**Workshop: Towards an Integrated Information System**

|  |  |
| --- | --- |
| **Meeting:** | workshop |
| **Date and Time:** | 01 December 2011  |
| **Venue:** | Amsterdam |
| **Agenda:** | https://www.egi.eu/indico/conferenceDisplay.py?confId=654 |

Participants 3

Actions 3

List of Requirements 4

Introduction 5

EMI Status, plans and release timeline 5

Top-level cache-registry 6

Information quality 7

Timeline 7

IGE Status, plans and release timeline 7

Current use cases: Information System usage by the EGI operational tools 8

GOCDB 9

SAM 9

Metrics Portal 9

Discussion 9

Use of the Information System in WLCG Experiments 10

Operational problems (DMSU feedback) 11

Usage of information system clients 11

EGI operational requirements 12

EGI.eu 12

NGI\_NDGF 13

NGI\_PL 14

NIG\_DE 14

GPGPU support and generic resource support 14

EGI requirements for the implementation of a federated cloud 15

Discussion 16

Registries (GOCDB, EMIR, IIS) 16

Top-level cache for ARC, gLite, GLOBUS, dCache, UNICORE 16

Information validation 17

Migration to GLUE 2.0 17

# Participants

|  |  |  |  |
| --- | --- | --- | --- |
| Name and Surname | Abbr. | Organisation |  |
| Pedro Andrade |  | CERN | SAM |
| Andreozzi Sergio |  | EGI.eu | EGI policy team |
| Backer Bruce |  | South African NGI | NGI\_ZA |
| Stephen Burke |  | STFC | NGI\_UK |
| Daniele Cesini |  | INFN  | EGI-InSPIRE JRA1 |
| David Collados |  | CERN | SAM |
| Drescher Michel |  | EGI.eu | EGI-InSPIRE SA2 |
| Tiziana Ferrari  |  | EGI.eu | Chairman, notes |
| Field Laurence |  | CERN | EMI |
| John Gordon |  | STFC, UKI | APEL/GOCDB |
| Vera Hansper  |  | CSC/NDGF | NGI\_NDGF |
| Mattias Ellert |  | Uppsala Universitet  | EMI/IGE |
| Emir Imamagic |  | SRCE, NGI\_HR | EGI-InSPIRE SA1 |
| Balazs Konya |  | Lund University | EMI |
| Kostas Koumantaros  |  | GRNET, NGI\_GRNET | EGI-InSPIRE SA1 |
| L’Orphelin, Cyril |  | CNRS | Operations Portal |
| Memon Shiraz |  | Jülich Supercomputing Center | EMI |
| Dimitri Nilsen |  | KIT, NGI\_DE | EGI-InSPIRE SA1 |
| Meredith David |  | SRCE | GOCDB |
| Mats Nylen |  | NGI\_SE | EGI-InSPIRE SA1 |
| Paolini Alessandro |  | INFN | EGI-InSPIRE SA2/DMSU |
| Marcin Radecki (remotely) |  | CYFRONET, NGI\_PL | EGI-InSPIRE SA1 |
| Salomoni Davide |  | INFN | Workload Mgm WLCG TEG |
| Saverchenko Ilya |  | LRZ | IGE |
| Simon Alvaro |  | CESGA | Accounting Portal |
| Peter Solagna |  | EGI.eu | notes |
| Ulf Tigerstedt |  | CSC | TSA1.8 |
| Anders Waananen |  | UCPH | NGI\_NDGF |
| John Walsh (remotely) |  | TCD | NGI\_IE |

