



CONTRACT NUMBER 508830

DEISA
**DISTRIBUTED EUROPEAN INFRASTRUCTURE FOR
SUPERCOMPUTING APPLICATIONS**

European Community Sixth Framework Programme
RESEARCH INFRASTRUCTURES
Integrated Infrastructure Initiative

Accounting facilities for DEISA

Deliverable ID: DEISA-DSA5-3.1
Due date : November 24, 2006
Actual delivery date: Month, DD, YYYY
Lead contractor for this deliverable: SARA, Netherlands

Project start date : May 1st, 2004
Duration: 4 years

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	X
RE	Restricted to a group specified by the consortium (including the Commission	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of Content

Table of Content.....	3
1. Introduction.....	4
1.1 Executive Summary.....	4
1.2 References and Applicable Documents	4
1.3 Document Amendment Procedure	4
1.4 List of Acronyms and Abbreviations	4
2. Accounting facilities	6
2.1 Introduction.....	6
2.2 Strategy for general accounting facilities.....	6
2.3 Data provider tool	7
2.4 Data retrieval	10
2.5 Reporting Tools	11
2.6 Content of accounting records.....	11
2.7 Units for charging usage.....	12
2.8 Status	13
2.9 Conclusions	13

1. Introduction

1.1 Executive Summary

In the second project year work started for the collection of accounting information about DEISA users. In an early stage procedures were defined for the IBM Loadleveler sites for the exchange of usage records between sites.

For the general case, including all sites, general facilities have now been developed for the collection and distribution of usage records using a specific common format. These facilities allow every DEISA site to fetch accounting information provided by any batch scheduler, to convert them into the specific format and push them into a local database system. Other sites, users and project managers can retrieve from these databases only those usage records for which they are authorized. Sites can retrieve records from other sites for users or projects that they are responsible for. So it is possible to account for the total usage of users or projects and if a user exceeds the usage of allocated resources by DEISA, further access to the DEISA infrastructure can be blocked by the responsible site. Sites also can import usage records of their home site users into local accounting repositories for long term storage.

1.2 References and Applicable Documents

- [1] <http://www.deisa.org>
- [2] [DEISA deliverable DSA5-2.2 - Implemented improvements of security infrastructure](#)
- [3] <http://www.psc.edu/~lfm/PSC/Grid/UR-WG/UR-WG-Spec-20050925-tracked.pdf>
- [4] <http://exist.sourceforge.net/>
- [5] <http://xmlbeans.apache.org>
- [6] <http://www.globus.org/wsrf/specs/ws-wsrf.pdf>
- [7] <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML>

1.3 Document Amendment Procedure

The initial document amendment procedure is via communication between members of DEISA User Administration workgroup team. The document is then submitted for review to the DEISA Executive and an Executive appointed DEISA reviewer. The document is then amended according to comments received from the Executive and the DEISA appointed reviewer. It is subsequently re-submitted to the DEISA Executive for submission to the EU.

1.4 List of Acronyms and Abbreviations

eXist open source native XML data base [4]

GPFS	General Parallel File System
Home Site	the DEISA partner that is responsible for the administration of a DEISA user. Home sites register locally known DEISA users into the DEISA User Administration System and provide local support for these users.
JDK	Java Software Development Kit
JMX	Specifies a protocol for communication between an MBean and a client
MBean	a Java Bean exposing a JMX compliant management interface
MC-GPFS	Multit-Cluster General Parallel File System
LDAP	Lightweight Directory Access Protocol
LL	Loadleveler, a batch scheduling system from IBM
LSF	A batch scheduling system from Platform Computing Corporation
NQS II	Network Queuing System, a batch scheduling system
OGF	Open Grid Forum
PBSpro	Portable Batch System professional, a batch scheduling system
UR-WG	Usage Record Working Group
WSRF	Web Services Resource Framework
XSD	XML schema definition
X.509	Standard for digital certificates

2. Accounting facilities

2.1 Introduction

For DEISA it is mandatory to have global information on the usage of resources by projects and users of the infrastructure. A DEISA user gets an assignment of resources (like CPU time, memory) in units of the home site of the user (the home site is the computer centre of the partner that is responsible for the administration of the user). The resources consumed by the user typically are on other systems than that of the home site and the usage information has to be exchanged with the home site to enable the control by the home site of the used resources. Also project management needs information on the global usage of resources.

