Accounting Facilities and Policies used by DEISA

Johannes Reetz

RZG, johannes.reetz@rzg.mpg.de

MLA Workshop, HPDC 2009,

Munich, Germany

June 10, 2009

www.deisa.eu

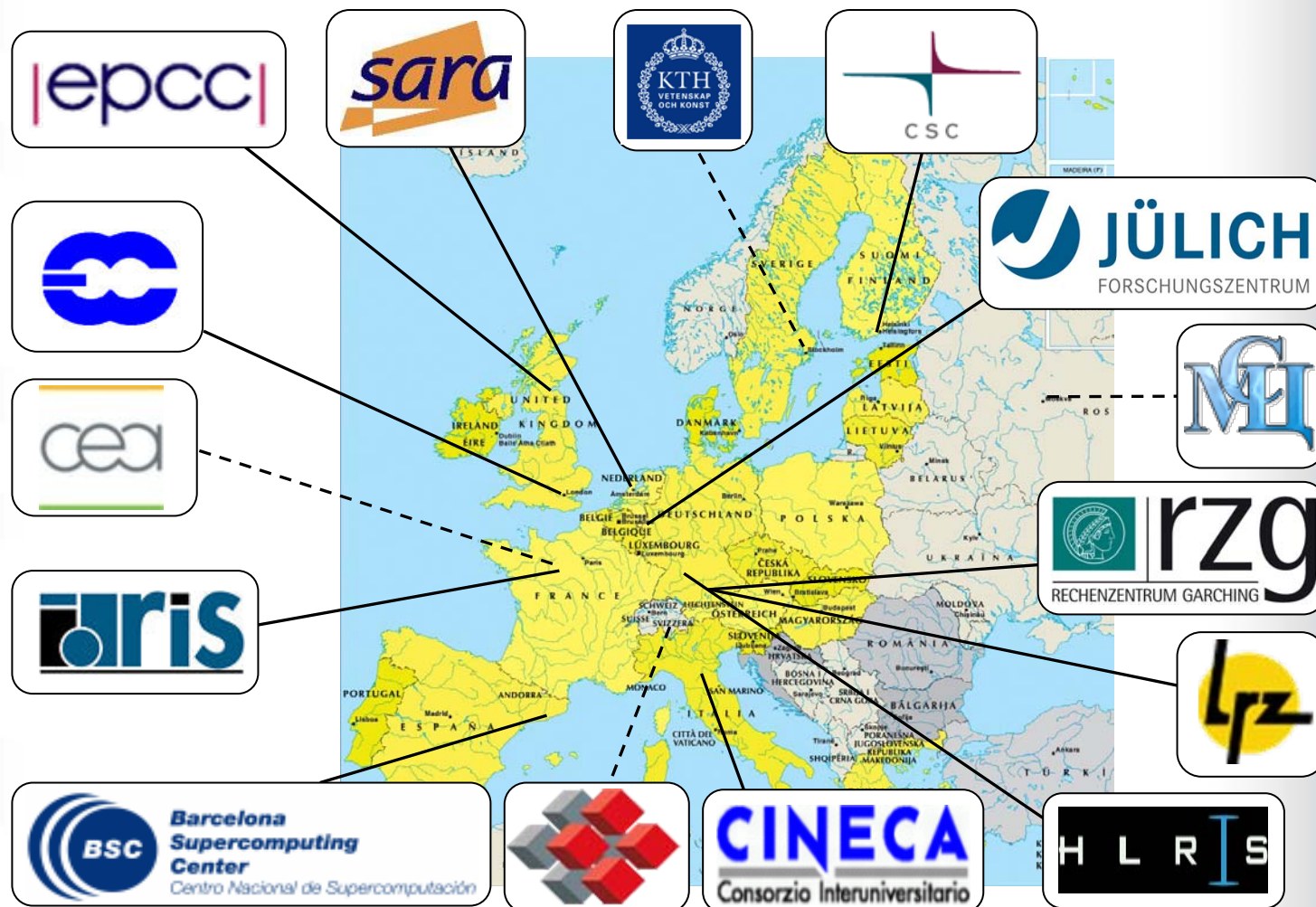# Agenda

⇨ DEISA – HPC Resources, Projects, Accounting Requirements

⇨ User administration: Projects, Users, Budgets

⇨ Accounting facilities

- Site-local accounting tools (Accounting Data Provider)

- Utilization of the OGF Usage Record format v1.0

- Accounting Information Service

- Accounting Reporting Tool (DART)

⇨ Procedures and Policies

⇨ Lessons learned and future Improvements

RI-222919

Agenda

⇨ **DEISA** – HPC Resources, Projects, Accounting Requirements

⇨ User administration: Projects, Users, Budgets

⇨ Accounting facilities

   ▪ Site-local accounting tools (Accounting Data Provider)

   ▪ Utilization of the OGF Usage Record format v1.0

   ▪ Accounting Information Service

   ▪ Accounting Reporting Tool (DART)

⇨ Procedures and Policies

⇨ Lessons learned and future Improvements

RI-222919

DEISA consortium

Principal partners   and associate partners

RI-222919

DEISA HPC compute resources

Site	Architecture	CPU@Clock	Tf/s	#Cores	´Memory
BSC	IBM PowerPC	PPC@2.3	94.2	10240	20.0
CINECA	IBM Blade Center	Opt. DC@2.6	26.6	5120	10.0
	IBM Power5-575	P5@1.9	3.9	512	1.0
CSC	Cray XT4 AMD 4-core	QC@2.3	32.3	4048	4.4
	Cray XT5 AMD 8-core		54.4	6816	6.1
ECMWF	IBM P5-575+	P4+@1.9	33.0	4552	
EPCC	Cray XT4	DC@2.8	59.0	11328	33.2
	Cray X2	X2	2.9	112	0.9
	IBM Power5-575	P5@1.9	15.4	2560	5.2
FZJ	BlueGene/P	PPC450+@0.85	223.0	65536	32.0
	IBM Power6-575	P6@4.7	8.4	448	1.8
HLRS	NEC-SX8	SX-8@2.0	12.7	576	9.2
IDRIS	BlueGene/P	PPC450+@0.85	139.0	40960	20.0
	IBM Power6-575	P6@4.7	67.3	3584	10.8
LRZ	SGI-Altix Itanium2	I2 DC@1.6	62.3	9728	39.0
RZG	BlueGene/P	PPC450+@0.85	54.0	16384	4.0
	IBM Power6-575	P6@4.7	120.0	6560	18.5
SARA	IBM Power6-575	P6@4.7	60.2	3328	15.6

