cases the owners of scientific instruments and their users are not aware that it is possible to put their resources into the Grid network and make them remotely available for research. The possibility of remote access will allow for **better utilisation** of the instruments, therefore making the resource usage more efficient. The owners will also have a chance to **cooperate** on their research with specialists from many different remote locations. The most obvious advantage of this new approach is that one can **exploit the storage and processing** capabilities that the classical data/computing Grid offers. By representing the instrument as a service and integrating it with other services through well-understood protocols, it becomes (conceptually, but also technically) straightforward to directly store experimental data to arbitrary locations world-wide, replicate them in multiple locations, and perform post-processing that might previously take days or months in only a fraction of the time. The massive Grid capabilities, surpassing any kind of supercomputer performance for specific categories of scientific problems, are a perfect match for experimental and applied science.

However, the gains stemming from this integration extend far beyond this obvious advantage.

The very sharing of resources that lies at the heart of the Grid concept is of utmost importance in the case that we are studying. Securing the sharing of instruments through industry standard methodologies (PKI, Shibboleth, etc.) offers an unprecedented way for scientists to cooperate efficiently through standardized security means, without the hassle of implementing interoperability layers for each different experiment and site combination.

Instrumentation as a service, an analogy to software as a service, allows composing atomic (in the software sense) experimental actions into measurement chains or long-standing experimental processes, irrelevant of the location of the cooperating instruments. Workflow execution mechanisms bringing together multiple different instruments, possibly geographically dispersed, can raise experimental science to new levels. The hard-coding of previous-generation scientific applications meant that either someone had to do all the experiments manually, one by one, or that an experiment involved a single instrument on a single site. Using RIS, the composition of multiple different experiments into a single application in the form of a workflow sets a new paradigm.

4 MODEL USE CASE

We performed our model use case analysis in two complementary ways: by the examination of three specific prototype installations, and by study of questionnaires collected from instrumentation-related scientists. Eventually, results from prototypes and experiments were combined with a wider selection of use cases from various scientific worlds.

4.1 PROTOTYPE INSTALLATIONS

Three prototype installations were built (based on previous project's results) as a final stage of research conducted within the framework of the RINGrid project. The main goal of this implementation was evaluating the suitability of existing solutions for the purpose of Remote Instrumentation, according to the expectations and impressions of scientists themselves. We describe the setups below in this section. The geographical locations of test bed parties is presented in Figure 1.

Figure 1. Geographical locations of users and test beds in prototype installations

The GRIDCC Device Farm experiment

The core element of the GRIDCC middleware (for more details, see http://www.gridcc.org) is the Instrument Element (IE), which offers a standard Web service interface to integrate scientific and general purpose instruments and sensors within the Grid. The second key component of GRIDCC is the Virtual Control Room (VCR), which was introduced to provide remote users with a virtual area from where they can control and monitor the instrumentation and where they can collaborate with each other, even if located in different physical sites. The third main component is constituted by the Execution Services. They consist of the Workflow Management System (WfMS) - which provides a workflow engine able to handle BPEL workflows interacting both with the new features of GRIDCC and with traditional computational and storage grid services based on EGEE's gLite – and of the Agreement Service (AS), which performs advanced resource reservations. The WfMS is optional and may be left out of a GRIDCC platform installation. The single essential GRIDCC component is the IE; the VCR is needed for collaborative purposes, and an interactive experimentation environment also requires the latter. Security services provide authentication/authorization functionalities.

The experiment conducted with users in RINGrid on the GRIDCC platform was related to the evaluation of the effects of noise and fading over a video transmission performed on a wireless channel. The setup is illustrated in Figure 2.

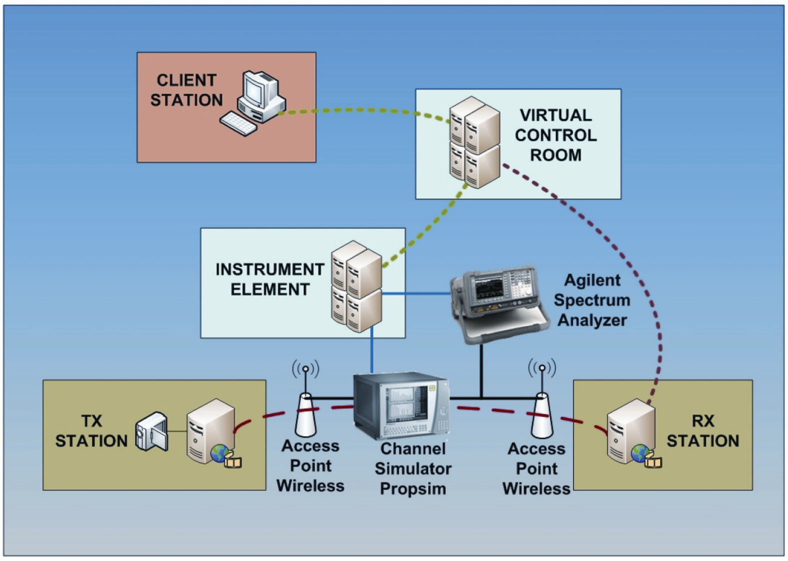


Figure 2. Device Farm experimental setup

VLab

The Virtual Laboratory project, developed in the Poznań Supercomputing and Networking Center (PSNC), assumed the creation of a distributed workgroup environment, with the main task of providing remote access to various kinds of rare and expensive scientific laboratory equipment and computational resources. The main website of the project is http://vlab.psnc.pl/, which also allows to access the inner part of the web portal (login) that plays the role of user interface with the system.

A testbed (presented in Figure 3) with an instance of the VLab system was prepared for the RINGrid project's purposes. The preparation of the test bed, its maintenance and the remote users' training sessions and assistance engaged the PSNC/VLab staff, together with employees of the Institute of Bioorganic Chemistry (IBCH) and their equipment, namely, a NMR spectrometer.

The VLab setup prepared for RINGrid tests consists of:

- A VLab system server with all its services installed in the PSNC computers;
- A VARIAN 300 MHz NMR spectrometer with its controlling workstation in IBCH and job submission and execution VLab software modules.

The installation is accessible trough the VLab web portal, which provides the user interface for submitting, executing and monitoring the tasks and also for manipulating the experimental results.