In the second project year an intermediary solution for the exchange of usage records was set up for all sites that use Loadleveler (LL) as batch system (see section 2.4 of DEISA deliverable DSA5-2.2 [2] for a full description). For these sites the format of the records is the same, and the exchange of the usage records was realized simply by exchanging filtered LL *history* files via the DEISA MC-GPFS file systems. LL sites can import the usage records that their home site users have produced on other DEISA sites.

This works well if all the sites are using the same batch scheduler. On the other hand, some necessary information such as the name of the project the user is working for and the subject name of the X.509 certificate can not be provided by this procedure. Also, as the accounting must be extended to DEISA sites using a different batch system a more standardized way of interfacing and distributing all the relevant accounting information has to be employed.

2.2 Strategy for general accounting facilities

In general, DEISA sites are using different schedulers that are producing accounting information in a vendor specific format. The formats of usage records produced by the LL, LSF, PBSpro or NQS II schedulers, all in use today by DEISA partners, all differ between each other.

An OGF working group, the UR-WG, has developed a widely accepted recommendation for a format specification for usage records [3]. This format specification, currently version 1.1 that is expected to become a standard, has been adopted by DEISA. Based on this decision, DEISA follows a two-step approach for *collecting* and *distributing* global accounting information.

In the first step, each site extracts all the DEISA related usage records from the vendor specific accounting log files, and converts them into OGF UR-WG XML formatted documents. The information that can not be provided by the batch system (e.g., project name and X.509 subject name) is taken from alternative information sources and added to the XML documents. The XML documents are stored into a local database (see Figure 1).

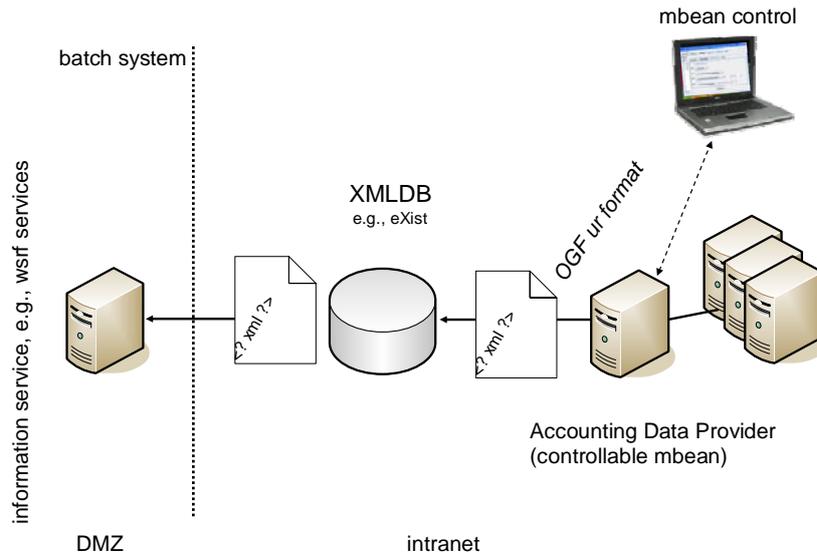


Figure 1 configuration of an accounting data provider at a single site

In the second step, entities such as users, project managers or site accounting managers query the distributed accounting information via secured information services located at every DEISA site. The queries can be performed by means of a WSRF client tool. The WSRF based information services are accessible only for authorized entities (See Figure 2).

