PARTNER
Grant Agreement Number 215840

WP22 – Deliverable 3 – 03/2011
Data-driven platform for hadron therapy

Faustin Laurentiu Roman

Host Organisation

CERN

Supervisors
Prof. Manjit Dosanjh
Prof. Jose Bernabeu

April 2011
Data-driven platform for hadron therapy

PARTNER Work Package 22, Milestone 3
Prototype grid hadron therapy test bed
Make the common elements available in the EGEE/EGI infrastructure,
demonstrate added value

Faustin Laurentiu Roman
PARTNER ESR, CERN
Faustin.Roman@cern.ch

April 2011
Abstract

This is the Deliverable 3 of the PARTNER Work Package 22 within the Marie Curie Initial Training Fellowship of the European Community’s Seventh Framework Programme under contract number (PITN-GA-2008-215840-PARTNER). The relevance of information sharing to hadron therapy and technical challenges have been discussed in the first deliverable. The legal and ethical considerations with regard to collaborative applications in hadron therapy have been described in the second deliverable, common to WP 22 and WP 23. This deliverable, following the concepts presented in the previous two deliverables, presents the PARTNER grid platform, software components integration with the EGI infrastructure and added value.

PARTNER grid prototype is a data-driven platform that allows medical doctors and patients to record the side effects during and after hadron therapy treatment. In addition, doctors can view a record of the patient’s side effects across institutions and researchers can request access to statistical information. Original medical data are left where it was created, and accessed using federated queries via a Web 2.0 portal.

This report describes the platform architecture and the software components implementation: databases, database federation, portal and the security framework. Finally, the report summarizes the platform status, with screenshots of main features.

Supervisors: CERN (host institute): Prof. Manjit Dosanjh, manjit.dosanjh@cern.ch
IFIC (associated institute): Prof. Jose Bernabeu, bernabeu@ific.uv.es

Co-Supervisors: IFIC: Dr. Jose Salt Cairols, salt@ific.uv.es
IFIC: Dr. Gabriel Amoros, amoros@ific.uv.es
Contents

Overview 1
1.1 Introduction .................................................. 1
1.2 Platform architecture ........................................ 2
1.3 Workspace analysis ........................................... 3

Software implementation 6
2.1 Databases ...................................................... 7
2.2 Database federation ........................................... 7
2.3 Portal .......................................................... 9
2.4 Security framework and grid services ....................... 10

Conclusion 12
Appendix 13
Glossary 18
Bibliography 18
Overview

This work is a joint project of the PARTNER Wp 22, 23, 24 Early Stage Researchers D. Abler, V. Kanellopoulos, F. L. Roman (CERN) within the framework of the CERN PARTNER Grid Project (M. Cirilli, A. Di Girolamo, M. Dosanjh, J. Shiers, A. Valassi), in collaboration with IFIC (G. Amorós, J. Bernabeu, A. Fernández, V. Méndez, J. Salt Cairls), University of Oxford (J. Davies, S. Harris, K. Peach), University of Surrey (K. Kirkby, N. Kirkby) and the Cambridge University Hospital NHS Foundation Trust (R. Jena).

1.1 Introduction

Health grids are infrastructures comprising applications, services or middleware components that deal with the specific problems arising in the processing of biomedical data using grid technology. “Grid” computing means coordinated resource sharing and problem solving in dynamic, multi-institutional organizations whilst ensuring strict confidentiality [1]. Resources in grids can be databases, computing power, and even medical devices[2].

The PARTNER [3] (Particle Training Network for European Radiotherapy) ICT Packages objective is to create such a grid test-bed for hadron therapy, a collaborative project among Work Packages (WP) 22, 23 and 24. WP 22 [4] concentrates on core services such as databases, federation, portals and the security of the infrastructure.

The relevance of information sharing to hadron therapy and technical challenges have been discussed in the PARTNER Deliverable 1 of WP 22 [5] and WP 23 [6]. The legal and ethical considerations with regard to collaborative applications in hadron therapy, namely patient referral and cancer research, have been described in the PARTNER Deliverable 2 common to WP 22 and WP 23 [7].