# Actions

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Action Owner** | **Content** | **Status** |
| **1** | T. Ferrari/E. Imamagic | to assess the status of gstat probes and verify if any of the probes should be migrated to SAM | OPEN |
| **2** | J. Gordon | to assess the number of sites for which weighted HEP SPEC 06 averages are used for accounting normalization | OPEN |
| **3** | T. Ferrari | to assess the status of deployment of the site BDII and to discuss with NGIs an upgrade plan to EMI versions | OPEN |
| **4** | I. Saverchenko | to collect information about Glue 2.0 usage in Gridway | OPEN |
| **5** | B. Konya | to provide documentation on how to enable the publishing of GLUE 2.0 information in ARC CE  | OPEN |
| **6**  | V. Hansper | to start an upgrade campaign of ARC CEs in NGI\_NDGF | OPEN |
| **7** | U. Tigerstedt | to report to the group results with GLUE 2.0 publication enabled on ARC CE | OPEN |
| **8** | D. Cesini | to discuss plans for usage of GLUE 2.0 information with the operations tool PTs | OPEN |
| **9** | T. Ferrari | to contact E. Yen to gather information about gstat plans for migration to GLUE 2.O 🡪 IN PROGRESS. Mail sent on 15/12/2011. | IN PROGRESS |
| **10** | T. Ferrari  | To assess the interest in publication of GPU information, and in case of positive feedback, set up a task force with GLUE experts to understand how to use the schema | OPEN |
| **11** | B. Konya, L. Filed (for EMI) | To discuss the contribution of EMI to OGF of a proposal for a registry interface  | OPE N |
| **12** | I. Saverchenko (for IGE) | To discuss the possibility to expose a LDAP interface and use GLUE 2.0 as service Level Interface in GLOBUS | OPEN |
| **13** | T. Ferrari | To request a second workshop on information system during the User community forum 🡪 proposal submitted, being evaluated. | CLOSED |
| **14** | T. Ferrari | To investigate a GLUE 2.0 EGI profile to define what is needed to published by EGI services | OPEN |

# List of Requirements

**Requirement 1. HEP SPEC 06 information should be made available accurately. Currently it’s a manual value, and configuring it manually is error prone. HEP SPEC 06 is used for the estimation of installed compute capacity and for normalization of accounting data.**

**Requirement 2. Topological information about Grid resources and Cloud resources should be made seamlessly accessible from the information service.**

**Requirement 3**: **static/semi-static/dynamic information collected from all deployed middleware stacks (ARC, dCache, gLite, GLOBUS, UNICORE) has to be made available by the information discovery service and by the top-level cache. Information on deployed middleware across the entire infrastructure is necessary for monitoring of software deployment and estimation of installed capacity.**

**Requirement 4: a single client which is able to extract information from the UNICORE registry (or a generalization of it such as EMIR) as well as the top-BDII.**

**Requirement 5**: **The state of a service and site (downtime, suspended, in production etc.) should be advertised by services consistently, i.e. status information should be consistently published by the service itself, the information system and the GOC DB (for example a service that is downtime should appear as such in the information system). Clients should make use of downtime information to help prevent unnecessary failures. Currently end-points that are in downtime or are uncertified are published in the central cache regardless of their status in GOCDB.**

**Requirement 6. GPGPU need to be advertised to be made accessible to grid jobs. GPU information is similar to CPU information: processor model, benchmark, number of GPU cores, GPU clock rate, etc. The type of API (e.g. CUDA, OpenCL) has to be advertised as well.**

**Requirement 7. A top-level cache is needed for publication of dynamic information about the cloud. Examples of static/quasi-static information are: hypervisor, images available, authz authn mechanisms, network, available virtual images, the possibility to upload own images and the corresponding format, the data interface supported and its format, the resource capacity (number of cores, storage, memory per node). Dynamic information is needed to select where to submit the virtual machines at a given point in time. Virtual machines themselves do not necessarily need to advertise themselves but the services running on them may publish themselves.**

**Requirement 8. GOCDB, EMIR and IIS are all registries and these should agree on the interface and the data format.**

**Requirement 9. GOCDB should be able to handle multiple GIIS endpoints (LDAP url) for every service that needs to be polled by the top-BDII cache to let top-BDII query these services.**

**Requirement 10. The proposal for GLUE 2.0 information is the following: (1) validation should be run before service bootstrapping (2) in case of errors, the service should not start (3) in case of errors during service operation, the error should be reported in a log.**

# Introduction

Tiziana Ferrari/EGI.eu presents the agenda of the day and the objectives of the workshop: exchange of information about existing plans, discussion of requirements, definition of an action plan to address these.