Status: Apr 2009

RI-222919

Projects and Science Communities

DECI call 2005

29 proposals accepted   12 mio core-h granted*

DECI call 2006

23 proposals accepted   12 mio core-h granted*

DECI call 2007

45 proposals accepted   30 mio core-h granted*

DECI call and Science Communities 2008

42 proposals accepted   50 mio core-h granted*

3 virtual communities   7 mio core-h granted*

DECI call and Science Communities 2009

75 proposals   >200 mio core-h requested*

>5 virtual communities

---

*) Core-h normalized to IBM P4+@1.6GHz

DECI: **D**EISA **E**xtreme **C**omputing **I**nitiative

Yearly call for proposals

Communites: Virtual Scientific Communtities

Usage of DEISA resources

DEISA user
SSH / GSISSH client
UNICORE   client
DESHL   client
WS-GRAM   client

Distributed
European
Infrastructure for
Supercomputing
Applications

JÜLICH FORSCHUNGSZENTRUM

IDRIS

HLRS

BlueGene/P   IBM P6   IBM P6   BlueGene/P

NEC SX8

LL   LL-MC   LL-MC   LL

NQS II

epcc

IBM P6   IBM P5

LL

PBS Pro   Cray XT4   LL

Cray XT4/5   PBS Pro   SGI ALTIX   PBS Pro   LRZ

CSC

IBM P5   IBM P6   BlueGene/P

LL-MC   LL-MC   LL

CINECA
Consorzio Interuniversitario

Maui/Slurm   LL

IBM PPC   IBM P6

BSC   sara

rzg
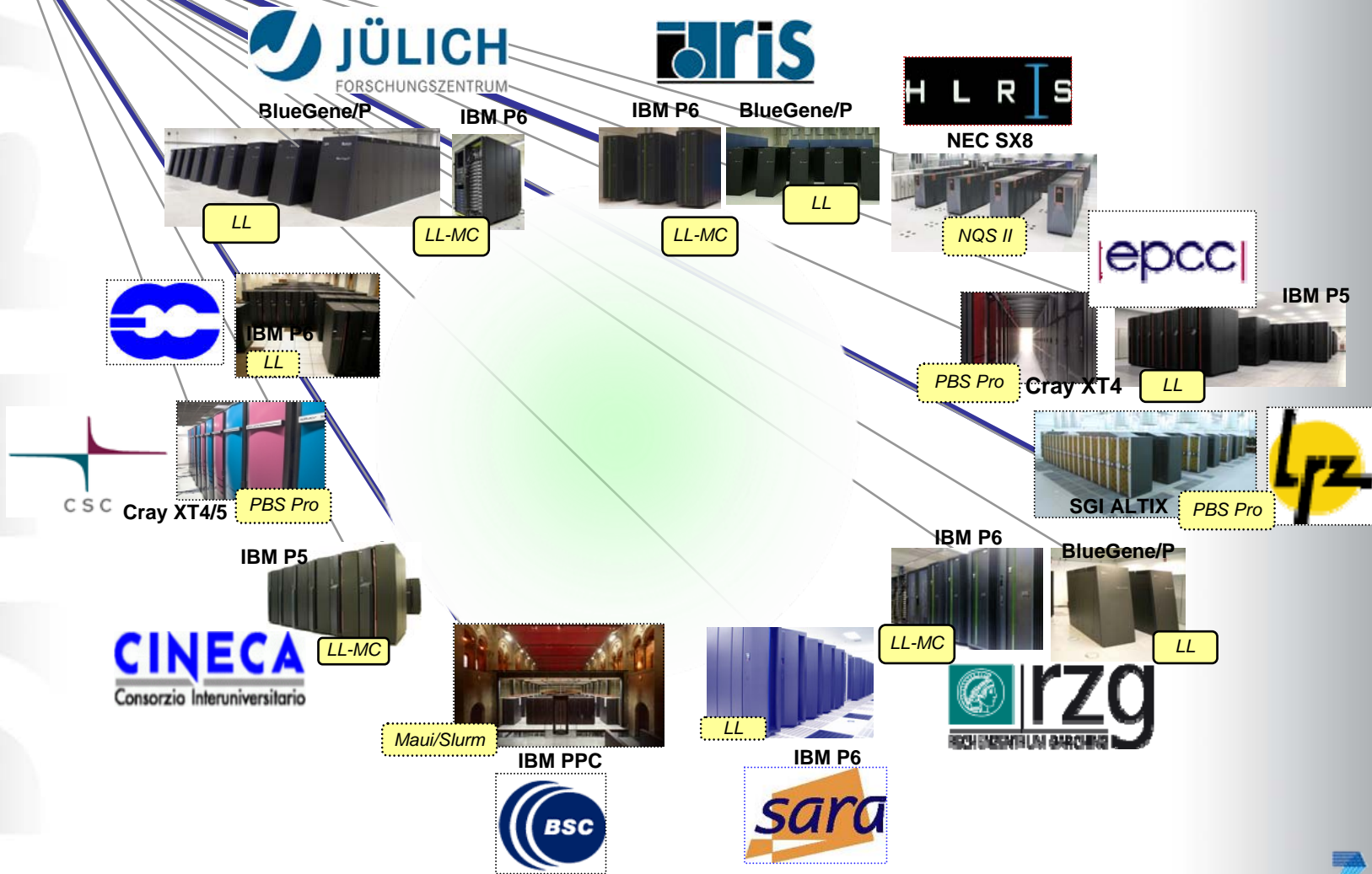
RI-222919

LL remote job submission

DEISA user
SSH / GSISSH client
UNICORE   client
DESHL   client
WS-GRAM   client

Distributed
European
Infrastructure for
Supercomputing
Applications

JÜLICH FORSCHUNGSZENTRUM

IDRIS

HLRS