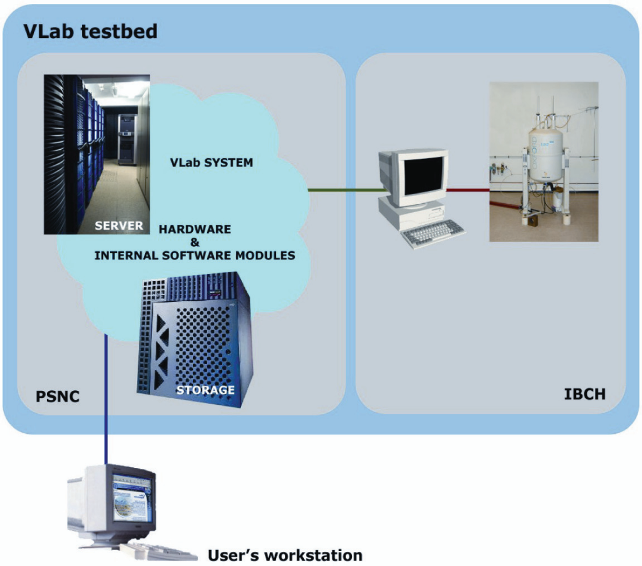


Figure 3: VLab testbed representation

UCRAV

UCRAV (*Uso Colaborativo de Recursos de Alto Valor*) is a pilot platform built with a set of tools designed for the Internet, whose purpose is to offer remote instrumentation services. The UCRAV solution uses the Grid concept for the research and development of the scientific-technological activity, allowing the remote visualization of instruments available within an environment of collaboration between researchers and users. All the UCRAV applications are built with open code standards. An outstanding feature is the use of Globus Toolkit tools for the construction of the services' grid.

UCRAV, in its current form, follows a simple approach to enable features for the collaboration of remote scientists with scientists or assistants at the site of the instruments. It includes facilities for the reservation of remote instruments, accompanied by tele-conferencing facilities, which allow collaborating scientists to have real-time discussions, also sending video feeds. The authentication for the use of he instruments is taking place using standard grid security models (X.509 certificates). As such, UCRAV can be considered as a barebones platform providing the absolutely necessary features for using an instrument remotely, without remote control, but with the help from a local assistant.

The test bed implemented for the RINGrid project utilized a Bruker AC-250P spectrometer, with a GNU/Linux server attached to it and the complete UCRAV software installed.

4.2 GENERAL USE CASE

In order to devise a general model use-case for remote instrumentation, we prepared a questionnaire, which solicited descriptions of experiments from the scientists who perform them. The questionnaire was addressed to many different communities, including the people who participated in the preliminary phase of identification of instruments and user communities and definition of requirements, the participants of the Open Grid Forum's "Remote Instrumentation Services in Grid Environment" Research Group [RISGE], the external collaborators of the consortium. The distribution of scientific disciplines the questionnaires come from is shown in Figure 4.

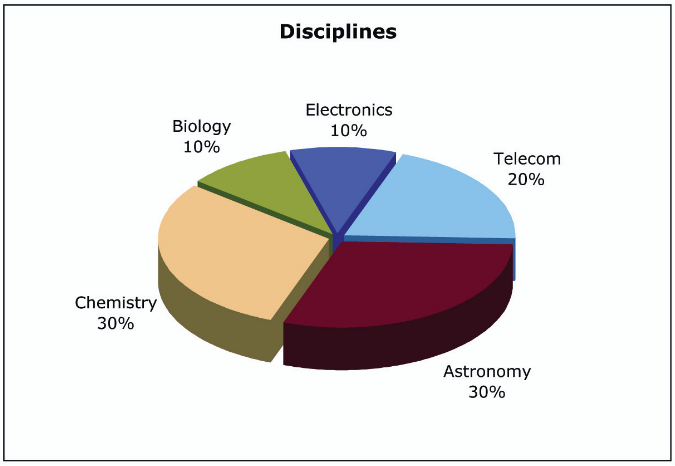


Figure 4: Distribution of disciplines for received use cases

We performed an analysis of the received responses and identified patterns of experiments. Then, based on these patterns, we defined a model use case, which is equivalent to a typical workflow for remote experiment execution.

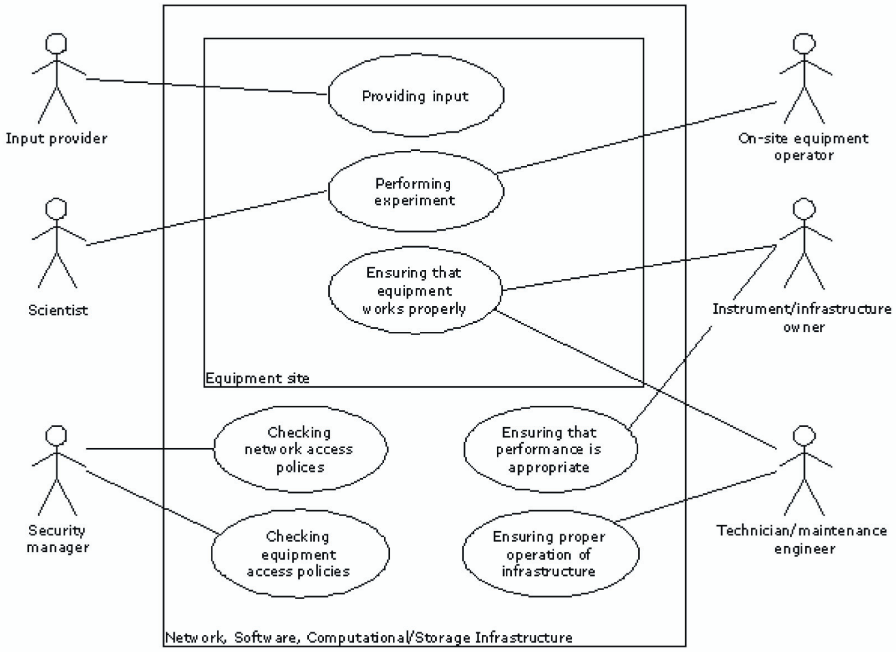


Figure 5: Use Case diagram of model workflow

In designing the general model use case, we had two options as regards the approach to follow: Coming up with the “least common denominator” of the responses we received, i.e., to include only the compulsory steps which are necessary and at the very core of the experiment, or to include also the optional steps. The optional steps may be completely irrelevant to a specific use case, or may be part of the experiment workflow in one instance but not in another, or they may even take place in one cycle of an exeriment but not in others (within the same experiment instance). Therefore, we ended up with a superset of actions, relatively to the ones of the experiments described in the responses we received. This superset defines a workflow with a possible loop and plenty of optional nodes: 1. Schedule the experiment; 2. Enable the experiment (preparatory actions); 3. Define access policies; 4. Provide the input; 5. Calibrate the instrument; 6. Execute the experiment; 7. Evaluate experimental output; 8. If needed, loop to step 3; 9. Clean-up activities.

UML and flowchart representation. The model use case (workflow) is schematically presented below. We are using Unified Modelling Language (UML) use case diagram notation in Figure 5.