Applications were derived from brainstorming discussions among the PARTNER Grid project, and a list of requirements (Annex, Fig. 2.13) was drafted and sent to the European Grid Initiative (EGI) [8] by the PARTNER-grid collaboration. EGI represents a follow-up of previous grid infrastructures for science in Europe, different in concept and implementation[9]. Valuable clinical input was provided by Dr. Rajesh Jena (Cambridge Univ. H. NHS Foundation trust), Prof. Ramona Mayer (EBG MedAustron), Prof. Bledwyn Jones (Gray Inst., PTCRi) and discussions with Valencia & Oxford hospitals to which we are grateful.

This report gives an overview of the platform and describes the present status of the test-bed. The remainder of this report is structured as follows: Section 1.2 introduces the platform architecture, describing the history and requirements of the platform based
on a specific scenario; next section presents the portal structure, basic services and the
roles of users. Following the introduction, the actual software components are presented,
the databases, database federation, user interface and the security solutions, their imple-
mentation and characteristics. Finally, the report summarizes the platform status and
outlines next steps, with screenshots of main platform features.

1.2 Platform architecture

The initial design of PARTNER grid platform, called HISP (Hadron therapy Informa-
tion Sharing Platform) [10] implied two distinct scenarios:

- a clinical scenario in which a patient goes from a local hospital to a hadron therapy
  center for treatment (referral)

- a research scenario in which clinical data are used to draw scientific conclusions

The scenarios expose the medical domain challenges from different perspectives: in the
first case you want to follow the patient around Europe, and make sure that his record
is up to date and persistent, in the second case we are talking about analysis on a set of
data scattered in different places, recorded in different formats, at various time frames.
The scenarios helped us to identify an initial set of services and their functions for HISP.
Considering the project challenges in getting real data, timescale, manpower, after dis-
cussions it has been decided to restrict the scope of the project to a subset of meaningful
functions in the form of a side effects recording system.

Side effects [11] recording is part of daily treatment quality assurance, in clinical trials
is used to determine the effects for specific therapies within the scope of treatment and
in research to understand the pathology of these effects to develop strategies for their
prevention. Data recording is based on different scoring systems and the actual recording
could be done in multiple locations (if patient is referred to a hadron therapy center and
is monitored home), periodically, over significant time periods (~5 years and more).

The side-effects scenario uses HISP as a data-driven platform for recording and sharing
medical data, allowing medical doctors and patients to record the side-effects during
and after treatment. Doctors can access patient’s side effects across institutions and
researchers can request access to statistical information. The conceptual architecture is
shown in Fig. 1.1, result of a meeting of the PARTNER-grid collaboration.

For this specific project, data which normally resides in hospital systems such as Hos-
pital Information Systems (HIS), Picture Archiving and Communication System (PACS),
Treatment Planning System (TPS), will be first generated from published clinical trials
and then imported in a database which emulate the hospital environment. Data recorded
by patients will reside in a specific patient database.

An important medical requirement is to provide secure cross-border access to het-
erogeneous repositories while leaving data where it was created and not copying it cen-
trally (data warehousing), due to data privacy and ownership concerns. HISP uses fed-
erated queries to provide users with a unified view over distributed resources as a virtual
database. Domain specific data recording forms are generated and linked to a metadata
registry, this will be described in PARTNER Deliverable 3 of WP 23.
1.3 Workspace analysis

In HISP the roles of users are grouped into workspaces, group of pages displaying specific tools depending on the user role:

- New Users: generic workspace used as main portal space, login and registration
- Hospital follow-up: medical doctors view patient history and record side-effects, part of clinical follow-up
- Patients follow-up: patients record their side-effects and quality of life during and after the treatment
- Research: researchers can analyze interesting data sets.

The initial description of the pages, roles and (basic) actions various actors have in the platform is grouped in Fig. 1.2.
The web portal login page provides the HISP description and based on the role the user has, after successful sign-in, it will be redirected to a specific service within a portal workspace:

- **New User**
  - see portal description
  - register
  - sign-in

- **Medical Doctors**
  - search patient record from Hospital1 HIS or create entry
  - search/read data entries
  - insert (update) data entries - local or from H2 DB
  - create forms for follow-up
The roles are implemented both in Liferay portal and Vine toolkit, using the administration interface of the portal (Annex, Fig. 2.8).
Software implementation

The part of HISP software implemented in the context of WP22 can be split into the following components:

- databases and federation
- portal and user interface
- security framework and grid services

The various software components were chosen on the basis of an open source, standard, free to use and enterprise ready philosophy.