Access to the data can be limited using authorization services based on X.509 certificates. Accounting data can be used to trace the activities of a single individual and EU rules and national laws from countries of participating partners protect the distribution of this kind of information. For instance see the EU directive [95/46/EC \[7\]](#) on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

So it is important that access to the data can be limited as the privacy of the data must be carefully protected. We must comply with the rules set by the different national laws and EU rules.

Details of the different parts of the DEISA accounting system are described in the next paragraphs.

2.3 Data provider tool

The accounting data provider is a service developed within DEISA that extracts usage records from a specific batch system, converting them according to the OGF UR-WG XML schema recommendation. It is implemented in Java and uses the MBean technology.

The MBean server provides a container (MBeanContainer) which can execute various service beans. Figure 1 shows the deployment of the accounting data provider at one site. The design of the corresponding service bean, the AcctDataPushService and associated classes, is shown in Figure 3.

Every DEISA site runs an MBean server. This server is configurable (figure 5) and the software is designed to facilitate the development of implementations for other batch systems or to implement site-specific utilities for accessing additional information sources.

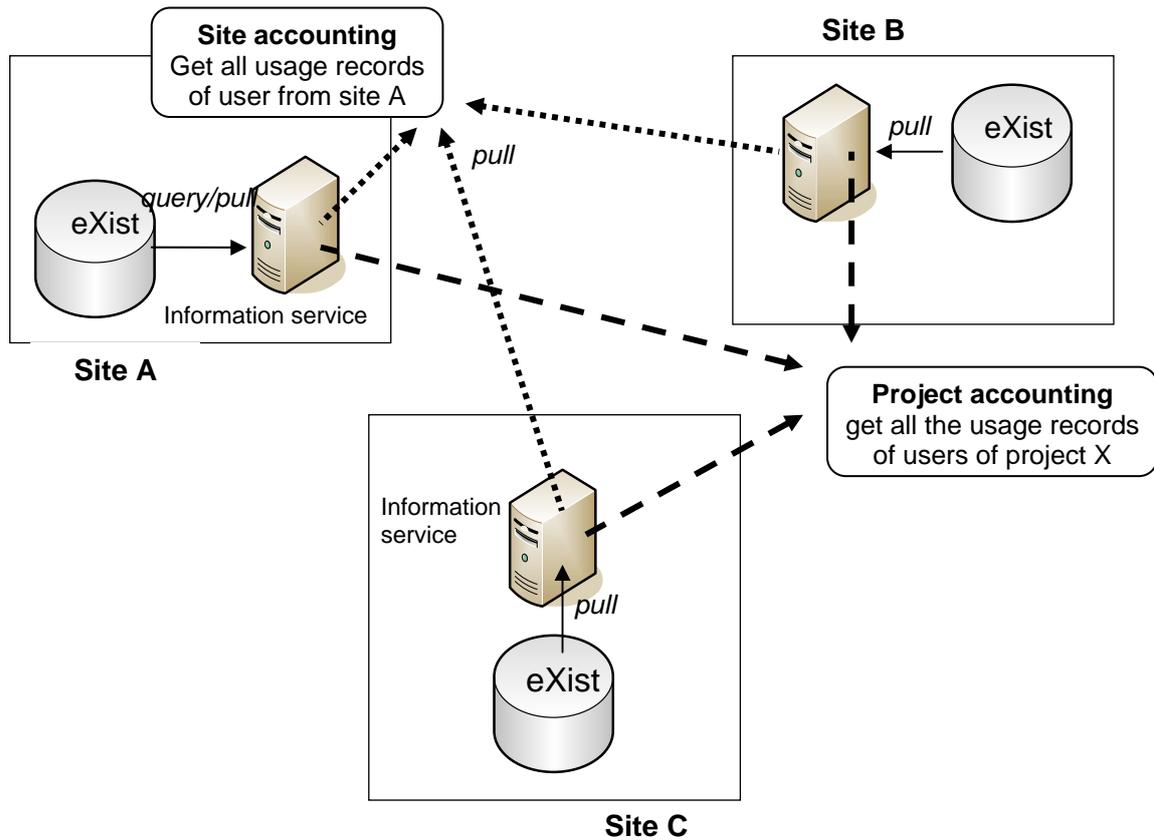


