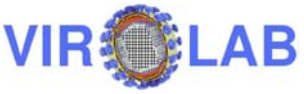
Data Management in EU Project ViroLab



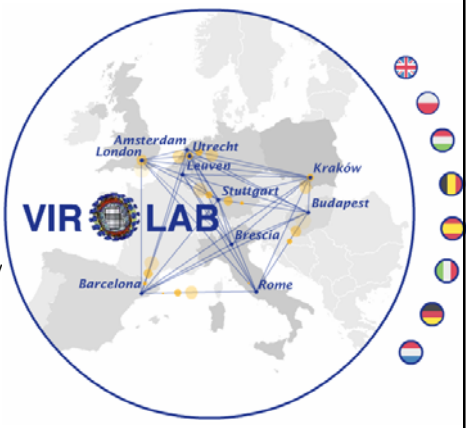
Matthias Assel, Bettina Krammer, Aenne Löhden
University of Stuttgart
High-Performance Computing-Center Stuttgart (HLRS)
www.hlrs.de

Outline

- Introduction to ViroLab
- Data Management in ViroLab
- Security
- Future work
- Conclusion

Project Overview

- EU-funded STREP (6th Framework)
- Project duration: 1 march 2006 – 28 feb 2009
- Partners:
  - University College London, UK
  - Gridwisetech, Poland
  - ACK Cyfronet AGH, Poland
  - Eotvos Lorand University, Hungary
  - Catholic University of Leuven, Belgium
  - IRSICAIXA Foundation, Spain
  - Catholic University of Rome, Italy
  - University of Brescia, Italy
  - University of Stuttgart, HLRS, Germany
  - University of Amsterdam, Netherlands
  - University Medical Center Utrecht, The Netherlands
  - Virology Education, Netherlands

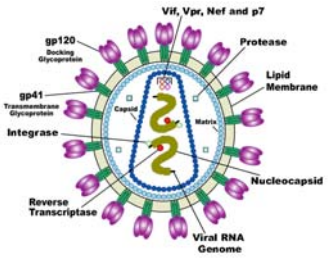


Mission of ViroLab

- provide researchers and medical doctors in Europe with a **Virtual Laboratory for Infectious Diseases** that facilitates medical knowledge discovery and decision support
- enable easy access to distributed resources as well as sharing, processing and analysing virological, immunological, clinical and experimental data
- prototype for this virtual laboratory: **HIV drug resistance**

Treatment of HIV

- In 1996, several novel anti-HIV drugs became available
  - Several classes of antiretrovirals available
  - Only combination of at least 2 different classes successful in the long-term
  - Dramatic reduction of mortality among HIV-infected patients



- **BUT**
  - Serious side effects
  - Expensive medication
  - Genetic variation of HIV very high
  - Drug resistance occurs frequently
    - Good adherence is key
    - Drug resistance interpretation tools have been developed, e.g. Retrogram

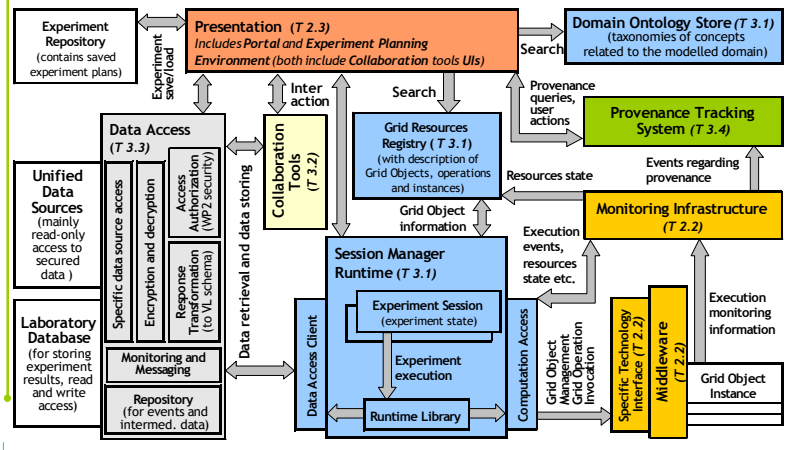
Mission of ViroLab

- prototype for this virtual laboratory: **HIV drug resistance**
  - increasing problem world-wide
  - integrate biomedical information from
    - **viruses (proteins and mutations),**
    - **patients (e.g. viral load) and**
    - **literature (drug resistance experiments)**

    resulting in a rule-based distributed decision support system (DSS) for drug ranking
  - include advanced tools for
    - **(bio) statistical analysis,**
    - **visualization, modelling and simulation**

    to enable prediction of the temporal virological and immunological response of viruses with complex mutation patterns for drug therapy

