

Towards Innovative Healthcare Grid Solutions: *ViroLab - A Virtual Laboratory for Infectious Diseases*

M. Assel¹ and B. Kramer¹

HLRS - High Performance Computing Center Stuttgart,
Nobelstr. 19, 70569, Stuttgart, Germany
email: {assel, kramer}@hlrs.de
phone: (+49 0711) 685 62515, *fax:* (+49 0711) 685 65832

Abstract

With more rare and critical diseases, medical diagnoses and disease prevention constitutes one of the challenging and at the same time most difficult fields in modern health care. Using interdisciplinary approaches where different areas of research such as biology, chemistry, mathematics, and computer science meet together in order to analyze, understand and reconstruct complex medical processes, those projects typically achieve excellent research results but are inefficient and unfeasible for daily medical workflows.

The EU funded project *ViroLab* should address this important issue by combining different scientists together into one project in order to design and develop a working platform for researchers and especially for doctors, which facilitates medical knowledge discovery and decision support for HIV drug resistance for daily clinical usage.

In this paper, we will present the overall concept of the *ViroLab* project. The core functionalities will be described in detail and shall give the readers an idea of an innovative system for disease prevention, diagnosis and treatment based on integrated biomedical data and information on several levels.

1 Introduction

Every year, thousands of people die from HIV¹ and millions become infected. Currently, AIDS² is the fastest growing infectious disease on earth and its progression cannot be simply stopped or even reduced. Figure 1 provides an overview of the devastating problem of the HIV pandemic and shows the dramatic situation all over the world.

Medical treatment of HIV-infected patients has made significant progress during the past years. Almost twenty antiretroviral drugs are available, divided over four classes. Patients should take a combination of at least two different classes to achieve complete suppression of the virus. Such combination therapies have

¹Human Immunodeficiency Virus

²Acquired Immunodeficiency Syndrome

reduced mortality among infected patients but meanwhile, health professionals are fighting against a new phenomenon of modern medicine: *Drug Resistance*[1].

	HIV-infected adults and children	HIV prevalence among adults (%)	New infections per day	Daily deaths from AIDS
Subsaharian Africa	25,000,800	7.2	8,700	6,500
South and Southeast Asia	7,400,000	0.7	2,700	1,300
Eastern Europe and Central Asia	1,600,000	0.9	740	170
Latin America	1,800,000	0.6	550	180
East Asia	870,000	0.1	380	110
North Africa and Middle East	510,000	0.2	180	160
North America	1,200,000	0.7	120	50
Caribbean	300,000	1.6	80	70
Western and Central Europe	720,000	0.3	60	30
Australia, New Zealand and Pacific Region	74,000	0.5	20	10
Total	40,300,000	1.1	13,530	8,580

Figure 1: The AIDS Epidemic³

About one quarter of a million HIV individuals in the USA and Europe are living with drug resistant viruses and approximately ten percent of new infections in these countries occur with viruses that have at least one drug resistant mutation. Even a small proportion of patients die because there are no drugs to inhibit their viruses[2].

Drug resistance arises surprisingly quickly if viral replication continues in the presence of drugs. Mutations conferring resistance to antiretroviral drugs commonly result in a virus that replicates more slowly than HIV fully susceptible to therapy[3]. In order to determine drug resistance, genetic information such as nucleic acid sequences is needed. This information is widely available and has been used in Europe and beyond for a considerable number of years.

ViroLab[4] is a Specific Targeted Research Project of the EU 6th Framework Programme for Research and Technological Development in the area of integrated biomedical information for better health. The mission of the project is to provide researchers and medical doctors in Europe with a virtual laboratory for infectious diseases.