Figure 2 distributed accounting information services

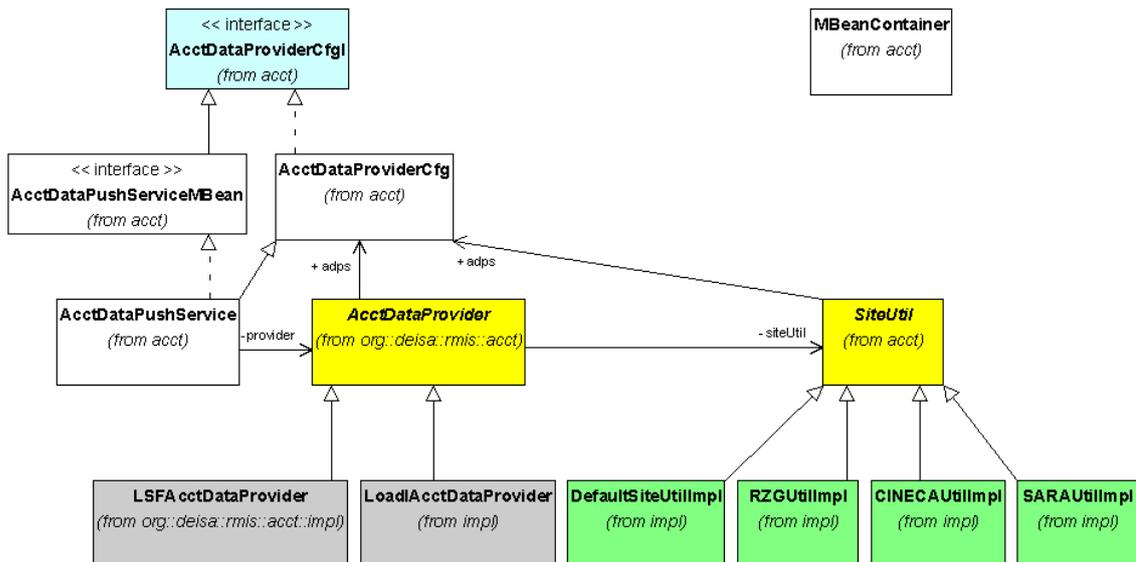


Figure 3 UML class diagram of the accounting data provider software. `AcctDataProvider` and `SiteUtil` are abstract classes. The implementation of the first is adapted to the specific batch systems. Implementations of `SiteUtil` provide additional, site-specific access procedures.

The implementation of the abstract Java class `AcctDataProvider` must be adapted to the local batch system. The abstract class provides several methods for creating XML elements according to the OGF UR-WG recommendation based on the

xmlbeans API [5]. Implementations of *AcctDataProvider* have already been developed for Loadleveler by RZG and CINECA and for LSF by SARA. In addition, each site has to adapt for some site specific data access procedures and this is achieved by implementing the abstract class *SiteUtil*. With this class the values for the user's X.509 subject name and the project name are added to the XML documents. Implementations for SARA, CINECA and RZG are already available and can be used as templates for other sites. *RZGUtilImpl* is adapted to the user administration MYSQL data base as used at RZG. The two other implementations follow a more generic approach and extract the missing information from a text file that is created from the DEISA LDAP based administration system.

An MBean server can be controlled (Figure 4) and monitored (figure 5) by means of an external client (Figure 1), e.g. *jconsole*, a tool that is included in Sun Java Development Kit, JDK 5.