Architecture



VIROLAB

HLRS

Objectives of ViroLab

- develop a virtual organization (VO) that binds the various components of the ViroLab virtual laboratory and guarantees a maximum of security and trustworthiness
- develop a virtual laboratory system infrastructure for transparent workflow, data access, experimental execution and collaboration support
- virtualize and enhance state-of-art in genotype resistance interpretation tools (applications), integrating them into the virtual laboratory
- establish epidemiological validation that ViroLab correctly and quantitatively predicts virological and immunological outcome, and disseminate the results to other European medical experts

- Role of HLRS: Data Access

VIROLAB

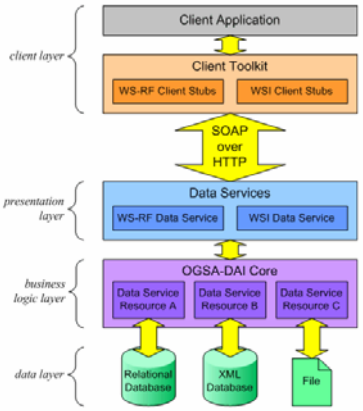
HLRS

Data Management

OGSA-DAI

- **OGSA-DAI = Open Grid Services Architecture Data Access and Integration**
- Middleware product allowing database access (e.g. Oracle, IBM DB2, MySQL, PostgreSQL, XML DB, etc.) on the Grid (GT4, Unicore)
- Offers data integration services to clients
- Allows data to be queried, updated, transformed and delivered
- Compliant with two popular web service specifications
  - WS – I (Web Services Inter-operability)
  - WSRF (Web Services Resource Framework)
- Java API
- Developed by University of Edinburgh
- Open Source, available at www.ogsadai.org.uk
- Used in a number of projects

OGSA-DAI architecture



Data Virtualization and Access in ViroLab

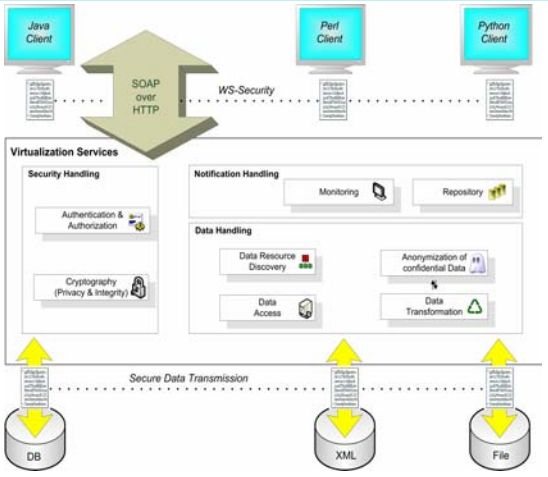
Objective: *Single point of access with standardized interfaces that provides transparent, consistent and data resource independent access to distributed heterogeneous data resources (clinical, virological and drug DBs)*

User Example: *"Find all records for patients with specific symptom and specific age group"*