BlueGene/P   IBM P6   IBM P6   BlueGene/P

NEC SX8

LL   LL-MC   LL-MC   LL

NQS II

epcc

IBM P6   IBM P5

LL

PBS Pro   Cray XT4   LL

IBM P5

Cray XT4/5   PBS Pro

SGI ALTIX   PBS Pro

LRZ

IBM P6   BlueGene/P

LL-MC   LL-MC   LL

CINECA
Consorzio Interuniversitario

rzg

Maui/Slurm   LL

IBM PPC   IBM P6

BSC   sara

RI-222919

Collecting usage records

RI-222919
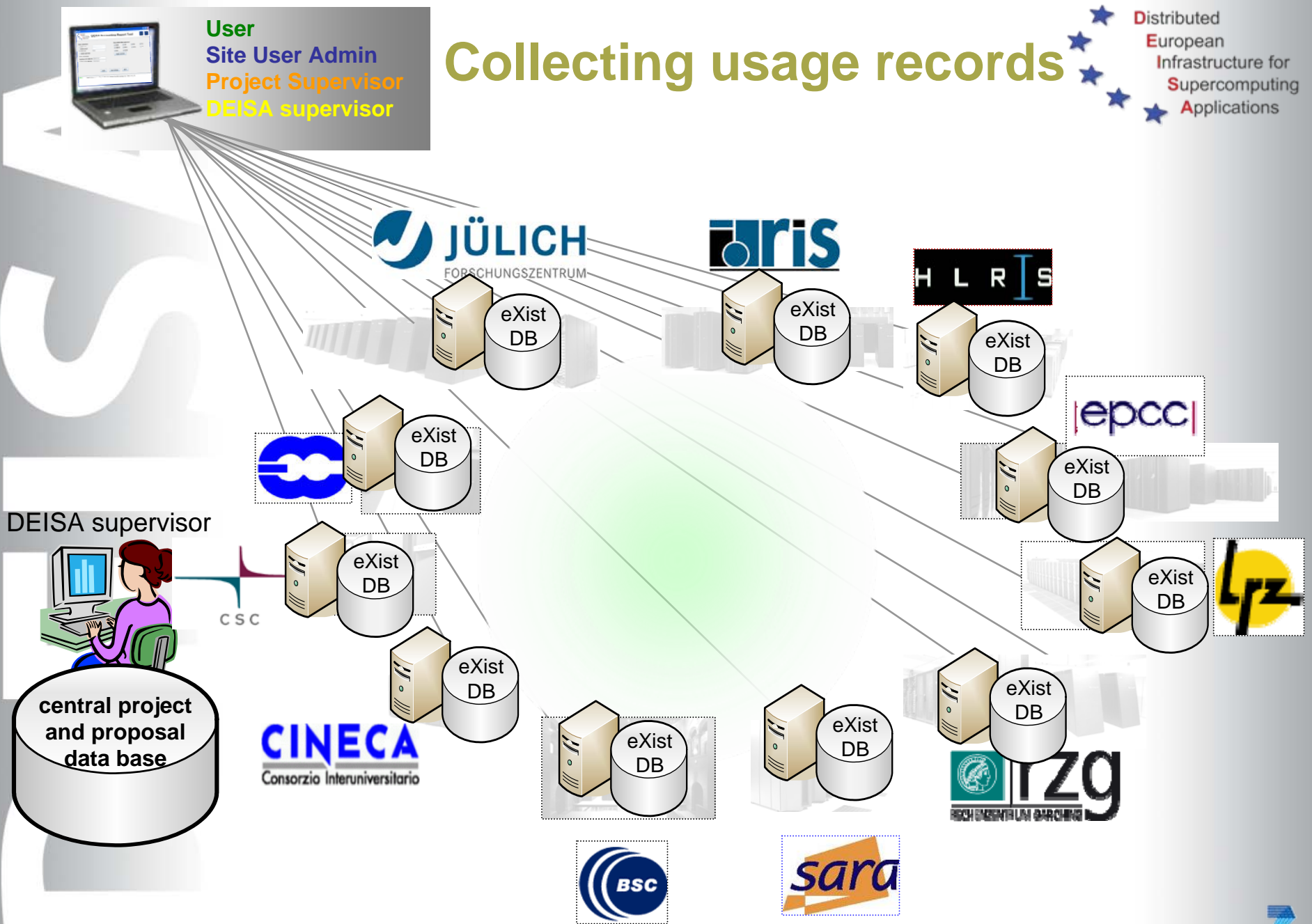
Collecting URs via Web interface

RI-222919

Results obtained via Web interface

Agenda

⇨ DEISA – HPC Resources, Projects, Accounting Requirements

⇨ **User administration: Projects, Users, Budgets**

⇨ Accounting facilities

- Site-local accounting tools (Accounting Data Provider)

- Utilization of the OGF Usage Record format v1.0

- Accounting Information Service

- Accounting Reporting Tool (DART)

⇨ Procedures and Policies

⇨ Lessons learned and future Improvements

RI-222919

User and Project Administration Concept

1. Each user has an administrative DEISA **Home Site**

2. Each Project (or community) is administrated by an **Home Site**

3. A project is mapped to one more **Execution Sites**

4. The budget of compute cycles granted to a project is assigned to one or more **Execution Sites** for a given period of time.
→ the Execution Sites know about the budget of allocated core cycles

5. The consumed computing resources are collected for each project monthly to be registered into the centralized **project and proposal data base** for project controlling purposes.