# EMI Status, plans and release timeline

L. Field/CERN (see slides)

EMI activities are aiming at the consolidation of various components of the information system:

* At a service level, consolidation of the information providers 🡪 EMI Resource Information Service (ERIS), LDAP interface
* Of the UNICORE and ARC registry functionality, which will be developed into a registry service for all stacks, called EMI Information Registry
* Of the information model (GLUE 1.3, Nodugrid schema and Glue 2.0) 🡪 all schemas will be replaced by GLUE 2.0
* Of the data model 🡪 LDIF (the decision was driven by the fact that both gLite and ARC are based on LDIF)
* Single global cache for dynamic information of ARC, dCache, gLite, UNICORE services 🡪 top-BDII
* ARC and UNICORE do not have a concept of “federation” i.e. of hierarchy of information providers 🡪 federation will be supported in EMI
* LDAP v3 is the interface of choice (ARC supports both LDAP and a web service interface).

The EMIR service will be provided to store a minimal set of **static** information for the discovery of service end-points (ERIS instances), which on the other hand can provide dynamic information.

EMIR can be used to deploy a hierarchical tree of Domain Service Registries (DSRs). Global Service Registries (GSRs) are foreseen to be peer equivalent entities getting information from a federation. Getting information from other federations is possible by collecting it from other peer GSRs. Peer GSRs provide the same set of (cached) information.

From example, a top-level DSR is a NGI-level instance, while a GSR can be a registry providing information for the whole EGI (other registries can serve other peer infrastructures e.g. OSG, Naregi, etc.). Can top-level DSRs publish to two or more GSRs? This is not current foreseen but is an interesting use case.

EMIR (see also the specification document attached to the agenda) has the following properties:

* exposes a REST interface
* is a java service (json data structure)
* is based on a push model
* supports GLUE 2.0
* contains downtime information (propagation of downtime information should not be a problem as information propagation time is lower than minutes). If case of specific latency requirements, a requirement needs to be sent to EMI.
* Data persistency? Information was not provided during the meeting.

## Top-level cache-registry

The BDII client will keep talking to the top-level cache (the clients will keep using the same interface and data structure), while the mechanisms used to populate the top-BDII could change as follows

1. The service registers itself into the EMIR,
2. The global cache gets the list of ERIS end-points from EMIR (or GOCDB),
3. The global cache pulls dynamic information from the ERIS of the service.

Clients should use different services depending on the type of information needed:

* the registry is used for service discovery (static information)
* the top-level cache is used for monitoring purposes (to extract dynamic information). For service metadata (i.e. properties which change less frequently for example after reconfiguration, such as downtime status, middleware version etc.) another mechanism could be used.

## Information quality

Sanity checks of information are needed to improve its quality. EMI aims at introducing checks to automate the validation process. Inaccuracies can be caused by old software versions deployed, misconfiguration of services, etc. Only what is required should be published (it is better to have less information but accurate). For discussions on this topic see the last Discussion slot.

## Timeline

The code freeze of EMIR is due on the 14th of December. It will be released with EMI 2.0 as prototype (without undergoing certification). Currently both UNICORE and ARC services can register themselves in EMIR. With the EMI 2.0 release also gLite services will be able to register themselves. The stability of the service will have to be evaluated and the use cases will have to be understood. An example of use case (but beyond the timeline of EMI) would be that a service that consumes dynamic information (such as WMS), gets the list of CEs from EMIR, listens on the message bus for dynamic information of those CEs.