Flowchart presented below shows operations (actions) which must be made to conduct typical experiment in Remote Instrumentation system.

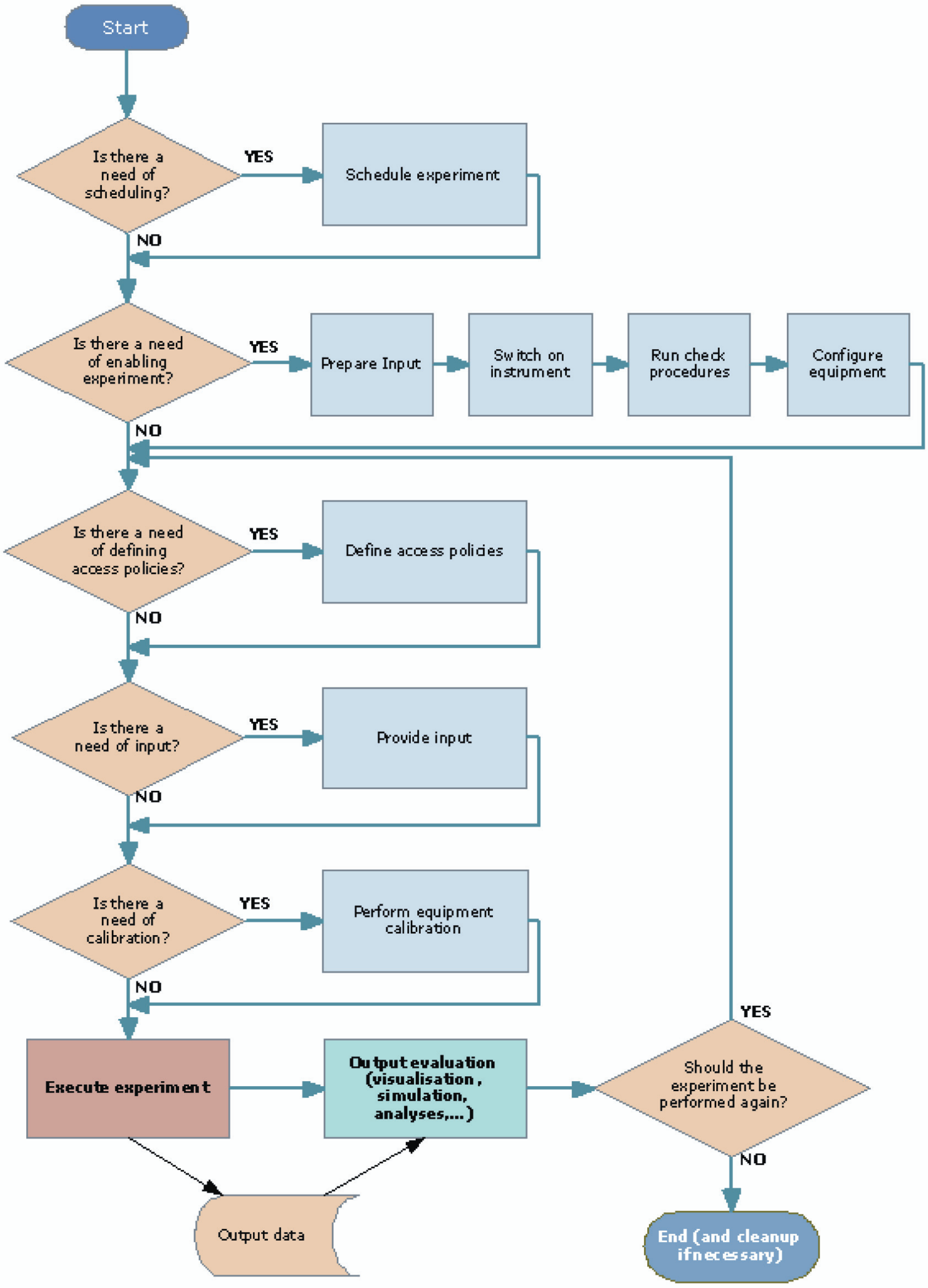


Figure 6. Flowchart of model use case

4.3 REQUIREMENTS STEMMING FROM THE USE CASE ANALYSIS

We recognize three functional requirements in the basic OGSA Grid service capabilities (also termed *Abstract Instrumentation Domain* in the reference architecture of conceptual design, reported in Figure 3, which will be discussed in Section 5) as **common to all types of remote experiments**: 1. Instrument virtualization; 2. Error monitoring and logging; 3. Accounting and monitoring/controlling experiment execution.

They belong to the classes of Execution Management and Resource Management Services. Among these, *Instrument Virtualization* or, in other words, the definition of appropriate *Service Providers* for Remote Instrumentation is of utmost importance and characterizes Remote Instrumentation Services with respect to other service types. This is an area where standardization work is certainly needed, and where a compromise between the generality of the interfaces and their specialization to instrumentation in different fields should most likely be sought, in the light of both end-users' and application developers' needs.

All these can be variously classified into the categories of Execution, Resource, Data and Information Management. Though all these functionalities exist in current Grid middleware, there is ample space for specialization to the Remote Instrumentation needs. In particular, we recall here the importance of QoS-aware workflows, producer/consumer models for efficient data transfer, cross-domain QoS-mapping, and exchange of information for control purposes.

Finally, there are some non-functional requirements that are of importance to Remote Instrumentation, as stemming from our use case description:

- Input management;
- The presence of (and interaction with) local operator(s);
- The presence of specialized visualization devices (and the corresponding software for data representation – which pertains essentially to the local host, but may have the need to interact with data services).

In particular, the possible need of interaction with a local operator is peculiar to a number of instances of Remote Instrumentation Services (e.g., in material science), and calls for the increased integration between general collaborative tools and user interfaces for Remote Instrumentation (as, e.g., in the VCR concept elaborated in previous and current projects like GRIDCC and DORII).

In essence, our testing with user groups on specific platforms – though necessarily limited by the resources available to the RINGrid project – has confirmed some of our previous conclusions, regarding aspects that pertain to all our three domains. In almost all cases, initial impairments caused by the interaction with local security policies had to be faced and solved before running the experiment. Thus, it is quite relevant for the real diffusion of Remote Instrumentation Services that Security Services be available to a large extent, in standard form, and easily implementable within the platform in use.