RI-222919

Projects and Proposal Data Base

RI-222919

Projects and Proposal Data Base (2)

Contains information about

- project acronym and unique project identifier

- project membership

- project supervisors

- project execution machines and corresponding budgets

- core cycles (job time) per project assigned to execution machine

- CPU normalization factors

RI-222919

DEISA User Administration System (UAS)
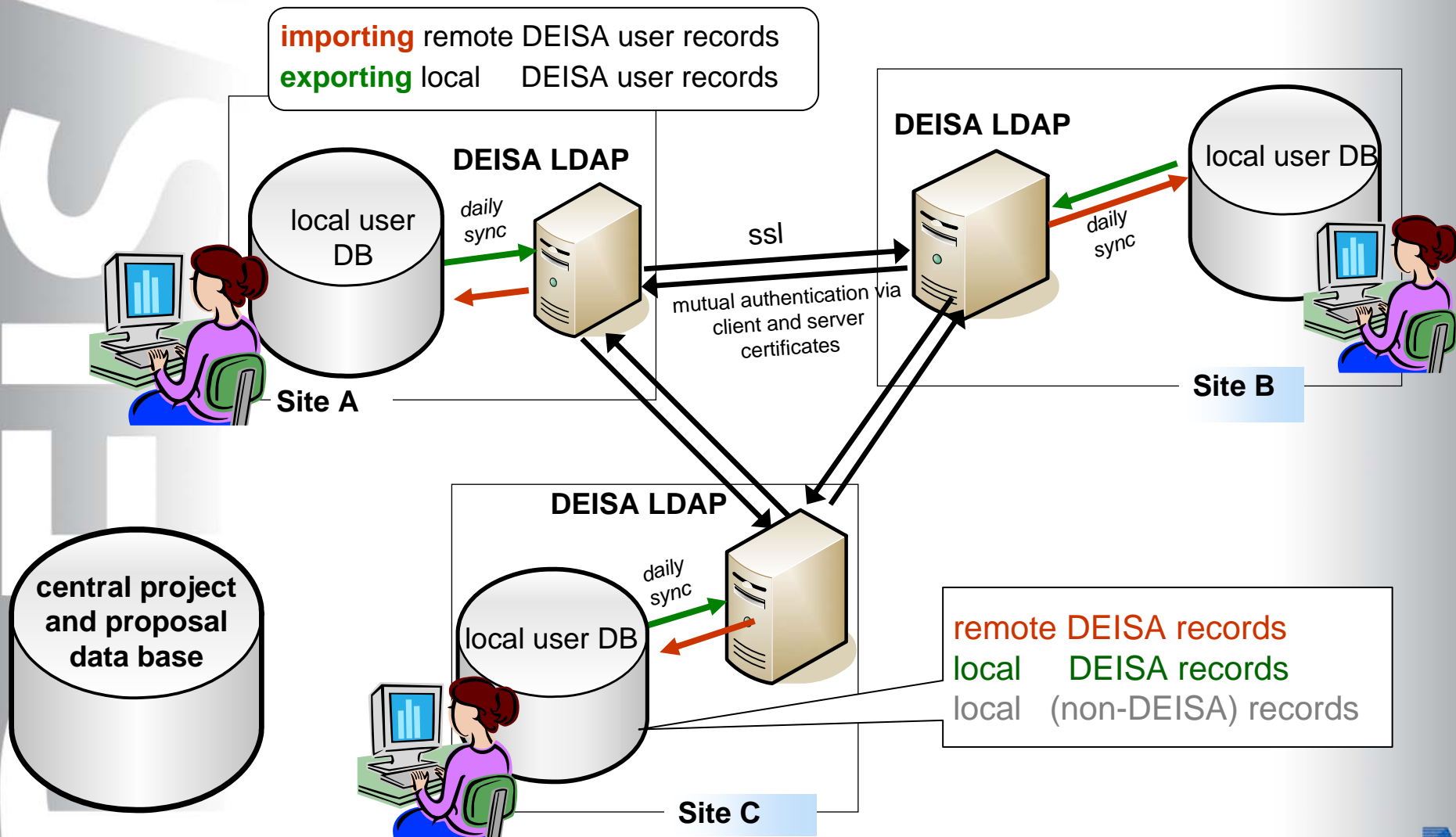
importing remote DEISA user records
exporting local   DEISA user records