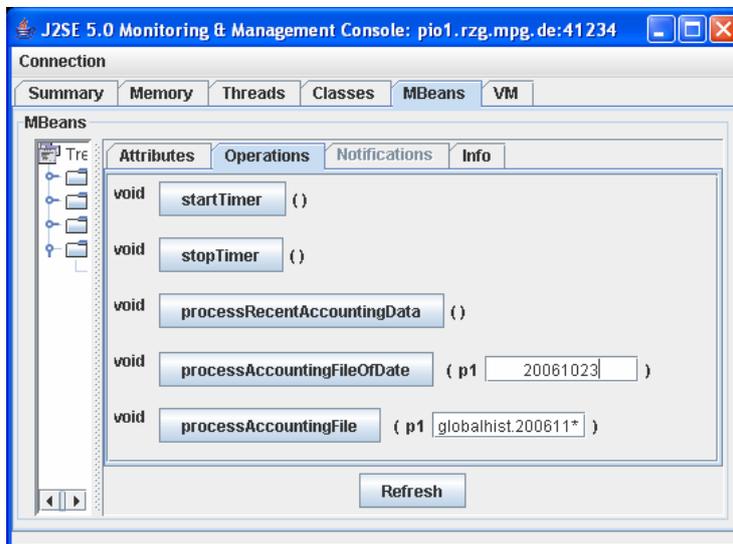


Figure 4 *jconsole*. Control interface for starting and stopping the MBean. If necessary, single or multiple proprietary accounting log files of the batch system can be processed repeatedly.

Each site is running a database where the local DEISA usage records are stored. Currently, the Open Source native XML database *eXist* [4] is used for storing the XML documents. The *eXist* database should not be considered as a long-term repository, but usage records should at least be available over a time period of several months. With the attribute *TimeInterval* the interval with which the database is updated can be configured (see Figure 5). It is required that the database is updated at least every 24 hours.