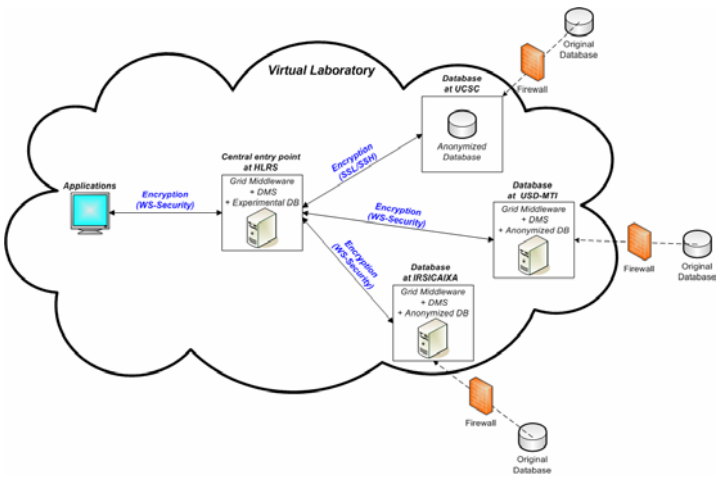
- Data access based on OGSA-DAI's core functionalities - mainly accessors for different resources (e.g. MS Access DB, MySQL, files)
- Extensions needed for ViroLab purposes
  - Discovery service queried with unified meta query language
  - Transformation service used to convert heterogeneous data sets
  - Security infrastructure adapted to OGSA-DAI's security mechanisms
- Management application facilitating the administration of access rights

Data Virtualization and Access

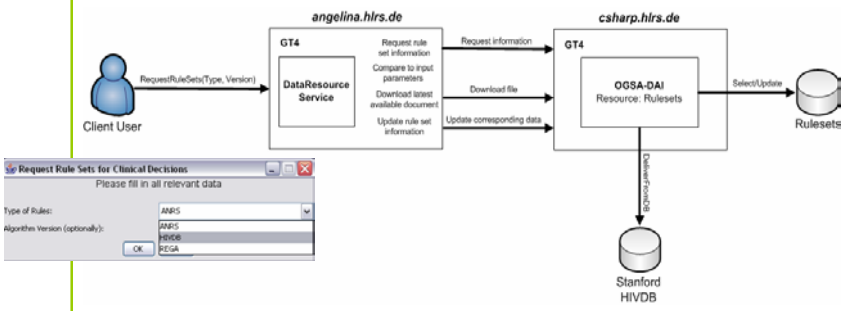
Virtualization Layer – Logical Components



Example: Data Access within ViroLab (hospitals)



Example: Request Ruleset for Drug Ranking Support



Service available at:

http://angelina.hlrs.de:8080/wsrf/services/DataResourceService

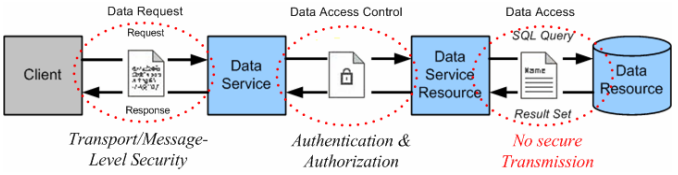
Security

Security and Privacy in ViroLab

- privacy: anonymized data sets to be provided by data providers (patient DB)
- authentication and authorisation grants access to activities and data service resources to certain users (Shibboleth)
  - authentication based on LDAP, performed at the Home Organisation (HO) site
  - authorization to be performed by PDP (Policy Decision Point)
  - user privileges based on attributes
  - access control of local data management systems to be performed by database administrators
  - usage of mapping functionalities from OGSA-DAI to map user attributes onto database user credentials
- data encryption
  - via GSI (Grid Security Infrastructure) provided by GT4
  - TSL/SSL

OGSA-DAI Security Infrastructure

- Globus Toolkit Security
  - Transport-Level Security (encrypts complete transmission)
  - Message-Level Security (encrypts only SOAP message)
- „Authorisation Interface“ where developers can plug in their own authorisation mechanism
  - Resource and Activity Authorisation
  - Role Mapping (based on X.509 certificates)



Future Work

- Security:
  - DAS needs to be shibbolized in order to perform attribute-based authorisation
  - Develop PDP for OGSA-DAI
- integrate meta query language to allow easier distributed querying
- Provenance
  - Data Access responsible for storing input as well as result data
  - linked via unique identifier
  - Provenance stores its own data (ontologies)

Conclusion

- Integration of biomedical resources (data, applications) into a collaborative working environment is a challenge
- ViroLab is in an early stage of development but promising
- Security is crucial!

Thanks for your attention