The most interesting conclusions, however, come from user observations that touch the sphere of **Experiment Execution** (or, in other words, of complex services to be built "on top" of basic abstractions). It is apparent that the existing platforms we have tried all have some limitations in this respect. There is a need for:

- *Flexibility and adaptation to the user-specific experimental environment*. This point pertains to the capability of customizing a Remote Instrumentation platform involving a certain type of devices to perform a specific experiment.
- *Flexibility and adaptation of the user interface to the user's skills and level of expertise*. This is another requirement that stemmed from user comments in our short experimentation phase. It would be highly desirable to have a Graphical User Interface (GUI) capable of adapting to the level of user experience or to the needs of a particular experiment.
- *Appropriate tools for self-training and learning-by-doing*. The usage of remote instrumentation to a large extent brings a novelty in teaching experimental sciences through distance learning sessions. A topic to be investigated regards the development of tools, by

means of which users can perform supervised or even unsupervised training on the instruments, and gradually be acquainted with their usage.

All these capabilities should be embedded in business-value Remote Instrumentation Services. As a matter of fact, they are just a small instance of flexible, user-friendly, workflow-oriented orchestration services.

5 CONCEPTUAL DESIGN

5.1 ARCHITECTURAL FRAMEWORK

The semi-layered, three-tier representation of OGSA was taken as a reference. This is reported in Figure 7, along with the identification of the relevance to our context, which we have added on both sides. With respect to RIS, we refer to the three tiers as ***Physical Instrumentation Domain***, ***Abstract Instrumentation Domain***, and ***Experiment Execution Domain***, respectively. On the right side of the figure, we have indicated the scope of the domains, plus the presence of a layer whose task is that of presenting a uniform service to those that need to build the instruments' abstractions (Meta-Instruments Convergence Layer). Starting from the top, we identify the necessity of high-level, value-added services for the end user in order to allow building up and describing complex experiments that possibly involve multiple instruments. We may note that this domain is outside the scope of OGSA. The capabilities offered should include finding and choosing the components that are necessary to perform the experiment, as well as describing its high-level workflow. The ground for this should be provided by tier 2 whose task is to build basic OGSA-compliant abstractions of the instrumentation, alongside with their service capabilities. Finally, the Physical Instrumentation Domain includes all physical instruments and their proprietary software, together with all computational, storage and networking elements that are necessary for their mutual interconnection, data transfer and processing. Though the networks are seen by OGSA simply as resources, disregarding their internal structure, we will also explicitly consider instances of their protocol architecture in order to identify and highlight the aspects that are more relevant to Remote Instrumentation Services.

5.1.1 Physical Instrumentation Domain

This sub-section deals with the physical infrastructure, and particularly with instrumental (including sensor networks), telecommunication, storage, computing, and visualisation resources.

Instrumental resources are not typical ones in a strict Grid sense, and thus most use cases we may find in the recent literature and today's life refer to computational resources, such as memory, disks, processors, etc. What seems to be missing in the OGSA specification is the capability to enable interoperability among such diverse, heterogeneous, and distributed resources and services as scientific instruments are. Instrumental resources comprise both human factors and interactivity in job control, steering and management. Human factor issues that are missing in [GFD29] use cases' documentation need to be introduced in the descriptions of requirements and capabilities.

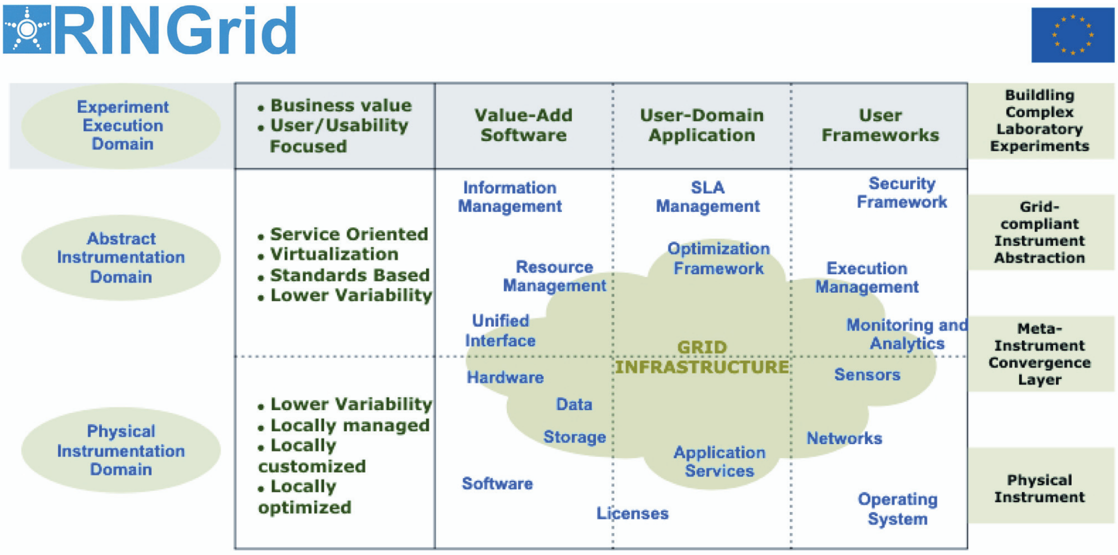


Figure 7. Three tier reference model and relation to the RIS environment

Storage resources may be a mixture of disk and tape from directly attached storage (DAS) that is used in the RIS environment. In RIS applications the major usage is for bulk data storage. The different type of data in one or more media can be stored and managed by the storage resource manager (SRM) and accessed via a uniform interface. The most important functions of the storage resource are reading and writing data, providing data security, and sharing bulk data via GridFTP. The bulk data storage requirements concern the provision of sufficient storage resource space to store data or files. The space is managed through the management interface provided by the storage service. Some instruments may need Multi-Terabytes for storage. Some typical applications involve permanent storage of experimental results from the instrumentation and digital libraries. Moreover, a sufficiently large storage space is required for storage resources and data to be transferred.

Computing resources. Efficient access to and movement of huge quantities of data is required in more and more fields of science and technology. An excellent example that came to the scene some time ago is eVLBI (electronic Very Large Baseline Interferometry), where networks of radio telescopes are used to produce detailed radio images of stars and galaxies. Not going much into details, telescopes and networks performing at data rates well beyond 1 Gbps are expected after 2010. Plans call for a 30 Gbps telescope capability before the end of the decade. The processing of multiple 30 Gbps data streams will clearly require an upgrade of the existing data formatting and transport mechanisms, as well as greatly increased data processor capacity. The above example shows how technological requirements push forward and the resources we have now at disposal may not be efficient in the nearest future. Thus, huge data transfers, their optimization, caching and replica services need to be guaranteed. Data gathering from instruments has to be reliable, and QoS parameters have to be ensured, respectively. Information is often very precious and losing it may mean repeating the whole experiment, which can take several weeks, or even months.