DEISA LDAP

local user DB

local user DB

daily sync

daily sync

ssl

mutual authentication via client and server certificates

DEISA LDAP

Site A

Site B

DEISA LDAP

daily sync

local user DB

central project and proposal data base

remote DEISA records
local   DEISA records
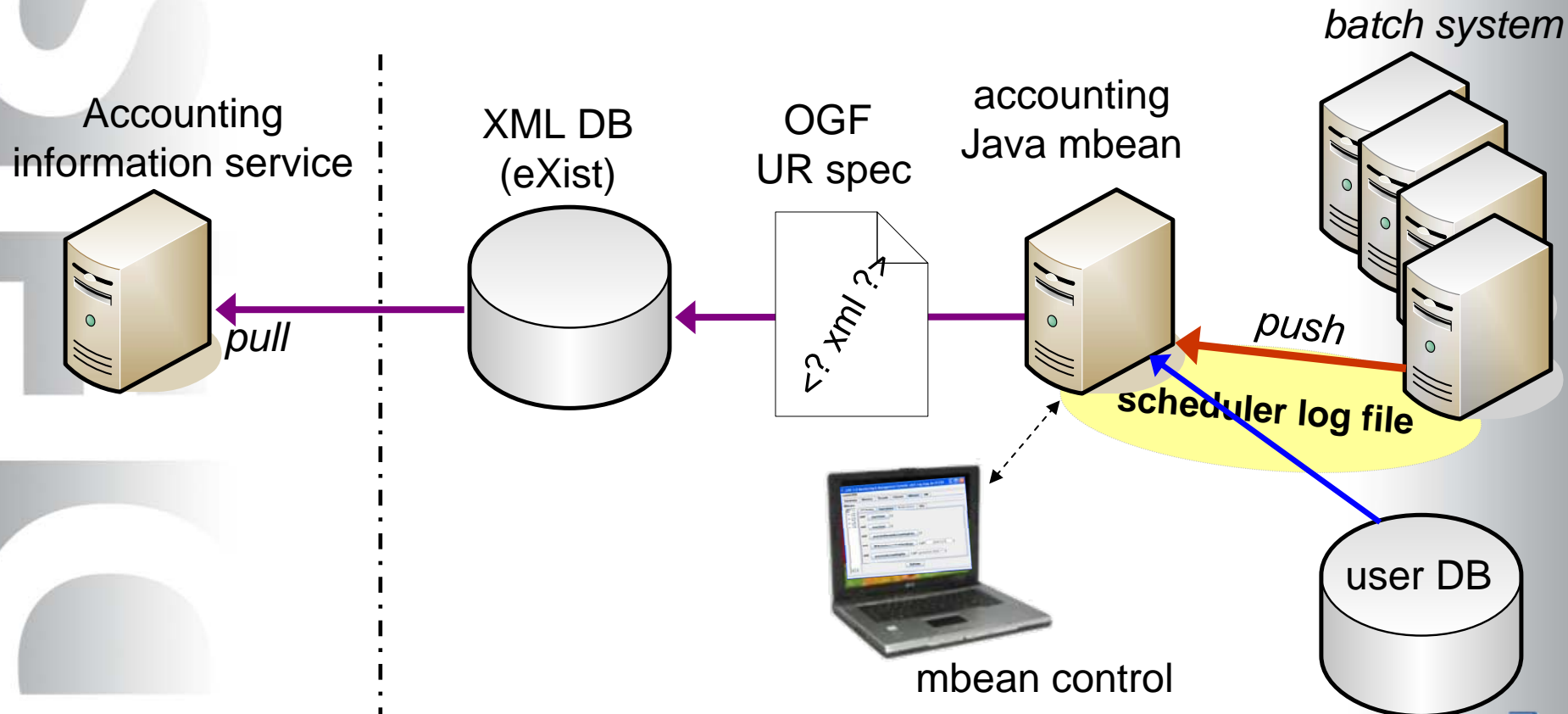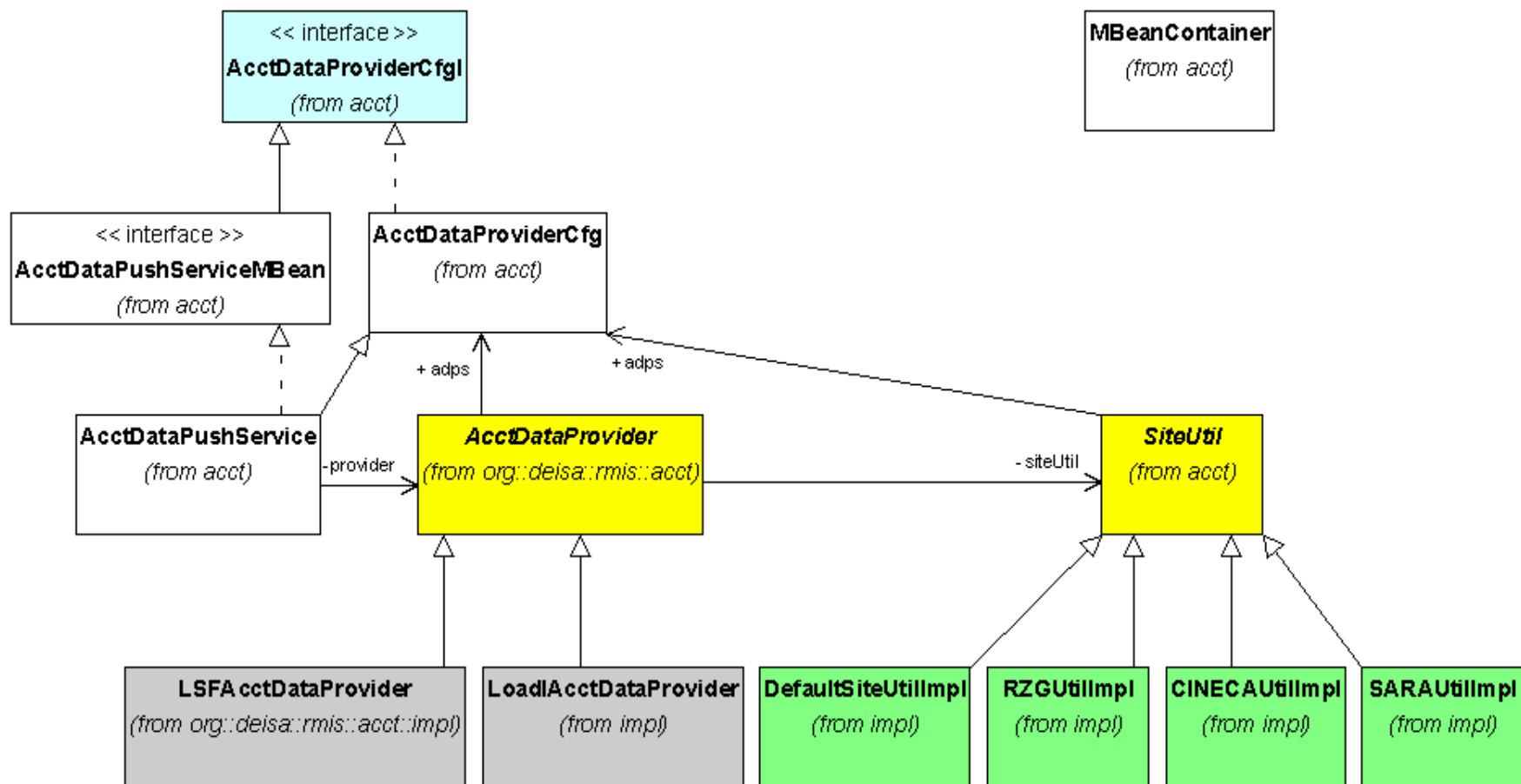local   (non-DEISA) records

Site C

RI-222919

Agenda

⇨ DEISA – HPC Resources, Projects, Accounting Requirements

⇨ User administration: Projects, Users, Budgets

⇨ Accounting facilities

- Site-local accounting tools (Accounting Data Provider)

- Utilization of the OGF Usage Record format v1.0

- Accounting Information Service

- Accounting Reporting Tool (DART)

⇨ Procedures and Policies

⇨ Lessons learned and future Improvements

RI-222919

Accounting Data Provider Tool (1)