As a prototype for this virtual laboratory for infectious diseases, the problem of HIV drug resistance is used. The virtual laboratory will integrate the biomedical

³Extracted from the HIV Medicine 2006 Flyer: <http://www.hivmedicine.com>

information from viruses (proteins and mutations), patients (e.g. viral load) and literature (drug resistance experiments) resulting in a rule-based decision support system for drug ranking. In addition, *ViroLab* will include advanced tools for (bio)statistical analysis, visualization, modelling and simulation, enabling prediction of the temporal virological and immunological response of viruses with complex mutation patterns for drug therapy[5].

In the following, we will explain the approach and the outcomes of *ViroLab*, point out the major components and their functionalities, as well as emphasize the advantages and benefits of this research project.

2 Approach and Requirements

The complex interplay in HIV treatment as shown in figure 2 results in an interdisciplinary collaboration of multiple sciences where all scales, all disciplines and all data is required. Extracting information from appropriate literature, simulating drug-protein complexes and immune response, analyzing temporal data behaviour and / or population trends, and finally deriving rules for decision support, *ViroLab* will integrate all these methodologies from different domains into one collaborative working environment to overcome the difficult interactions and to come up with an approach that cannot be used only by experts.

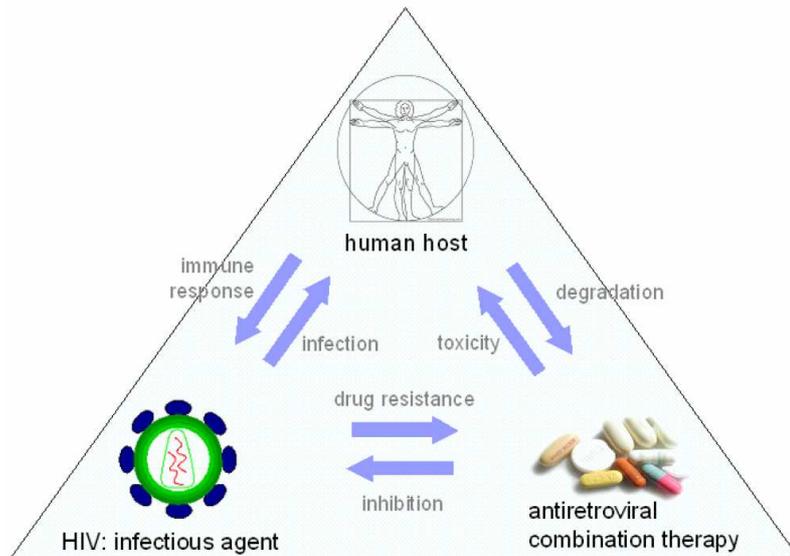


Figure 2: The Complexity of HIV Treatment⁴

⁴Adapted from *ViroLab* presentation at CGW'06, Poland by Prof. Peter Sloot

The long term objective of *ViroLab* is to provide a user-friendly and clearly arranged Graphical User Interface (GUI) available for clinicians and researchers distributed over Europe, which allows, on the one hand, the usage of well-defined rule sets in order to correctly predict virological and immunological response for all particular antiretroviral drugs, and on the other hand, the planning of experimental workflows for computing and analyzing genotypic resistance using specific development tools.

In order to develop an environment that enables easy access to distributed resources for sharing, processing, analyzing and simulating virological, immunological, clinical and experimental data, different technologies are required and more or less needed.

Mesoscopic simulation and analysis of biological processes rely on High Performance and High Throughput Computing due to their immense variety and complexity of chemical compositions. Therefore, access to resources providing lots of computational power such as local cluster systems and / or virtualized grid nodes should be made available and achievable to the laboratory users.

The federation and integration of heterogeneous and distributed data resources should be transparent to and hidden from the users by creating an abstraction layer that combines data disclosure and data fusion capabilities[6] to guarantee data access as well as data management in a simple but at the same time efficient way. Capabilities like distributed text / data mining algorithms necessary to perform semantic operations on heterogeneous data sets are also required and will be considered during implementation phase.