Data visualisation resources. Data of experimental results gathered from instruments as well as computed data derived from already existing datasets need to be visualized. This may be a complicated task since, depending on the experiment, this data may need to be shown as a video stream, possibly to be modified interactively. Even more problems arise when tactile interfaces (to be used for disabled scientists) come into play [D4.3]. It is very hard to predict – even for the near future – what changes need to happen to visualization services. The reason for this lies in the fact that there is no need for a different approach to visualization if there is no change in the type of data to visualize. However, for most data types there already exists a valid and satisfying approach to visualization. The only problems we currently see are those with interactivity and with the visualization of widgets (control elements for graphical user interfaces).

5.1.2 Abstract Instrumentation Domain

We now leave the physical instrumentation and networking domain, and take into consideration the major capabilities that are in place or should be developed (or simply adjusted) to address the specific needs of RIS.

Execution Management Services. A number of points in the part of the OGSA architectural document related to Execution Management Services (EMS) seem to be particularly relevant to RIS. EMS services deal with Resources, Job Management, and Resource selection services. An important resource is a Service Container which encapsulates running entities and exposes their resource properties and methods. A Job Manager, on the other hand, implements a manageability interface (for example, a WSDM collection which can expose some of the methods exposed by its members as its own), and it is responsible for orchestrating the services used to start a job or a set of jobs. We believe that the EMS architectural abstractions are connected to the extent instrument virtualization is carried out, and that careful consideration should be taken in trading off users' and developers' needs (e.g., in balancing generality of interfaces and ease of implementation of new Instrument Elements, one of the key abstractions for instrument virtualization).

Data services are intended to offer all the functionalities needed to interact with data produced and stored in the Grid system: they manage the access, update and transfer operations that have to be performed on physical resources, therefore acting as an interface to them. The typologies of resources that Data Services should be able to manage are numerous, and include, among others, repositories of files in several different formats, DataBase Management Systems (DBMS), Streams, File Systems, temporary aggregated data views: therefore, several different capabilities can be requested to the Data Services, spacing from the basic data transfer support to the configuration of the managed resources or to enforcing access policies and security rules.

Resource Management Services. According to the OGSA specification, some of the key issues related with resource management on the Grid need to be discussed. They are (advance) resource reservation, on-line access to the instruments, monitoring, accounting and optimization. Resource reservation is a technique ensuring the user that the resource he/she requires will be available at a certain instant and for a certain amount of time. Monitoring deals with controlling states and current instruments' status, etc., and helps other services responsible for system activity. Accounting allows the instrument owners or VO (Virtual Organization) executives to bill the regular user for instrument usage. Resource management can be provided on several levels.

OGSA **Security Services** aim to facilitate the enforcement of the security-related policy within a VO. The purpose of security policy enforcement is to ensure that the higher-level application objectives can be met. A distributed, interactive, and multi-user system, such as a scientific instrument, a virtual laboratory or a computing resource, is subject to the same threats as a resource-sharing environment built over a public network infrastructure, like the Internet. One of the critical differences between RIS security and host or site security is *site autonomy*. To meet the objectives of RIS deployed within the Grid, as well as the security policy of the VO, enforcement of the following aspects, as also specified in OGSA, might be required: message integrity and confidentiality, authentication of interacting entities, minimum authentication strength, secure logging and audit, separation of responsibilities, intrusion and extrusion detection, authorization policy checks, least privilege operations, mandatory access control mechanisms, discretionary access control mechanisms, avoidance of DoS attacks, redundancy, and training.

Self-Management Services. Self-management may be conceived as a way of reducing the cost and complexity of operating RIS and Virtual Labs. In a self-managing distributed environment, RIS components, such as scientific instrumentation, storage devices, software modules (e.g., operating systems, middleware, applications), network services and devices, should be self-configuring, self-healing and self-optimizing. These self-managing attributes suggest that all the operations involved in configuring, healing and optimizing a RIS can be automatically initiated and performed basing on situations, events or RIS requirements that the RIS components are able to detect. It is relevant to highlight that these features enable RIS providers to operate efficiently with fewer human resources, while decreasing costs and increasing the organization's ability to satisfy users' requests.

The main goal of **information services** is to enable efficient access and manipulation of information about applications, resources, services and remote instrumentation in the Grid environment. Concerning the term *information*, we refer to data or events generated by any component of the Grid system which are used for monitoring, logging, discovering, message delivering purposes, by any other actor of the Grid network. The information service does not collect physical data and events generated in the Grid system; in other words, it is not a centralized storage of data, but it rather contains the schema and description of such data and events using metadata.

5.1.3 Experiment Execution Domain

This is the domain of user applications and user frameworks for RIS, where users can build extended experimental platforms and Virtual Laboratories by utilizing the basic service capabilities provided by the other domains. The key point here is to mask the complexity of a whole physical and, to some extent, abstract infrastructure lying underneath the presentation layer, and provide a friendly interface to the regular user. We believe that the mechanisms to be provided in this Domain should allow:

- The creation of permanent or semi-permanent Virtual Organizations that collectively offer access to existing physical remote instrumentation federated within the VO, which announces itself as a Virtual Laboratory, or a set of Virtual Laboratories;
- Easy user access to high-level *pre-configured* Experiment Execution Services provided by such Virtual Laboratories;
- The capability on the users' part to build their own experiment, if not present among the pre-configured ones, by describing it in a high-level workflow description language, which would cause the system to discover the needed components, present alternatives, allow users' choices and adjustments, and finally create instances of the underlying Execution Services, as if configuring an ad hoc Virtual Laboratory "on-the-fly".

In all these cases, suitable enhanced services that allow for such high-level descriptions need to be created, by composition of the basic abstraction services provided by the OGSA-specific domains. The workflow description is one of the main aspects here. The workflows to be defined in this domain should provide the functional flow of the data obtained from the instruments on which further calculation and post-processing can be performed. Some existing WFM applications, e.g., the one stemming from the VLab (http://www.vlab.psnc.pl) experience, meet some of the OGSA requirements presented in [GFD29] and [GFD80].

5.1.4 Cross-Domain Issues

There are some aspects that involve the interaction between the different domains identified above, especially as regards the exchange of information for control and performance optimization purposes.

Service quality may become of paramount importance in many RIS activities. RIS users require reliable environments on the Grid, so that experiments can be carried out as if instruments in remote locations were locally available. Furthermore, advances in RIS research have recently resulted in commercial interest in utilising Grid infrastructures to support commercial applications and services of Remote Instrumentation. However, significant developments in the areas of QoS deployment are necessary before widespread adoptions can become a reality. Specifically, mechanisms such as requirement identification, cross-domain requirement translation, performance monitoring and performance auditing need to be incorporated into the RIS infrastructures in order to move beyond the best-effort approach to service provisioning that current Grid infrastructures follow.