# IGE Status, plans and release timeline

Globus 5 introduces many changes with respects to Globus 4 (which is based on MDS, Glue 1.3, and web services). The Integrated Information System (IIS) addresses the problem of service advertisement/discovery (it does not provide dynamic information). It is not fully compliant with Glue 2.0 (it implements it partially) – the full Glue2.0 compliance is expected in 2012 (first quarter) as IGE effort (an IGE release is expected in April 2012). It includes a registry and a servlet which exposes deployed registries through a RESTful interface or CLI interface (http/https are the supported protocols).

The capability schema is an xml document which describes entities such as applications, computers, services etc.). It’s a legacy that has nothing to do with Glue 2.0.

Multiple registries can be deployed to construct a hierarchical tree (at the lower level of the tree the IIS can be a site service). The beta version 0.3 can be deployed (IIS is being evaluated in other projects such as MAPPER as replacement of MDS).

In order to publish dynamic information into IIS an adapter needs to be developed.

E. Imamagic: is it possible to limit who is advertising the data? A: Yes, access control is available to limit the publication of information.

Q: How can dynamic information for GLOBUS resource be integrated into a global cache?

A: Glue 2.0 compliance will be released during the first quarter of 2012. Dynamic information could be collected by the cache by querying the service end-points.

Q: Will GLUE2 be used for the capability as well?

A: No. The capability schema will stay, and will be complemented by the Glue2.0 information.

Q: you use XML embedded database, is it scalable/reliable?

A: Different database solutions to the GIIS are possible (such as txt files).

Action (I. Saverchenko): to collect information about Glue 2.0 usage in Gridway.

# Current use cases: Information System usage by the EGI operational tools

D. Cesini/INFN presents the dependencies of operational tools on the information system. Several tools extract information from gstat instead of querying BDII directly to take advance from some internal gstat heuristics in estimating installed capacity. Gstat is a visualization tool that also runs probes (more probes that SAM itself runs).

Dependencies on gstat:

* SAM
* Operations Portal

Dependencies on information system:

* Metrics Portal
* SAM
* Gstat
* Operations Portal
* Accounting repository

Dependencies on GOCDB:

* Information system
* SAM (collection of topology information in ATP)
* Operations Portal
* GGUS
* Accounting repository and accounting portal (collection of topology information)
* Metrics portal

**ACTION (T. Ferrari/E. Imamagic): to assess the status of gstat probes and verify if any of the probes should be migrated to SAM**.

T. Ferrari: what is the impact of caching in BDII on operational tools? And what is the WLCG timeline for the deployment of this feature?

L. Field: test results are available, the timeline for deployment is still to be defined.

E. Imamagic: caching has no impact on the operational tools, they would actually benefit from this.

## GOCDB

B. Konya: usage of Glue 2.0 for the representation of information available from GOCDB is desirable.

D. Meredith: GOCDB has EGI-specific status attributes (such as “certified”) that are not present in Glue2.0, however for new service types – e.g. ARC, UNICORE – being registered the Glue 2.0 service type naming convention is adopted (the legacy service types were not adapted to avoid disruption on other tools/applications).

GOCDB and EMIR: some status information (such as downtime, and contact information) will be handled by EMIR.

## SAM

SAM relies on different information sources:

* gstat for information about installed capacity (extracted from BDII).
* REBUS (WCLG): which is used to collect T1 and T2 topology
* GOCDB and OIM for topology information

## Metrics Portal

gstat cannot be used directly because it is not flexible enough (for each federated Operations Centre, metrics have to be collected for each member NGI).