Dealing with confidential and private patient information, secure transmission and storage specifies one of the most important issues within the project. Building a Virtual Organization (VO) for secure exchange of information between several organizations using state-of-the-art grid technologies, this VO should form the "glue" for all components of the *ViroLab Virtual Laboratory* (refer to figure 4).

Furthermore, strong mechanisms for authentication and authorization also play a decisive role within *ViroLab* and demand particular attention while designing a flexible but sophisticated solution. Current developments in grid and distributed computing have been evaluated and the approach of Shibboleth[7] to build up an *Authentication Authorization Infrastructure (AAI)* was chosen.

The overall functionality provided by the virtual laboratory infrastructure will cover well-defined tasks for doctors' use in a clinical environment through the conception as well as design and implementation of novel studies with the available data resources and scientific tools. The main components of this infrastructure will enable features for session management, runtime system brokering, data archiving and access, and scientist collaboration and will be presented in the following section.

3 The *ViroLab* Virtual Laboratory

3.1 General Laboratory Architecture

The *ViroLab* Virtual Laboratory will be used by medical doctors to review previous results and rankings or by scientists to conduct new experiments and simulations starting from pre-defined process flow templates. To support and provide these different functionalities, the overall architecture consists of several components each serving a different purpose. In figure 3, an overview of involved modules and their main interactions are schematically shown. The diagram also proposes the main flow of information in the channels and the nature of data being pushed through them.

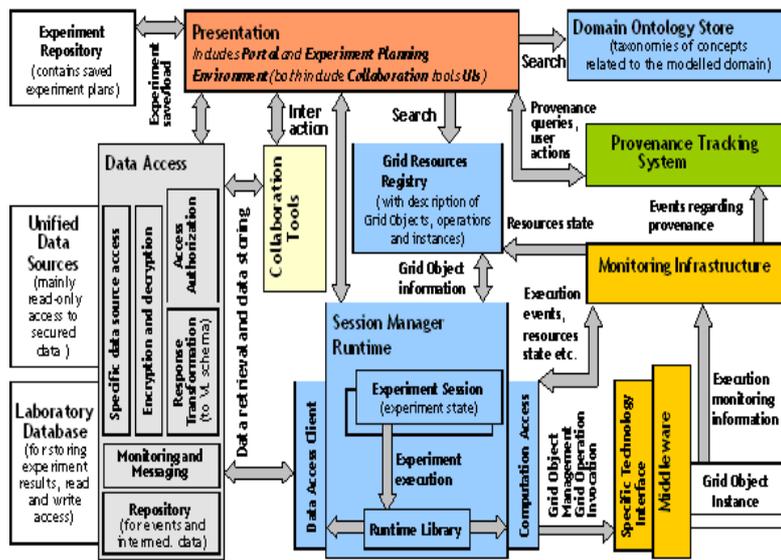


Figure 3: The General Laboratory Architecture

The topmost part is devoted to the presentation layer through which the user is able to interact with the system. Given three classes of users - experiment developer, experiment user, and clinical virologist - there are two main user interfaces provided by this layer: the *ViroLab* Portal dedicated to the medical and scientific users and the Experiment Planning Environment provided for the virtual experiments developers. The collaboration tools will have their graphical interfaces submerged in both these tools. The prepared experiments may be stored inside the Experiment Repository. Two persistent subsystems that are valuable in the process of experiment development are the Domain Ontology Store (which contains domain knowledge described in computer-legible formats) and the Grid Resources Registry. Both these modules

are there to help the developer choose proper computational and data resources for the planned experiment. Such an experiment may be executed afterwards using the Runtime system. The main responsibility of that part is to interpret the experiment plan script and, when necessary, to delegate any remote operations. All remote data access operations are sent to the Data Access server (there is a dedicated Data Access Client library to relay all the data retrieval and store communication). Similarly, all remote processing is done through the Computation Access library (which features both the Grid Operation Invoker and the Scheduler) that in turn contacts relevant software on the target computing element to pass along all the needed parameters and to retrieve all the results.