The HISP system (Fig. 2.3) is available at http://vopartner01.cern.ch:8080/web/guest.
2.1 Databases

Hospital systems present several challenges regarding interoperability and access policies which make an unified view over distributed resources difficult (see [7] for details). Examples of such challenges relevant for HISP are:

- hospital systems are normally not linked (except if parts of a single entity)
- medical data are rarely shared
- few systems offer to patients opportunity to input data
- access rights to clinical data is strongly controlled by varying local policies for user authentication
- policies regarding the use of medical data for research vary between countries and institutions.

As described in Sec. 1.2, due to difficulties in obtaining real data, data will be first generated from published clinical trials and then imported in a database which emulate the hospital environment. Data recorded by patients will reside in a specific patient database.

HISP uses MySQL[12], one of the most popular open source database. If needed, MySQL could be exchanged with a compatible database technology, both open-source or commercial. MySQL is used for the portal internal database as well as for the medical database emulation.

The Virtual DB schema, representing the medical domain necessary to analyze side-effects, is described in the PARTNER Deliverable 4 of WP 24[13]. As shown in Fig. 2.4 the schema has 4 parts, the first 3 being imported from the hospital systems and only the side effects part is actually recorded in HISP platform:

1. patient information: demographics, history, etc
2. tumor information: type, staging, etc
3. treatment information: therapies and characteristics
4. side effects information

The schema was implemented in a MySQL database available at: vopartner02.cern.ch:3306/hisp2

2.2 Database federation

An essential component of the platform is database federation: aggregation of heterogeneous data sources in a virtual database. The virtual database does not contain the data itself but just information about the actual data and their location. The original data are left where it was created therefore fulfilling the medical domain constraint.

Making a single call to multiple data sources and then integrating and organizing the data in a middleware layer is also called data virtualization, enterprise information integration (EII) or information-as-a-service.
HISP uses JBoss Teiid [14] to provide users with a unified view over distributed resources as a virtual database. Teiid uses standard JDBC interface to RDBMS, files and web services and the virtual databases can be federated:

- dynamically: using an XML file where all data sources are mapped statically, appearing as a single virtual database to client application - present status in HISP

- modeling the federation in Teiid designer, allowing virtual views - future step in HISP

Teiid is installed in a vopartner02.cern.ch JBoss Application Server exposing a virtual database view over 2 MySQL databases (Fig. 2.5):

- **HISP1**: jdbc:mysql://vopartner01.cern.ch:3306/hisp1
- **HISP2**: jdbc:mysql://vopartner02.cern.ch:3306/hisp2
- **dynamic Virtual DB (VDB) view**:
  
  | DynamicHISPVDB@mm://vopartner02.cern.ch:31100 |

- **Patient DB** is not part of the federation since it is directly accessible within the portal, but can also be integrated in the VDB View.

In the first simple case, the database schemas of the local hospital DBs (HISP1 and HISP2) will be the same. At a latter stage, we will introduce heterogeneous schemas that will be closer to the real hospital environment. This will test the federation and mapping capabilities of HISP.
2.3 Portal

A portal is a web based application that provides authentication, customization and content aggregation from different sources and allows users interact with information systems. Portal pages may have different set of portlets creating content for different users. A portlet is an application that provides a specific piece of content (information or service) to be included as part of a portal page. It is managed by a portlet container, that processes requests and generates dynamic content. Portlets are used by web portals as user interface components that provide a presentation layer to information systems.

Liferay portal[15] is an enterprise web platform, free and open source which has built-in user management, content management and conforms with many standards such as AJAX, JSR-168[16], JSR-286 (Portlet 2.0)[17], JSF-314 (JSF 2.0)[18]. Considering the medical domain requirements regarding standards, certification and support, Liferay portal provides a good alternative to other ‘grid’ portals, e.g. Gridsphere.