User interaction services exist in the Experiment Execution Domain. These services should enable users with exactly the same ability in performing scientific research that they would have when working in the same room with the scientific equipment. In order to make those remote instruments accessible as if they were locally available, an OGSA-compliant architecture is used, aiming at managing resources across distributed heterogeneous platforms.

Cross-domain issues are particularly worth mentioning because of the service quality demands of RIS applications. The general OGSA-compliant architecture includes QoS services. However, in order to provide QoS (or QoE[1]) to users, a cross-domain performance optimisation and monitoring component may be used. This component identifies the requirements from the Experiment Execution Domain, and translates service quality requirements to the Abstract Instrumentation Domain and the Physical Instrumentation Domain.

- **Identification of Performance Requirements of RIS Services**. A RIS should meet the expectations of users in terms of performance quality. Performance requirements can be expressed either explicitly or implicitly. Explicit expression of performance requirements means that the RIS platform should have a mechanism to allow users to express their expectations of service quality requirements. Implicit requirement identification means the service quality requirements follow default definitions or general acceptance of users. Once the service requirements are identified, they are passed to the performance optimisation and monitoring module, which interacts with the Abstract Instrumentation and Physical Instrumentation Domains to ensure that the specified performance can be met. Work in this direction is ongoing - no satisfactory or standard solution is available.
- **Performance Requirements Translation**. Accessing and using instruments in remote places may bring enormous benefits. However, some problems start to emerge, e.g. performance, security, economy, dependability, etc., when the scale of RIS experiments increases. Compared to conventional instrumentation environments, the performance quality of a RIS Service relates to two issues: instrument performance and platform performance. When users want to use an instrument on the RIS platform to support their experiments, they express the experiment requirements to the RIS platform. The instrument performance can be characterised by two aspects of parameters, intrinsically and extrinsically. Intrinsic characteristics, mainly limited by physical principles and manufacturers, are inherent in the instrument.
- **Performance Monitoring and Auditing**. Auditing is an important step towards commercialisation of RIS platforms. This function allows the platform to find out which part is liable when an error is identified by a final user. A monitoring system for the RIS platform has several components. It contains “sensoring”, monitoring information publication, and data analysis.
  - *Sensoring* (also termed *instrumentation*) is a process to insert probes into physical resources to measure the state of a physically available component, e.g., hard disk, telescope, or network switch.
  - Monitoring data publication: data to be monitored is huge in many cases. It is not likely to be forwarded to every user in its original form. Abstractions, standard schemes, publication mechanisms, and access policies are required.
  - The data analysis process is to reduce the size of raw monitoring data and provide clues for auditing. These can include statistical clustering techniques, data mining, or critical path analysis.

---

[1] Quality-of-experience (QoE) - in general it describes the subjective measure of a customer's experience with a service.

6 GAP ANALYSIS AND GUIDELINES FOR FURTHER DEVELOPMENT

6.1 LIST OF RECOMMENDATIONS FOR RIS

As was expected, use case formalization and data collection from users' experience do not contradict the observations and the conclusions that could be drawn on a purely technological basis [D6.2]. **Table 1** presents requirements for RIS as results of our study. The following sections elaborate on these requirements, to end with an identification of gaps and suggestions for future directions.

Column "Technological analysis" consists of considerations about the current status and the foreseeable evolution of RI-related technology in middleware and networking. Column "Experimentation" consists of information concerning technology used by three prototype implementations. Column "Use case analysis" stemming from both general considerations and the responses from direct user experience, which we collected and elaborated above. Column "Gap analysis" consists information about technology which needs further investigation to fulfil all requirements made by Remote Instrumentation systems. The detailed description can be found below in this chapter.

Domain	Requirement	Techno-logical analysis	Experimentation			Use case analysis	Gap analysis
			Device Farm (GRIDCC)	NMR Spectroscopy (VLAB)	NMR Spectroscopy (UCRAV)		
Physical Instrumentation	High throughput	X	X			X	
	QoS guarantees (especially on access networks)	X	X				X
	Performance monitoring and control	X	X			X	
	Virtual Private Networks	X			X		X
	Ability to reserve network resources / Bandwidth on Demand	X				X	
	High availability	X		X		X	
	Multicast	X					
	Use of IPv6	X					X
	Network transparency of applications (e.g., firewall independence)		X	X	X		
	Control of security parameters (policy decision and enforcement)	X	X	X	X	X	X
Abstract Instrumentation	Virtualization of instrument resources	X	X			X	X
	(Fine-grained, user-controllable) Scheduling of instrument resources	X	X	X	X	X	X
	Sharing of instrument resources	X		X		X	
	Service-related QoS provisioning	X					X
	Establishment and monitoring of Service Level Agreements	X				X	X
	Checkpointing and recovery mechanisms	X				X	
	Widespread use of publish/subscribe mechanisms	X			X		X
	Visualization of complex data types	X				X	

	discovery						
	Instrument resource query	X	X		X	X	
	Input management (data)				X	X	
	Advanced accounting and monitoring facilities			X		X	
Experiment Execution	Interactivity	X	X	X	X	X	X
	Variety of visualization widgets	X	X			X	
	High resolution visualization	X				X	
	Standardized interfaces for remote instruments and experiments	X	X			X	X
	Self-documenting, low learning curve interfaces	X	X			X	X
	Node monitoring and failure detection	X					
	Localization of interfaces		X	X			
	Easy workflow definition	X		X		X	
	Intuitive workflow execution monitoring	X		X		X	
	Experiments based on templates, repository of experiments		X				
	Generic visual interfaces for instruments		X				X
	Platform-independence of client-side interfaces		X	X	X		
	Workflow management: Dynamic workflows with task re-routing and QoS provisioning	X		X		X	X
	Modular, flexible, standardized systems which adapt to the user and the application		X	X		X	X
	Licensing issues on driver and post-processing software	X	X			X	X
	Presence of on-site operator		X	X	X	X	
	Input management (materials)			X		X	

Table 1: Consolidated view of final recommendations

The recently established Remote Instrumentation in the Grid Environment (RISGE) working group within OGF has undertaken the task of exploring “issues related to the exploitation of Grid technologies for conducting and monitoring measurement tasks and experiments on complex remote scientific equipment”. The following is a final set of recommendations that attempt to summarize our investigation in this project and to provide input to the RISGE activity. We do so in the light of the OGSA reference model, as adapted to the Remote Instrumentation environment.

6.2 ADVANCED NETWORKING SERVICES

The community should investigate the impact and the real needs in terms of advanced networking services. Among these: Mechanisms for QoS provisioning, VPNs, BoD and deployment of GMPLS functionalities for the dynamic creation of circuit-switched high-speed virtual connections on top of IP networks, Traffic Engineering and cross-domain interaction.