External processing also generates important events that may be useful for upper parts of the architecture. Therefore we include the Monitoring infrastructure that is able to acquire, store and provide interesting events in the form of notifications or responses to queries. Apart from the Runtime library and the Grid Resource Registry, an important peer of the monitoring module is the Provenance Tracking System. This component stores all the information relevant from the point of view of data (result) provenance tracking and publishes that information for interesting parties (after some optional reasoning). The main consumer of that data would be a user (particularly a virologist)[8].

The diagram shown in figure 3 is not guaranteed to be invariant - in fact, it is quite likely that in the course of project development some parts of the overall picture will change. Nevertheless, it presents the main subsystems that will be involved in the daily virtual laboratory operations and it identifies the main cooperative interfaces between these modules.

3.2 Virtual Organization Architecture

The Virtual Organization concept is used within *ViroLab* to provide users with access to requested resources no matter what institution they are from as long as they have appropriate access rights. The main advantage of using a VO-based approach is its manageability and flexibility without harming any resource provider's policy.

The architecture of *ViroLab's* Virtual Organization is depicted in Figure 4. The system is divided into three main layers (marked yellow): the presentation layer, which is the front-end for any *ViroLab* user, the security infrastructure including an AAI based on Shibboleth and mechanisms for data encryption, and a middleware stack that provides access to different grid middleware solutions like GT4, EGEE and others.

As already mentioned, the presentation layer offers two possibilities to enter the laboratory. The first one is the portal which the user accesses via a standard web browser. It gives access to all *ViroLab* resources for different kinds of users. The second tool is an Integrated Development Environment (IDE) for developing *ViroLab* applications named Experiment Planning Environment (EPE). It is based on the Eclipse RCP technology and is going to be used by scientific-oriented programmers. Thanks to EPE, one can efficiently implement *ViroLab*

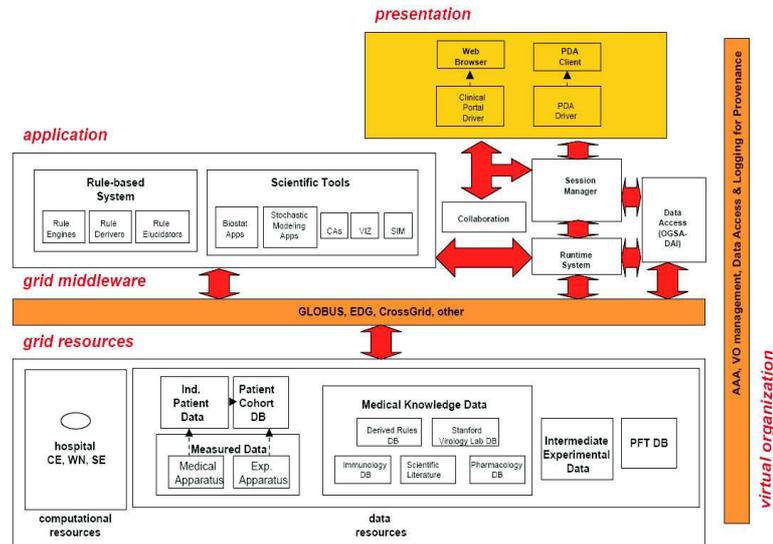


Figure 4: The *ViroLab* Virtual Laboratory from the Grid (VO) Perspective⁵

Both tools described contain all user interfaces and are responsible for authentication by interacting with the security layer which is split into two independent but essential functionalities. Authentication and authorization are based on Shibboleth techniques whereas secure communication is ensured through the concept of Grid Security Infrastructure (GSI) provided amongst others by the Globus Toolkit[9]. As *ViroLab* deals with personal medical data which belongs to the most sensitive data types in terms of privacy considerations reliability and robustness of the security solution is critical. The final decision whether the access for a given user to one specified resource should be granted or not must always belong to the owner of the resource. Therefore, the ultimate authorization decision is always taken by the service / data provider based on user's attributes. On the other hand, to ensure scalability and manageability of the system, attribute assignment is performed by user Home Organization's facilities. This indeed requires good level of trust between administrators of both systems.