An important reason for choosing Liferay is that grid frameworks such as Vine Toolkit[20] (see 2.4 section) are compatible with it. The combination of Liferay and Vine Toolkit provides an easy to use grid-enabled environment where additional services could be prototyped and tested.

The portal is installed at CERN data center (http://vopartner01.cern.ch:8080/web/guest/):

- operating system: Scientific Linux 5 x64
- portal database: local MySQL DB instance
- server container: Tomcat 6
User interface

Openxava[19] is a framework to fast produce Rich Internet Applications based on Java programming language. It is compatible with Liferay and can deliver CRUD\(^1\) interfaces as shown in Fig. 2.6, an example view of patient database.

![User interface example](image)

Figure 2.6: User interface example

2.4 Security framework and grid services

HISP is a prototype in the first phase of implementation, therefore security layer is not yet fully implemented. For the moment the portal authentication is done with a basic username/password method. Since the medical domain requires more sophisticated security methods and the research part of HISP could be computational intensive, the platform supports also the grid environment by means of the Vine Toolkit (VT)[20].

VT is a software framework that adds the Grid context to the web components and abstracts different middlewares implementation. In HISP, VT creates a link between the Liferay portal and VOMS\(^2\), the grid authentication and authorization grid service. A user that has a X509 grid certificate uploaded in the portal will be able to access transparently grid storage elements and submit jobs to computing elements configured in VT (see Fig. 2.8 for example of user with grid credentials mapping) in the EGI infrastructure. EGI is a follow-up of previous grid infrastructures for science in Europe, and the concept is built on each member state’s national grid initiatives which can represent the grid resources as well as the local user communities[9].

At present the HISP roles are implemented as defined in Sec. 1.3, with three main roles (doctor, patient, researcher) and two additional roles (new user and administrator). The

---

\(^1\)Create, Read, Update, Delete interface to databases

\(^2\)HISP VOMS is available at: [https://voms.ific.uv.es:8443/voms/vo.partner.eu/](https://voms.ific.uv.es:8443/voms/vo.partner.eu/)
roles are implemented both in Liferay portal and Vine toolkit, using the administration interface of the portal (Fig. 2.8), but a better synchronization between the two frameworks would be needed. Once the research part of the platform will be defined and analysis jobs can be launched in the grid environment, a better security mechanism, such as grid security using Shibboleth or robot certificates, will be preferable.
Conclusion

This report presents the PARTNER grid platform, first integration steps towards a prototype data-driven collaborative portal as described in previous PARTNER WP 22 deliverables.

Liferay portal provides an Web 2.0 interface to distributed MySQL databases which are federated using JBoss Teiid. The platform can be integrated into EGI environment using the Vine toolkit, that could be used in future to perform research analysis in the grid. An added value of the platform is that it uses a standard open source, free to use and enterprise ready or compatible software components, therefore ensuring software sustainability and easy transition to real-life environment.

Next steps in HISP are to populate databases with the generated data, design domain specific data recording forms and link the semantic part to the platform. For the research part grid-enabled workflow engines could be used. Also the grid framework interoperability could be tested by exchanging Vine Toolkit with other frameworks.
Appendix

HISP portal is available at: http://vopartner01.cern.ch:8080/web/guest/
The portal screenshots demonstrate:

- Portal first page (Fig. 2.7): provides platform description and login.

- User management (Fig. 2.8): the integration of the Liferay portal and Vine toolkit. Users can be added, deleted or configured. Roles can be associated to each user.

- Data view (Fig. 2.9): first database view portlet with generated data, showcasing the Openxava CRUD interface.

- Data selection (Fig. 2.10): example of data filtering.

- Data import (Fig. 2.11): interface to collect data.

- Simple form (Fig. 2.12): Liferay based simple-form.

- Requirements matrix for EGI (Fig. 2.13)
Welcome to Hydrotherapy Information Sharing Prototype (HISP) Platform.

HISP is a distributed platform for recording and sharing medical data within the hydrotherapy community.

**Scope**

The idea is to provide secure access, browse access to heterogeneous databases, while allowing data where it was created and not copying it centrally (data warehousing duties obviated and computing concerns). HISP provides tools to describe local data using metadata and use federated queries to provide users with a unified view over distributed resources as if in virtual databases.