Networking services should provide flexible connectivity support in a heterogeneous environment of instrumentation and access network technologies, in order to fulfill the requirements of QoS-aware workflows in the abstract instrumentation domain. To this aim, a tighter interaction is required between the Grid and the network management and control functionalities, to translate QoS requirements across multiple domains and ensure rapid reconfiguration and adaptation.

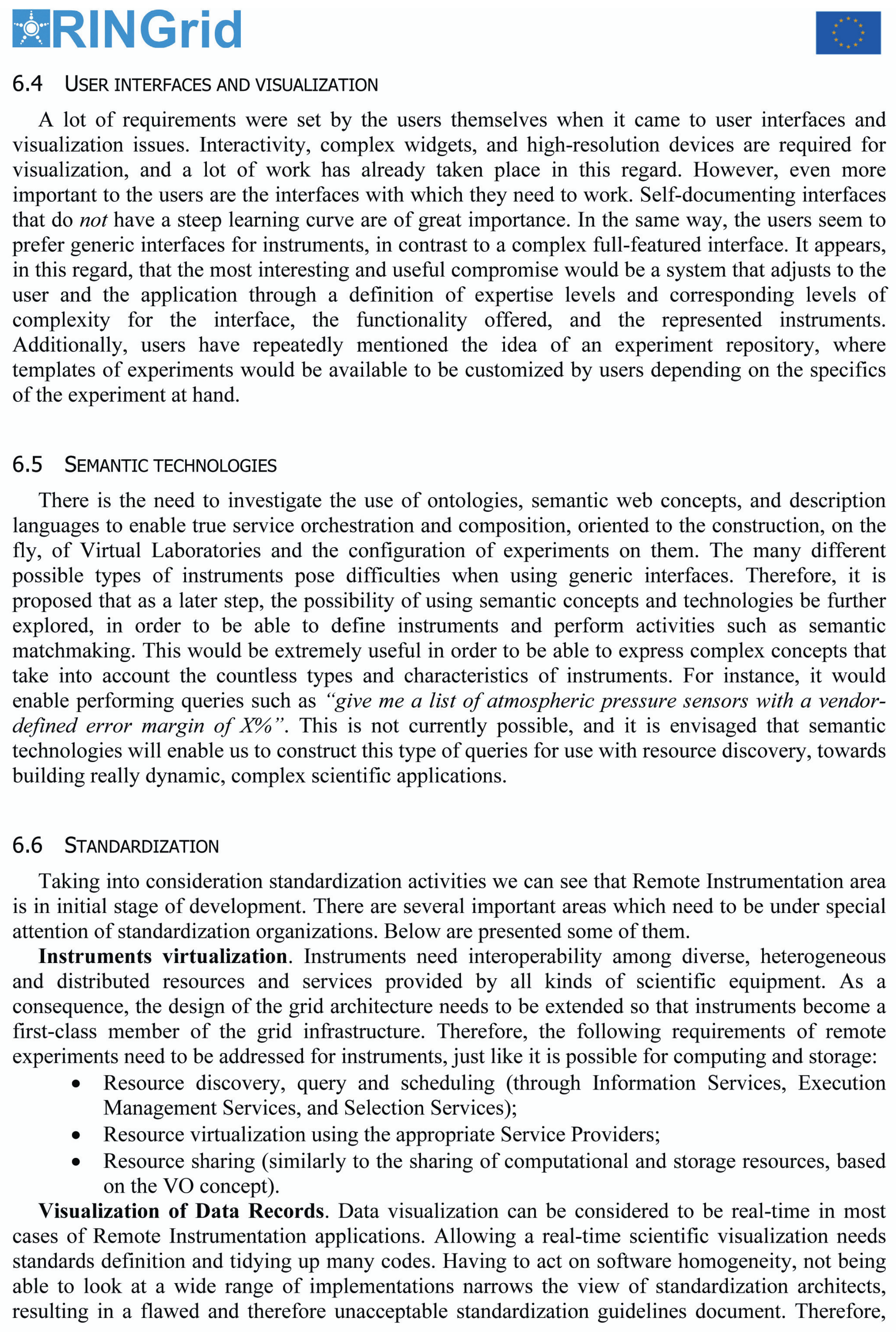
The impact and benefits of adopting **IPv6** should be investigated in terms of transparency, enhanced flow handling, and mobility support. We believe that networking services in support of RIS can constitute a valid workbench for the introduction of widespread IPv6 connectivity, also in consideration of numerous devices that should be made accessible over the network. These include both measurement instrumentation of different kinds and diffusion (i.e., not only large and expensive pieces of equipment, but also smaller and more common devices, which may be required to become part of a measurement chain), and sensor networks employed as large-scale data acquisition equipment. The latter may be exposed as Grid resources through data sink devices, but even individual sensors may need to be addressable and configurable.

The transport capabilities of **access networks**, and the possible limitations connected with wireless access, in both cases of user access and distributed data acquisition, should be assessed. In particular, maintaining QoS requirements over access networks is an important issue that is not yet completely addressed, though various different technologies (e.g., IEEE 802.11e, IEEE 802.16, ETSI DVB-S2/DVB-RCS, among others) offer possible mechanisms. Another important aspect regards the scalability of applications in relation to user access and equipment characteristics, an issue that relates to signal processing and networking topics. Finally, the role of satellites – sometimes the sole means of accessing remote or secluded areas – should be investigated in relation to RIS requirements, both with respect to terrestrial access and to controlling instrumentation in space laboratories. Moreover, in heterogeneous networking environments, comprising satellite, terrestrial wireless access, and high-speed backbone transport, RIS for network management – where the network is the instrument – should be highly recommended, e.g., for the purpose of testing new service capabilities.

6.3 GRID-SCOPE ENHANCEMENTS

OGSA services must be enhanced or complemented in the Abstract Instrumentation Domain, especially with regard to:

- *EMS Services*: Definition and implementation of appropriate Service Providers for instrument virtualization, taking into account the tradeoff between generality and ease of code development, understanding, and maintenance. Through the definition of a standardized interface for accessing (control, monitor, query) instruments, they will become first-level citizens of the Grid.
- *Data management*: Investigating relations between Digital Libraries and Remote Instrumentation experiments' data collection and storage. It is envisaged that such integration will allow for efficient indexing and re-composition of huge volumes of scientific data produced by instruments, thus facilitating their processing on the Grid.
- *Information Services*: Management of lists of instruments and their properties. This is the essential step towards integrating instruments on the Grid. Information models such as the GLUE schema must include instruments just like they include computing and storage resources.
- *Resource Management*: Mechanisms for enhanced (cross-domain) service discovery: Semantic matchmaking makes much sense with instruments, and it is a direction that the community should investigate. Additionally, service-level QoS must be further looked into in order to boost/enhance understanding of the QoS levels that a service can offer before trying to contact it for direct (best-effort) use or the establishment of an SLA.