- implemented in Java
- analyses the batch scheduler log files (e.g. LoadLeveler history file)
- adds missing information from local user DB or directly from DEISA ldap)
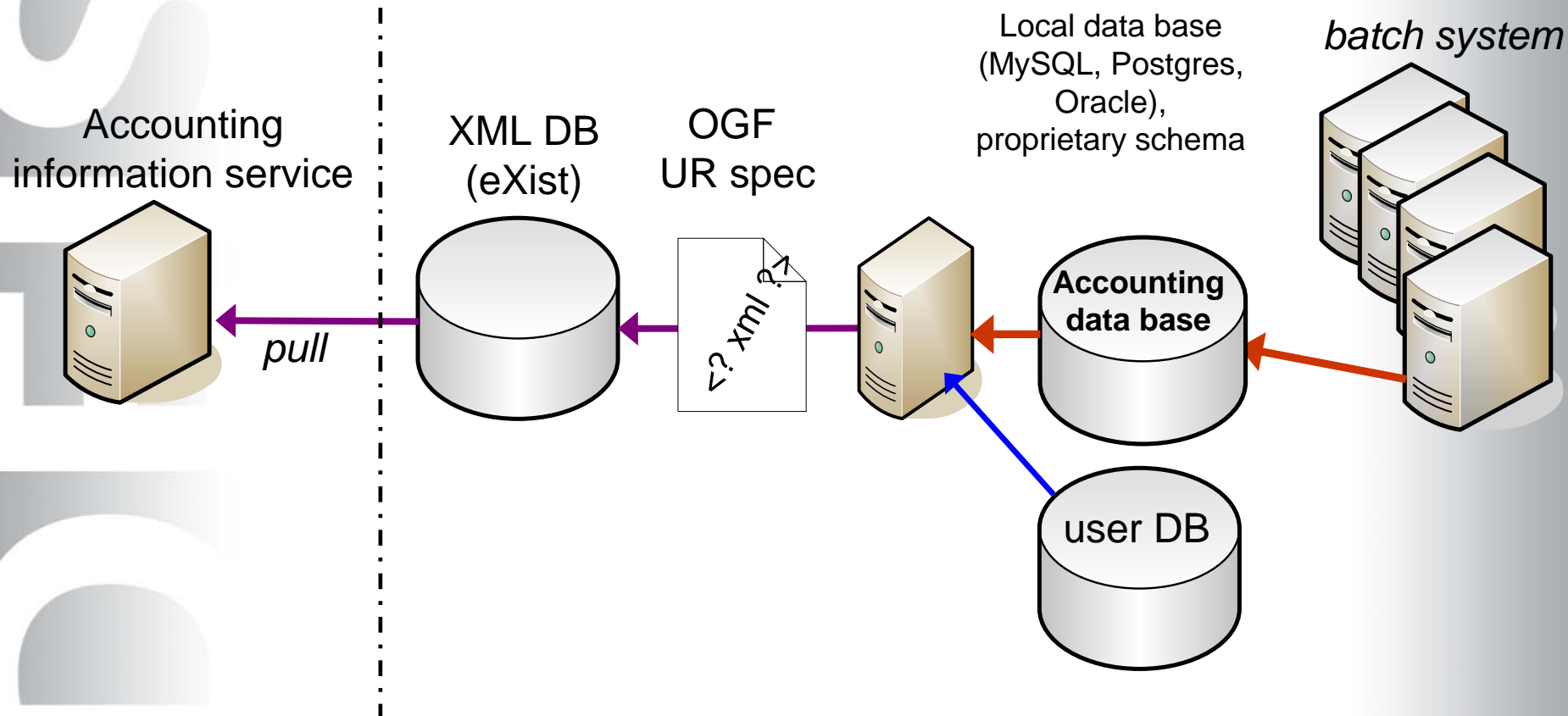- the result are XML records which are stored in eXist DB

batch system

Accounting information service

pull

XML DB (eXist)

OGF UR spec

`<?xml ?>`

accounting Java mbean

push

scheduler log file

mbean control

user DB

RI-222919

Accounting Data Provider Mbean design

RI-222919

Accounting Data Provider Tool (2)

- local accounting data base, stored information from batch system log file
- a second procedure (e.g. Perl script) creates XML usage records, missing information comes from user DB (or from DEISA ldap directly)
- XML records are stored in eXist DB

Local data base
(MySQL, Postgres,
Oracle),
proprietary schema

batch system

Accounting
information service

XML DB
(eXist)

OGF
UR spec

<?xml ?>

**Accounting
data base**

pull

user DB

UR Properties used by DEISA

| 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.

UR Properties used by DEISA

Not all attributes come from log files of the batch system

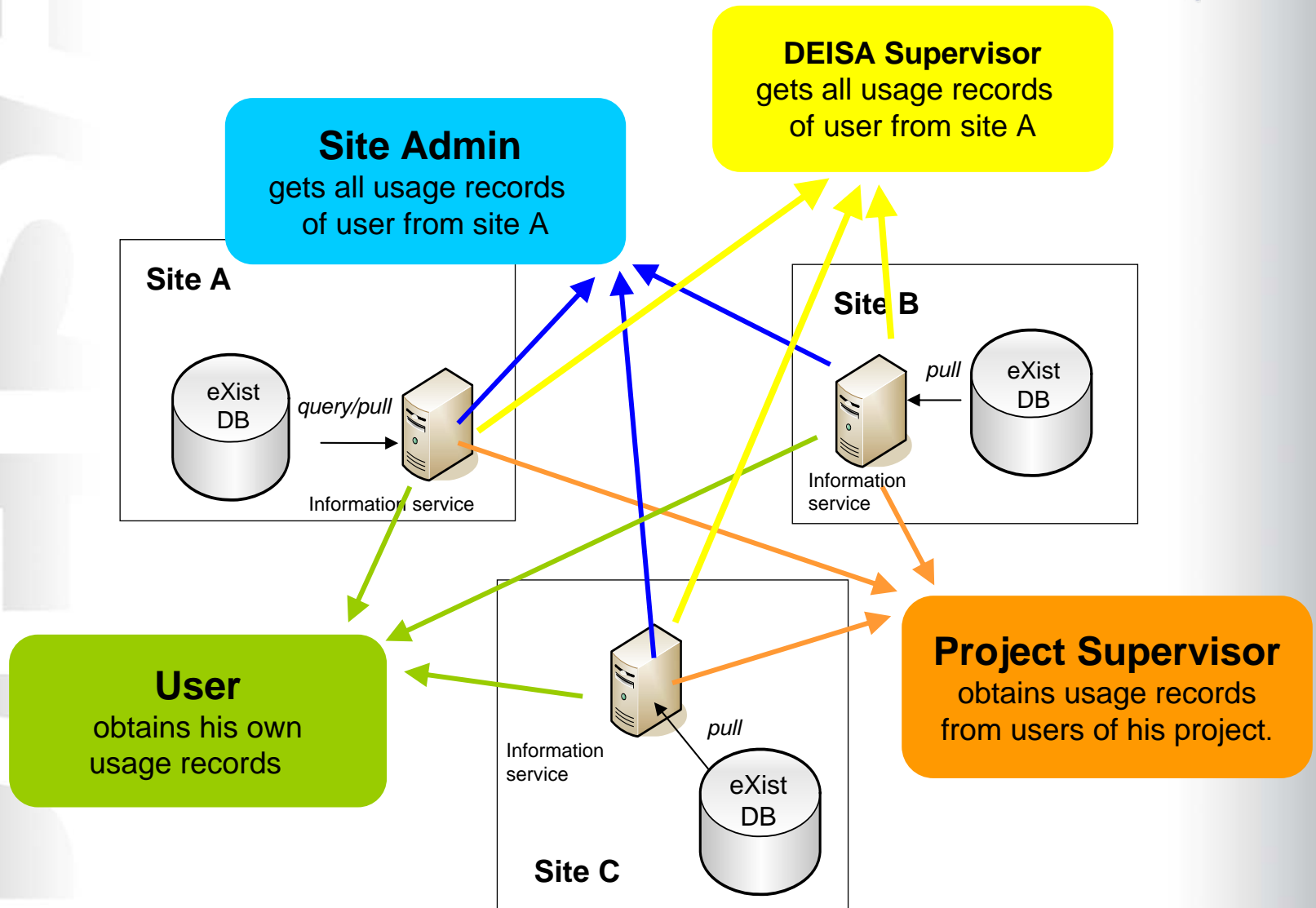- Key info (X.509 Subject name)
- Project name
- Machine name