## Discussion

* **Requirement. HEP SPEC 06 information should be made available accurately. Currently it’s a manual value, and publication is error prone. HEP SPEC 06 is used for the estimation of installed compute capacity and for normalization of accounting data.**
* **Requirement. Topological information about Grid resources and Cloud resources should be made seamlessly accessible from the information service (see also the discussion on federated cloud use cases later on the agenda).**

T. Ferrari: During the accounting workshop higher accuracy was requested in job normalization (when HEP SPEC 06 weighted average is used from BDII). Could the EMIR service be used as site service to keep information about WN end-points and the related hardware profile to easily associate the correct HEP SPEC 06 value to the WN on which the node is executed? In EGEE times the swat component was developed. S. Burke: swat was used for central monitoring of WN configuration.

S. Burke: the weighted average is used only when the batch system does not weight the CPU time itself (batch system scaling). T. Ferrari: in how many sites batch system scaling is enabled?

**Action (J. Gordon): to assess the number of sites for which weighted HEP SPEC 06 averages are used for accounting normalization**.

S. Burke: the CE subcluster feature could be used to publish the correct HEP SPEC 06 for a given subcluster of homogeneous WNs. Normalization could be implemented directly by the accounting probe before the UR is published (using the subcluster information), rather than centrally by the accounting repository at the end of the usage record publication.

In order to use EMIR for WN registration a special service (Worker Node) should be used to keep information at a site level. This use case will be further discussed after the assessment of how many sites are relying on published HEP SPEC 06 averages for normalization.

# Use of the Information System in WLCG Experiments

D. Salomoni/Workload Management TEG (see slides for details)

**ALICE**. BDII is used to regulate the flow of jobs (AliEn) – dynamic status information is used. The Alice VO-box checks the CE BDII to verify the site occupancy (more reliable dynamic info is needed in BDII).

T. Ferrari: how would the cached BDII impact this workflow?

A: the information is put in the cache file, if the information returned is stale, it will be marked as UNKNOWN. This will be treated as a problem with the corresponding CE (the workflow excludes the CE).

The ARC information system is used for ARC CEs.

**ATLAS**. A cache of experiment specific information is maintained for the experiment software components (PanDA, dashboards, etc.). ATLAS uses various information sources among which BDII, GOCDB, and OIM. BDII is used for bootstrapping of PANDA. BDII is periodically scanned and information is cached). Additional custom attributes are added to the BDII information.

**LHCb.** DIRAC does not use BDII. The DIRAC job submission framework statically submits to CEs. BDII is used by FTS.

**CMS.** The list of sites is not automatically updated based on BDII information. CMS verifies site functionality before enabling mass submissions. CMS does not use dynamic information for job submission. WNs are validated directly before a glidein starts. However, data management (FTS) relies on BDII.

As an overall summary, BDII is used by the experiments but just to support parts of the existing workflows.

L. Field: as WLCG interest is focused on quasi-static information, WLCG may consider the usage of a registry instead of a cache of dynamic information.

# Operational problems (DMSU feedback)

A. Paolini/INFN provides an overview of open problems concerning BDII deployment that were reported to the DMSU (see slides for details).

Site-BDII needs to depend on OpenLDAP 2.4 not on version 2.3 (as already done for top-BDII), many problems with OpenLDAP 2.3 being used by default by site-BDII have been reported and this badly impacts production. L. Field: this is planned for EMI 2.0.

Problem of bdii tmpfs filling up and silently causing a failure of the top-BDII was reported in <https://ggus.eu/tech/ticket_show.php?ticket=76337> and is badly impacting top-BDII performance of plenty of installations (CERN, Greece, Croatia to name a few). The problem affects recent gLite 3.2 versions as well as BDII 1.0.0 from EMI 1.0.0 (no information is available for version 1.0.1, so this version could be potentially affected too). This problem is now set as URGENT.

L. Field: Many of the issues reported by the DMSU affect gLite 3.2 and are fixed in emi-1 or will be fixed in emi-2.

Q: What's the plan for the use of GLUE2 by the clients?

A: in EMI-2 WMS and LFC will be able to consume GLUE2 information.