6.4 USER INTERFACES AND VISUALIZATION

A lot of requirements were set by the users themselves when it came to user interfaces and visualization issues. Interactivity, complex widgets, and high-resolution devices are required for visualization, and a lot of work has already taken place in this regard. However, even more important to the users are the interfaces with which they need to work. Self-documenting interfaces that do *not* have a steep learning curve are of great importance. In the same way, the users seem to prefer generic interfaces for instruments, in contrast to a complex full-featured interface. It appears, in this regard, that the most interesting and useful compromise would be a system that adjusts to the user and the application through a definition of expertise levels and corresponding levels of complexity for the interface, the functionality offered, and the represented instruments. Additionally, users have repeatedly mentioned the idea of an experiment repository, where templates of experiments would be available to be customized by users depending on the specifics of the experiment at hand.

6.5 SEMANTIC TECHNOLOGIES

There is the need to investigate the use of ontologies, semantic web concepts, and description languages to enable true service orchestration and composition, oriented to the construction, on the fly, of Virtual Laboratories and the configuration of experiments on them. The many different possible types of instruments pose difficulties when using generic interfaces. Therefore, it is proposed that as a later step, the possibility of using semantic concepts and technologies be further explored, in order to be able to define instruments and perform activities such as semantic matchmaking. This would be extremely useful in order to be able to express complex concepts that take into account the countless types and characteristics of instruments. For instance, it would enable performing queries such as *"give me a list of atmospheric pressure sensors with a vendor-defined error margin of X%"*. This is not currently possible, and it is envisaged that semantic technologies will enable us to construct this type of queries for use with resource discovery, towards building really dynamic, complex scientific applications.

6.6 STANDARDIZATION

Taking into consideration standardization activities we can see that Remote Instrumentation area is in initial stage of development. There are several important areas which need to be under special attention of standardization organizations. Below are presented some of them.

Instruments virtualization. Instruments need interoperability among diverse, heterogeneous and distributed resources and services provided by all kinds of scientific equipment. As a consequence, the design of the grid architecture needs to be extended so that instruments become a first-class member of the grid infrastructure. Therefore, the following requirements of remote experiments need to be addressed for instruments, just like it is possible for computing and storage:

- Resource discovery, query and scheduling (through Information Services, Execution Management Services, and Selection Services);
- Resource virtualization using the appropriate Service Providers;
- Resource sharing (similarly to the sharing of computational and storage resources, based on the VO concept).

Visualization of Data Records. Data visualization can be considered to be real-time in most cases of Remote Instrumentation applications. Allowing a real-time scientific visualization needs standards definition and tidying up many codes. Having to act on software homogeneity, not being able to look at a wide range of implementations narrows the view of standardization architects, resulting in a flawed and therefore unacceptable standardization guidelines document. Therefore, the following steps are required to successfully drive visualization in high-profile scientific environments:

- create alternative programming interfaces similar to glogin/GVid in non-GT middleware
- show by several examples/use cases that this implementation is en par/surpasses currently existing implementations
- based on this now-existing diversity of implementations, standardize the architecture of grid data visualization, taking into account currently existing or in-draft specifications
- refine the APIs from the first point to adhere to the specification
- show that the use cases also produce the desired results with the now standardized architecture
- if tests are successful, add use cases with higher requirements concerning bandwidth, latency or jitter
- if necessary, iterate over the standardization process

It would be also desirable to take the visualization of widgets into account when drafting the standard for visualization of data records, as these two visualization types go hand in hand.

Remote Instrumentation e-Learning. There are several organizations which develop e-Learning standards. It is recommended that the integration and standardization process between e-Learning systems and RI environment should be done after the case studies on the state of standards implementation within the particular e-Learning systems. This will allow avoiding the potential critical problems with data reusing and migration.

There were also recognized, besides mentioned above, two functional requirements as common to all types of remote experiments and which need RI-specific standardization activities: **error monitoring and logging**, **accounting and monitoring/controlling experiment execution**.

7 SUMMARY

This document is the final outcome of the RINGrid project. We have attempted here to provide the first set of elements for the conceptual design of a complete Remote Instrumentation Services Architecture. The OGSA reference model has been a guideline to our investigation. We have attempted to point out what are, in our opinion, the most relevant aspects that should be the subject of future study and development. The emphasis has been put initially on technological aspects, and on what is and will be shortly available in terms of middleware and networking technologies that may help in reaching the goals to be pursued in the RIS scenario. Then, the study has been complemented by use case analyses.

The main and absolutely necessary facilities identified are instrument virtualization, error monitoring and logging, accounting and monitoring/controlling of experiment execution. Additionally, some other functional requirements exist: the presence of selection services, the availability of site functional tests, policy decision and enforcement mechanisms, input and output data management, and workflow management. Further (non-functional) requirements involve input management (in the preparatory phase), the presence of (and interaction with) local operator(s), and the presence of specialized visualization devices.

The non-strictly-technological comments by users suggest some additional requirements to be addressed mostly conceptually and algorithmically: flexibility and adaptation to the user-specific experimental environment, flexibility and adaptation of the user interface to the user’s skills and level of expertise, and appropriate tools for self-training and learning-by-doing. More detailed information can be found in project documents that specifically address the issues of technology-related and user-related requirements [D4.3], [D6.2].

RINGrid

ACKNOWLEDGEMENTS

The RINGrid Consortium received funding from the European Commission in the 6th Framework Programme under the grant agreement no. RI-031891.

8 REFERENCES

[RINGRID] The RINGrid project website

http://www.rinGrid.eu

[D4.3] The RINGrid project deliverable D4.3 “Final report”

http://www.ringrid.eu/public/deliverables/RINGRID-WP4-D4_3-2008-02-13-Final.pdf

[D6.2] The RINGrid project deliverable D6.2 “Report on requirements verification”

http://www.ringrid.eu/public/deliverables/RINGrid-WP6-D6_2-2008-04-20-1-GRN-Report_on_requirements_verification.pdf.

[RISGE] OGF research group

http://www.ogf.org/gf/group_info/view.php?group=risge-rg

[GFD29] GFD-I-029, *Open Grid Services Architecture Use Cases*, Open Grid Forum, 2004,

http://www.ogf.org/documents/GFD.29.pdf.

[GFD80] GFD-I.080, *Open Grid Services Architecture*, Open Grid Forum, 2006.

http://www.ogf.org/documents/GFD.80.pdf