→ additional information from local user DB or DEISA LDAP

- Key info and project name can be retrieved from user DB
- Machine name from local file

RI-222919

Accounting Information Services

RI-222919
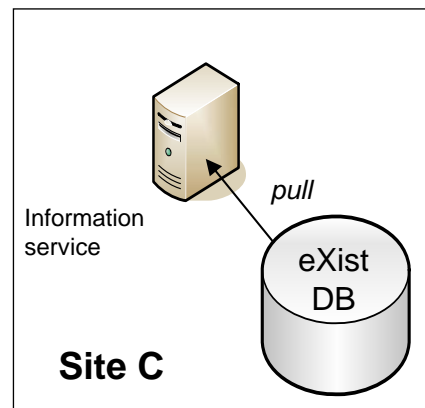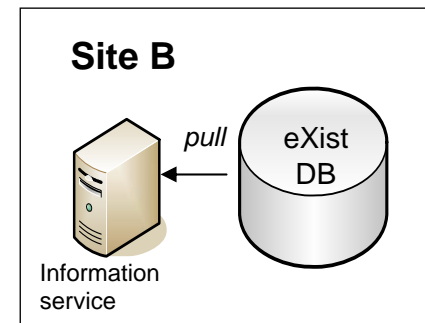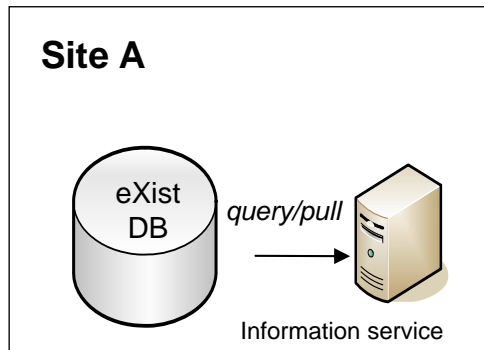
Information Requestors

RI-222919

Accounting Information Service
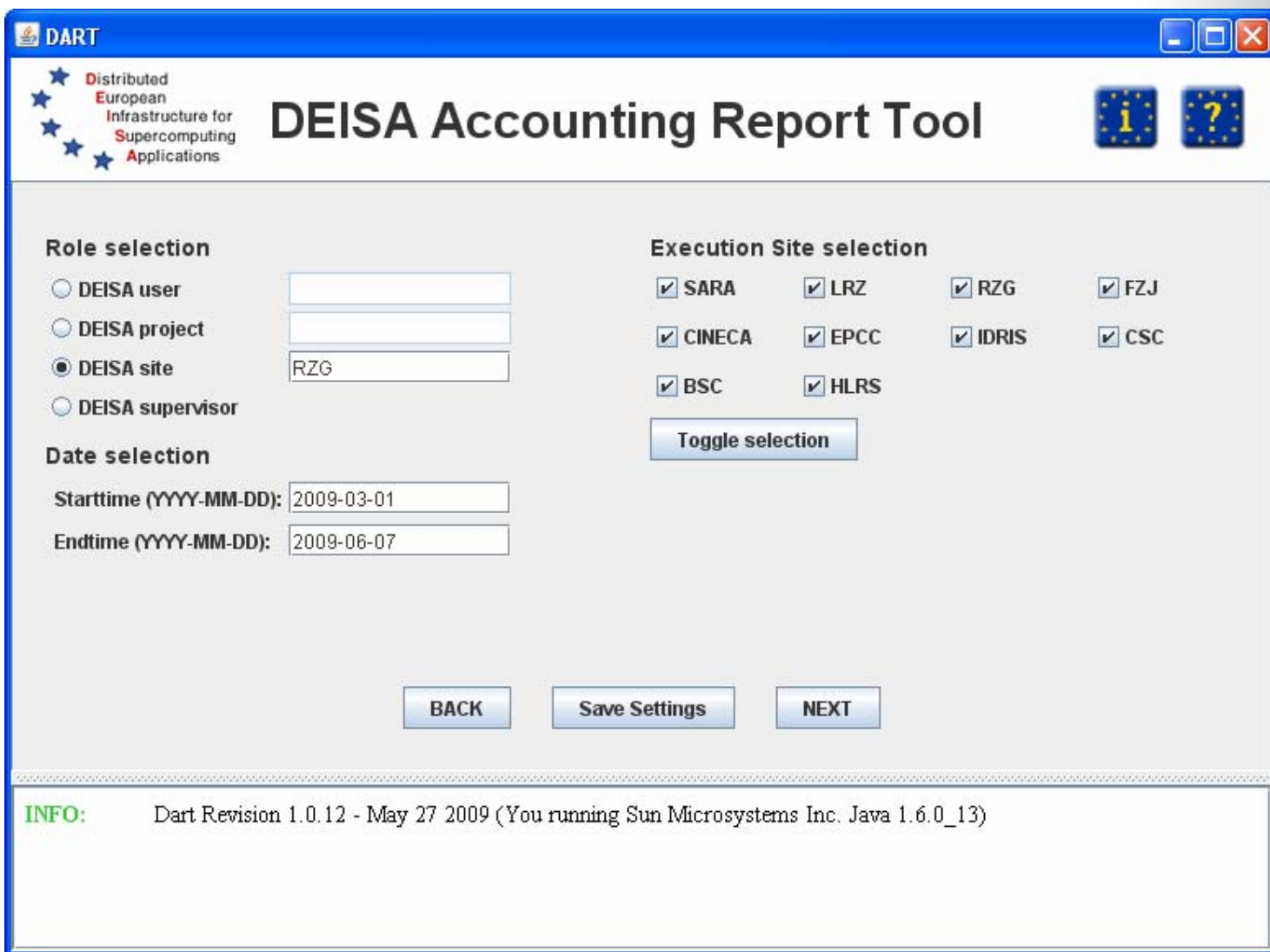
- CGI script in Apache HTTPS server (REST like interface)
- Protected with X.509 certificates
- 4 roles: User, Site administrator, Project supervisor, and DEISA supervisor
- Access control via *accounting-grid-mapfile*
- *accounting-grid-mapfile* can be automatically generated from information of DEISA UAS

```
"/O=dutchgrid/O=users/O=sara/CN=Jules Wolfrat" site-SARA
"/O=dutchgrid/O=users/O=sara/CN=Jules Wolfrat" user-sar00005
"/O=dutchgrid/O=users/O=sara/CN=Bart Heupers"  project-xyz
```