The middleware layer provides interfaces between the laboratory components and the underlying grid infrastructure (view figure 3 and 4). In a typical grid architecture, there are three basic aspects included: resource management (access to computation), information management (access to monitoring) and data management (access to data). Since the data access is covered by separate task

⁵Adapted from *ViroLab* presentation at CGW'06, Poland by Prof. Peter Sloot

within the project only the two former ones are in the scope of the middleware task. An important assumption for designing the *ViroLab* Virtual Laboratory was that there are many already existing middleware solutions. The goal of the middleware subsystem is then to reuse the existing grid middleware and infrastructures that are accessible to the partners without developing specific solutions from scratch.

4 Summary

Designing and developing applications for supporting and even improving medical diagnoses and treatment constitutes one of the most challenging fields in computer sciences. *ViroLab*, an European research project funded by the European Commission will accept this challenge in order to support doctors and researchers in addressing the problem of HIV drug resistance.

This paper focuses on the idea and the main concepts of *ViroLab*. It provides the readers with an insight into a particular and at the same time interesting research field within the Grid Computing area based on the integration of biomedical resources (data and applications) into a collaborative working environment in order to improve medical diagnoses and treatment of HIV-infected patients.

We have presented appropriate requirements that should not to be considered only, but necessarily realized to overcome the complexity and variety of biomedical processes. The core functionalities of the *ViroLab* Virtual Laboratory were described in a high-level detail due to the early stage of development as well as their difficult internal interactions. Nevertheless, we hope that *ViroLab* can successfully achieve its objectives and that the present paper encourages people to focus their research on complex medical interplay.

Acknowledgements

The research presented in this paper has been made possible through the support of the European Commission *ViroLab* Project Grant 027446 and all members of the *ViroLab* consortium. The authors want to thank all who contributed to this paper.

References

1. Deeks S. Treatment of antiretroviral.drug-resistant HIV-1 infection. *Lancet* 2003; 362(9400):2002-2011.
2. Jaggy C., von Overbeck J., Ledergerber B., Schwarz C., Egger M., Rickenbach M. et al. Mortality in the Swiss HIV Cohort Study (SHCS) and the Swiss general population. *Lancet* 2003; 362(9387):877.878.
3. Nijhuis M., Deeks S., Boucher C. Implications of antiretroviral resistance on viral fitness. *Curr Opin Infect Dis* 2001; 14(1):23-28
4. ViroLab - EU IST Project (IST-027446). <http://www.virolab.org>

5. Sloom P., Boucher C., Bubak M., Hoekstra A., Plaszcak P., Posthumus A., van de Vijver D., Wesner S. and Tirado-Ramos A. VIROLAB - A Virtual Laboratory for Decision Support in Viral Diseases Treatment. Cracow Grid Workshop 2005, Cracow, Poland, November 2005.
6. Assel M., Krammer B., Loehden A. Management and Access of Biomedical Data in a Grid Environment. Cracow Grid Workshop 2006, Cracow, Poland, October 2006.
7. Barton T., Basney J., Freeman T., Scavo T., Siebenlist F., Welch V., Ananthakrishnan R., Baker B., Goode M., Keahey K. Identity Federation and Attribute-based Authorization through the Globus Toolkit, Shibboleth, GridShib, and MyProxy. Proceedings of 5th Annual PKI R&D Workshop. March 2006.
8. Bubak M., Gubala T., Malawski M., Kasztelnik M., Bartynski T., Nowakowski P. Virtual Laboratory in ViroLab. Cracow Grid Workshop 2006, Cracow, Poland, October 2006.
9. Filipovic B., Straub T. Grid Security Infrastructure - Ein Ueberblick. 20. DFN-Jahrestagung. Heilbronn, Germany, June 2006.