



# **Comparison of European Grid Projects**

**Jarek Nabrzyski, Ariel Oleksiak (PSNC)**

**Project:**

GEMSS

**Area:**

Data Management

**Table of Contents**

- 1.Introduction.....3**
- 1.1.Objective and Structure.....3**
- 1.2.Uniform description.....3**
- 2.Data Management.....4**
- 2.1.General.....4**
- 2.2.Details .....4**
- 2.3.External.....5**

## 1. Introduction

### 1.1. Objective and Structure

This document is one of thirteen templates that have common goal to gather information related to main European Grid Projects in order to make their accurate comparison in the framework of GRIDSTART initiative. We believe that the participation of particular projects members in preparation of this document will allow comparing all activities in a credible and exhaustive way.

The proposed structure of the description consists of two parts. The former is concerned with the general overview and architecture together with the contents of layers (the first template). The latter includes the main components of the Grid infrastructure (remaining 12 templates). Since information regarding the project architecture is to be quite general, more detailed description should be provided in the review of the main aspects of the Grid infrastructure. In order to prepare uniform description for each project, we identify the important issues that have to, should or can be included into particular components. Common issues for all components and these specific for this component are briefly described in the next section.

We ask you to proceed according to this schema. However, a feedback is obviously welcome. For some projects the document has been partially completed on the basis of descriptions found at the official web pages. In this case, we ask you to revise already filled in sections, correct and complete them if necessary.

You should take into consideration future plans while you fill in particular sections. Actually they are even more important than the current state of the project components. If you are not going to design some elements in the scope of the project at all, please, note it in the proper section.

### 1.2. Uniform description

All the descriptions of the Grid infrastructure components are divided into three parts: **General** section includes main requirements and functionality, **Details** section relates to the issues specific for particular component and **External** defines its connections with other components and users.

As it was mentioned above, some of the issues are common for all components or at least repeat for many of them. Such issues, appearing for many or even all areas are shortly characterized below.

In **General** section:

**Main requirements** determine the objectives and requirements of the workpackage or the software module responsible for the design of functionality related to the particular domain of the Grid infrastructure.

**Functionality** contains a set of operations provided by the project in the given area.

In **External** section:

**Interfaces** define services, SDKs, APIs and so forth which can be used in order to access the functionality of the component.

**Low level Grid middleware** is the middleware providing basic Grid functionality as for example Globus or UNICORE.

**Relations with other components** determine components that utilize or are utilized by component being described as well as data and information flow between them.

Issues that are specific for this particular domain of the Grid infrastructure are presented in the sequel. Some of them, which we consider to be clear, have been skipped, however, if they turn out to be vague, please, contact the authors of this document ([ariel@man.poznan.pl](mailto:ariel@man.poznan.pl)).

The **Details** section contains all issues that allow coping with large data sets in a distributed grid environment.

Among them **Data Movement** defines methods for fast data transfer between different sites. **Data Migration** is concerned with the situation when a job is checkpointed and migrated into another machine. Job migration is accompanied by data migration. **High level data management** paragraph determines whether any solution is being developed to add semantics to data, create an ontology, keep track of data transformations etc.

## 2. Data Management

### 2.1. General

This document presents a snapshot of the current development. The GEMSS design is still being discussed by the project and not all issues have been finalized yet.

- **Main requirements**

Data is submitted along with the Web Service request to the remote site and results are pulled back when the service has finished. No data storage at remote site is foreseen due to legal and privacy constraints when working with patients' medical data. User is responsible for data storage at his/her local site.

- **Functionality**

No special functionality for data management has been identified. Applications will upload and unpack data as part of their service.

### 2.2. Details

- **Data movement**

This issue is not currently addressed in the project. It may be examined later.

- **Data storage**

This issue is not currently addressed in the project. It may be examined later.

- **Replica management**

Not addressed in project.

- **Optimization and prediction of data access**

Not addressed in project.

- **Data migration**

Not addressed in project.

- **Metadata management**

Not addressed in project

- **Queries and data search**

Not addressed in project

  - **Query language**

  - **Database access**

### Databases

**Interfaces**

**Security**

**Other**

- **High level data management (ontology, semantics)**

Not Addressed in project

**Concept**

**Definition**

**Language**

### 2.3.External

- **Interfaces**  
The interfaces are application specific and will be developed later in the project.
- **Relations with other components**