DART Client Tool

- **DART** = DEISA Accounting Reporting Tool

- Java 1.5 program → can be used on most platforms

- Supports PKCS12 and Java Keystore for authentication

- Connects and processes accounting data from all sites simultaneously

- Java WebStart application. No local configuration needed. URLs of accounting information servers and processors normalization factors always up to date

- User friendly GUI

Obtaining Usage Reports

RI-222919

Usage Reporting (2)

RI-222919

Agenda

⇨ DEISA – HPC Resources, Projects, Accounting Requirements

⇨ User administration: Projects, Users, Budgets

⇨ Accounting facilities

- Site-local accounting tools (Accounting Data Provider)

- Utilization of the OGF Usage Record format v1.0

- Accounting Information Service

- Accounting Reporting Tool (DART)

⇨ **Procedures and Policies**

⇨ Lessons learned and future Improvements

Procedures and Policies

- LDAP server and user DB of the DEISA UAS synchronize in less than 24h
    → changes in a local DB propagate within 24h
    → this includes modifications of user roles

- Accounting information service (eXistDB) is up-to-date within 24h
    → local accounting data provider must sychronize with 24h

- eXist DB data remains persistent for at least one year at all sites

RI-222919

Agenda

⇨ DEISA – HPC Resources, Projects, Accounting Requirements

⇨ User administration: Projects, Users, Budgets

⇨ Accounting facilities

- Site-local accounting tools (Accounting Data Provider)

- Utilization of the OGF Usage Record format v1.0

- Accounting Information Service

- Accounting Reporting Tool (DART)

⇨ Procedures and Policies

⇨ **Lessons learned and future Improvements**

RI-222919

Lessons learned

Advantages:

- Standardized protocols and XML schema
- No central database: distributed XML databases, information services
- X.509 based authentication, authorized access
- Simple system architecture, easy to used by users, supervisors, …
- Open Source
  (Apache HTTP Server, eXist DB, Java, PERL)

Problems, open topics:

- All the requested usage records must be transferred from the eXistDB
  to the client for report generation
  (1UR ~ 1,5 kB $\Rightarrow$ 10.000 URs ~ 14,7 MB $\Rightarrow$ 1.000.000 URs ~ 1,4 GB)
  large numbers of records $\rightarrow$ high performance load of the eXistDB
- Risk of incomplete reports, if a site is not reachable
  (1 site: 99% availability => 11 sites: 89,5% availability)
- Risk of incomplete reports if eXist data is complete due to failures of the
  local accounting data provider.

RI-222919

Further Improvements

- Allow Supervisors to see only summaries over a longer period
  - → less details
  - → more user data privacy

- Aggregation of UR Properties

- Introduce Budgeting
  using the LDAP infrastructure to propagate execution site specific project budgets

- Report generation, e.g. summation over URs on the server side
  Summarized values (e.g., jobtime within a given time frame) sufficient for client
  → transfer of less amount of data

- Improve availability of service (Data Replication? Central Repository?)
  Possible strategies:

  - save remote accounting data into databases at the homesite of the user and let users obtain their data from their homesite only

  - central database for aggregated accounting information

Summary

- Accounting data is gathered
  → directly from **batch system logs** (Loadleveler, PBS, NQS, Slurm)
  → or from an **accounting database** (e.g. mysql, postgres,…) that was previously fed by information from the batch systems.
  Proper project assignment requires information from the UAS.

- Accounting data is stored in **eXist XML** database using **OGF UR-WG** format

- **Distributed eXist DBs and information services** provide accounting information.
  Information server realized as HTTPS server with CGI script.
  Each site controls who gets access to which usage records

- Authorized access based on X.509 certificates, using secure communication (https), **different roles**: user, site admin, supervisors

- **DART Client tool** collects all data simultaneously and creates reports (e.g. taking processor performance factors into acount).

- **Jobtimes** (#cores $*$ elapsed time) are **normalized** before summing up across different computer platforms

- Accounting information from all sites is regularly collected by the **DEISA supervisor** and fed back into the **projects and proposal database** for project progress controlling purpose.