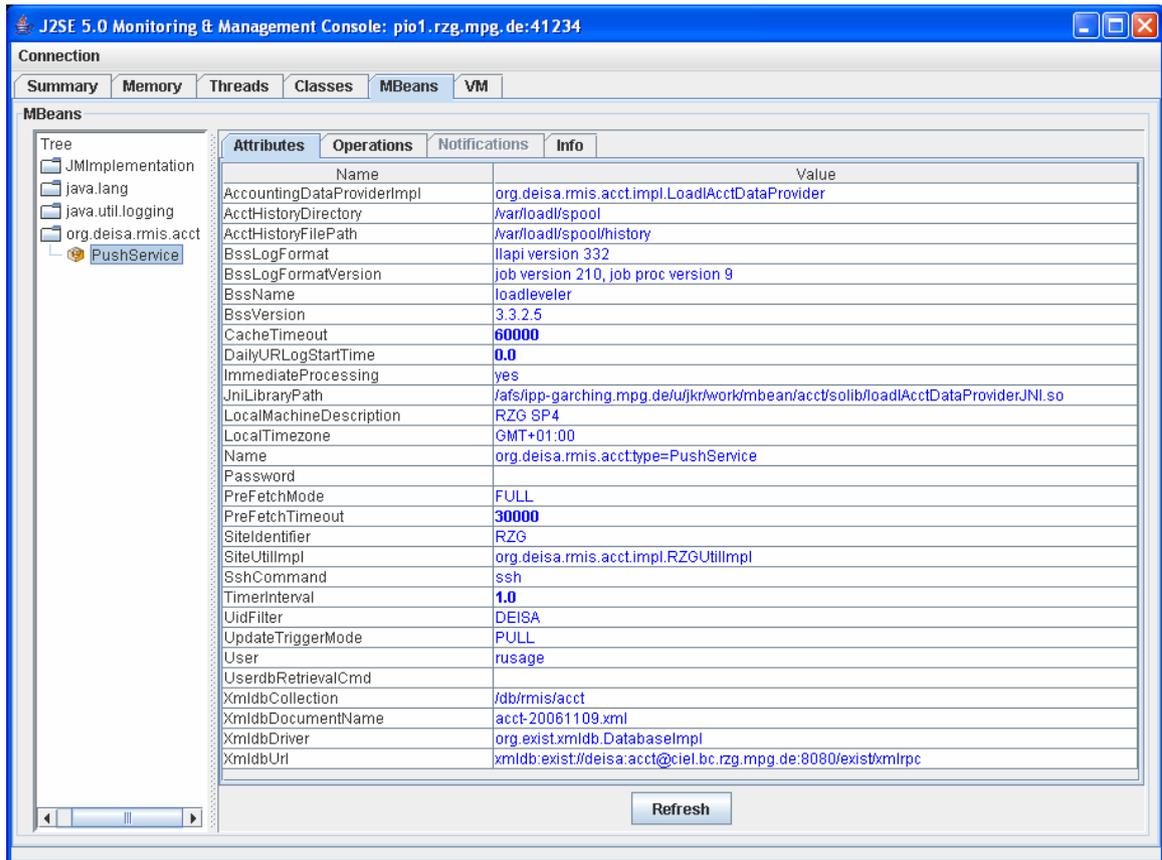


Figure 5 : jconsole interface for monitoring the accounting data provider. The attributes shown are either required to configure the MBean server (e.g., jniLibraryPath, TimerInterval) or they are representing the current state of the server (e.g., BssVersion, AcctHistoryFilePath).

A site can use other tools to store the usage records in the local eXist database. The only requirement is that the records have the standard format used by all sites.

2.4 Data retrieval

For the retrieval of data from the eXist database, WSRF based information provider has been developed. WSRF defines the framework for interoperability of grid services; see “The WS-Resource Framework” [6] for more information.

Each DEISA site runs the WSRF based information service and other sites can extract information for updating local accounting repositories or producing reports. If the information service receives a request it will access the eXist DB to retrieve the requested information and send it to the requestor, see Figure 2.

Each information service aggregates and provides the accounting information from their local repository according to specific views:

- **User view:** users can query accounting information related to their own jobs.
- **Project view:** registered managers of user projects can get a report about the usage of DEISA resources by project group members.
- **Management view:** registered DEISA managers, e.g. the ATASKF team leader, can request for a report on the global usage of DEISA resources by all the projects.
- **Site view:** in order to archive the usage records of their users, home sites can fetch all the usage records related to their home site users in XML format.

The privacy and integrity of the data is guaranteed using encrypted communication channels and digital signatures.

2.5 Reporting Tools

A Perl script has been developed with which information can be extracted directly from the eXist DB. Reports can be made for a chosen period for the number of jobs, the total wall time used, and the total CPU time used.

An example output with names made anonymous is shown below:

DEISA usage report for year 2006 and month 10

Project	Nr Jobs	WallDuration (secs)	CpuDuration (secs)
Project1	64	2618.00	2137.85
Project2	13	35.00	10.63
Project3	71	13511068.00	3524281.42
Project4	7	43.00	20.54
Project5	32	34509.00	31540.13

User	Nr Jobs	WallDuration (secs)	CpuDuration (secs)
User1	7	17.00	6.34
User2	14	132.00	8.31
User3	1	3.00	1.41
User4	71	13511068.00	3524281.42
User5	1	4.00	1.68
User5	7	43.00	20.54
User6	2	4.00	1.13
User7	14	2269.00	2096.26
User8	2	129.00	1.54
User9	20	147.00	13.56
User10	3	14.00	3.94
User11	1	2.00	0.94
User12	4	8.00	1.97
User13	7	26.00	7.22
User14	9	10.00	2.81
User15	1	22.00	0.36
User16	3	748.00	3.50
User17	11	40.00	16.48
User18	4	47.00	4.03
User19	1	1.00	0.35
User20	1	3.00	0.80
User21	3	33536.00	31516.00

Total	Nr Jobs	WallDuration (secs)	CpuDuration (secs)
	187	13548273.00	3557990.58

This Perl script can easily be adapted to extract from remote sites the information it is authorized for, using a certificate, from the WSRF tool described above.