**Usage**

The portal and an easy-to-use tool for all the actions involved in the hydrotherapy domain. The actions are grouped into categories that have specific tools for:

- Follow-up: medical and patient records for follow-up information, clinical tools and quality of life during and after the treatment.
- Discover: researchers can search for interesting data sets and construct complex queries over federated databases. MAPPers will be the tool...

Once logged in, you can register in one of the available communities. Your registration will be then approved by an administrator.

**References and tools**

- Figure 2.7: Portal Login Page
- Figure 2.8: User administration
Figure 2.9: Data view

Figure 2.10: Data selection
Figure 2.11: Data import

Figure 2.12: Simple form
Figure 2.13: Requirements matrix for EGI

<table>
<thead>
<tr>
<th>REQUIREMENTS FOR</th>
<th>ARCHITECTURE</th>
<th>FUNCTIONAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROJECT</td>
<td>DISCIPLINE</td>
<td>Service interoperability</td>
</tr>
<tr>
<td>Hadoop Therapy (HSP)</td>
<td>Medicine</td>
<td>C = B P = 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The platform must follow standards.</td>
</tr>
</tbody>
</table>

- **C score** = criticality score
- **P score** = priority score

Textbox: Explanation relevant to each use case.

<table>
<thead>
<tr>
<th>NON-FUNCTIONAL</th>
<th>Software Development Kit</th>
<th>Open Source</th>
<th>Response time of services</th>
<th>Reliable (robust) service</th>
<th>Resilience</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comprising documented API and sandbox environment</td>
<td>Standards based open source code that can be re-used and changed if necessary.</td>
<td>Data between submission of a request and obtaining the reply, low scores indicate that response times in minutes to hours are acceptable.</td>
<td>Reliability of service guaranteed that each request processed a reply is obtained (nothing 'gets lost'). This includes human readable messages in case of failure.</td>
<td>Ability of system to function when parts of it are not functioning. System depends on predictable and well behaved way.</td>
</tr>
<tr>
<td></td>
<td>C = 7 P = 5</td>
<td>C = 10 P = 8</td>
<td>C = 7 P = 5</td>
<td>C = 7 P = 5</td>
<td>C = 7 P = 5</td>
</tr>
</tbody>
</table>

- **C score** = criticality score
- **P score** = priority score

Textbox: Explanation relevant to each use case.
# Glossary

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AJAX</td>
<td>asynchronous JavaScript and XML</td>
</tr>
<tr>
<td>CERN</td>
<td>European Organization for Nuclear Research</td>
</tr>
<tr>
<td>CRUD</td>
<td>Create, Read, Update, Delete</td>
</tr>
<tr>
<td>DB</td>
<td>Database</td>
</tr>
<tr>
<td>EGI</td>
<td>European Grid Infrastructure</td>
</tr>
<tr>
<td>EGEE</td>
<td>Enabling Grids for E-Science</td>
</tr>
<tr>
<td>ESR</td>
<td>Early Stage Researcher</td>
</tr>
<tr>
<td>HIS</td>
<td>Hospital Information Systems</td>
</tr>
<tr>
<td>HISP</td>
<td>Hadron therapy Information Sharing Platform</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technology</td>
</tr>
<tr>
<td>JDBC</td>
<td>Java Database Connectivity</td>
</tr>
<tr>
<td>PACS</td>
<td>Picture Archiving and Communication System</td>
</tr>
<tr>
<td>PARTNER</td>
<td>Particle Training Network for European Radiotherapy</td>
</tr>
<tr>
<td>PROM</td>
<td>Patient-reported outcome measures</td>
</tr>
<tr>
<td>RDBMS</td>
<td>Relational Database Management System</td>
</tr>
<tr>
<td>TPS</td>
<td>Treatment Planning System</td>
</tr>
<tr>
<td>VDB</td>
<td>Virtual DB</td>
</tr>
<tr>
<td>VOMS</td>
<td>Virtual Organization Membership Service</td>
</tr>
<tr>
<td>VT</td>
<td>Vine Toolkit</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
Bibliography