2.6 Content of accounting records

In the GGF UR recommendation [3] 26 base properties are defined. Two properties are mandatory, the record identity which defines the record in a unique way and the status property which specifies the status of the job (aborted, completed etc.). With the extension property additional data can be specified that is not covered by the

other properties. The OGF UR-WG recommendation addresses the properties of records on a job level. The format definitions are given in XML and free format style. Within DEISA not all properties are used. Currently we deploy the list given in Table 1. The left column gives the element names as used in the specification and in the right column the properties used by DEISA are described (some elements define more than one property).

UR-WG Element name	Description
RecordIdentity	Identifies uniquely the usage record
JobIdentity	Contains local job identifier (LocalJobId) as assigned by the batch queue and a GlobalJobId (may be LocalJobId with a sitename prefixed)
UserIdentity	The username the job has run under (LocalUserId) and the Subject name of the X.509 cert (Keyinfo)
JobName	The global job name
Status	Completion status of job, e.g. completed, aborted.
WallDuration	Total wall clock time that elapsed while the job was running.
CpuDuration	Total CPU time used, summed over all processes of the job
MachineName	A descriptive name of the system on which the job ran
Host	The system hostname on which the job ran
SubmitHost	The system hostname from which the job was submitted
ProjectName	The name of the project that the job was run under
Processors	The number of processors used or requested (reserved)
EndTime	The time at which the usage ended
StartTime	The time at which usage started
NodeCount	The number of nodes used
SubmitTime	Not a UR-WG defined property. It gives the time the job is submitted to the system the job has run on.

Table 1 - Usage Record properties used by DEISA.

DEISA always can include more properties. The important thing is that there is agreement on the minimum number of properties. The **RecordIdentity** property must be unique. Currently we use a hash of the usage record (without the RecordIdentity property). We have defined one additional property, **SubmitTime**, which currently is not part of the UR-WG recommendation. We will propose to include this property in a future version of the recommendation.

2.7 Units for charging usage

Each DEISA site has its own method for charging the usage of their system. Different parameters can be used and also the weight of the parameters can differ. And even if the same formula would be used the result will differ due to differences in system performance. For instance I/O and CPU performance will influence the wall clock time measured at different sites for the same program.

Before site B can calculate the charge in their units for a job that has run at site A there must be agreement on the conversion to apply to account for the differences in system performance. Currently we use numbers based on some benchmark programs. The numbers in use are relative to a IBM Power4 CPU running at 1.7

MHZ. Sites always can propose another value for their particular situation, as can be the case for instance for a vector processor system.

A list of systems with their corresponding conversion numbers is published in an XML schema on the DEISA internal website and this information is used if reports are produced or if records are imported to other repositories with accounting data. The **MachineName** property from the usage records is used to look up the number that must be used.

2.8 Status

At the end of October 2006 3 test sites, CINECA, RZG, and SARA are running the new accounting facilities. Other partners will install the tools in November and December of 2006 or early 2007. By that time all information about DEISA usage will be available.

The tools for reporting will be further improved. Each site also will have to develop tools for importing usage records from remote sites into their local accounting repositories.

Users must also be able to query their personal use. It's yet not implemented that they can do this directly from the eXist databases and they will need assistance of their site to receive their personal information. But with the current facilities it will be possible to implement the functionality that they can query their personal usage.

2.9 Conclusions

Facilities have been developed for storing usage records in a common format following the UR-WG recommendation. Using this format also enables the interoperability with other grid infrastructures using the same format.

The data is stored locally at a site and authorised access can be managed by each site.

A strong requirement is the security of the data, on the one hand the data must be well protected because of privacy reasons, and on the other hand access must be enabled for those that need the information. With the facilities that we have developed we can manage the authorisation for accessing the data on a fine grained scale.

The set up is modular and most facilities can be replaced without disturbing the service and facilities can be adapted to local needs.