From a deployment point of view, many site-BDII are old and cannot handle Glue 2.0 information, so currently just a subset of sites are publishing GLUE 2.0.

Action (T. Ferrari): to assess the status of deployment of the site BDII and to discuss with NGIs an upgrade plan to EMI versions.

# Usage of information system clients

P. Solagna/EGI.eu: usage of information system clients was assessed by contacting the user community (survey), the community of operational tool developers (EGI-InSPIRE JRA1) and the usage was assessed within the operations community during operations meetings. The results of this survey are provided to EMI for future consolidation of clients.

* lcg-info\* (gLite) is widely used by NGI’s operations teams and user support teams
* ARC sites report usage of direct LDAP queries. ARC lib was not reported as used. It is used by SAM according to EGI-InSPIRE JRA1.
* UNICORE cli is heavily used
* SAGA SD API/CLI: no NGIs reported usage of this client
* GLUE2 support: many services will start to publish GLUE2 data from EMI-2, and the clients should be able to use the related schema.

As to the user survey, feedback was received from members of the following VOs: AEGIS, AEGIS, ALICE, CMS, COMETA, GRIDIT, IN2P3-CPPM, INFNGRID, LSGC, LSVRC, SEE, SEEGRID, WeNMR, aegis, atlas, belle, biomed, cms, dgkg, esr, euasiagrid, ipv6, moldyngrid, ops, trgridb, virgo, vlemed, vo.cta.in2p3.fr, voce.

lcg-info\* (gLite) was reported to be used by almost all the users who replied to the survey and it was ranked as the most important information system client. ARC lib was reported to be used by 1 user. UNICORE cli was reported to be used by 1 user. SAGA SD API/CLI was reported by 3 VOs as second client for importance. [See slides for extra details on the survey]

# EGI operational requirements

## EGI.eu

T. Ferrari presents the operational use cases for information discovery.

* **Requirement**: **collection of static/semi-static/dynamic information about all deployed middleware versions and stacks.** Information on deployed middleware across the entire infrastructure is important to support various processes/functions:
	+ Transition planning: are obsolete versions of software being deployed? Which Resource Centres (RCs) need to upgrade?
	+ Incident management: a critical vulnerability is detected, which RCs are affected? Which RCs are patching, and which are not following advisories and hence are eligible to suspension?
	+ Service Operation: which stacks/products need to be supported by the Service Desk? Is the Service Desk adequately staffed?
* **Requirement : assessment of installed capacity. Assessment of installed capacity is important to gain information about the number of Resource Centres, the number of logical CPUs and the amount of storage installed.** This use case supports various processes/functions:
	+ EGI-InSPIRE project management (metrics)
	+ Capacity management: comparison of installed capacity and utilization (accounting data)
	+ Capacity planning

Currently none of the two use cases is fully supported as there is no top-level cache that can provide dynamic information about ARC, gLite, GLOBUS and UNICORE services. Top-BDII is only limited to gLite and a fraction of ARC resources in EGI.

The existing top-level BDII cache is affected by various deployment problems. One of them is the discrepancy between information provided by different top-level instances (different information can be served by different instances). EGI is working toward the definition of a list of authoritative top-BDIIs and the A/R of the service is being monitored since September 2011. A list of authoritative instances is important to ensure applications are using a supported top-BDII instance.

T. Ferrari: is there a technical reason for the existence of the site-bdii? Can’t it be replaced by an instance aggregating at an NGI level? This would have the benefit of reducing the set of site services.

L. Field: the current deployment model based on a site bdii was defined to ensure that fresh dynamic information can be collected by BDII. The limitation is related to the number of service end-points, i.e. to the number of parallel threads that BDII can effectively support. If BDII would just provide static and semi-static information then a change in the deployment model would be possible (currently it pulls dynamic information every 5 min). Another motivation is related to security. By deploying a site BDII only one host needs to be accessed from external consumers, instead of each service end-point.

B. Konya: who is using site bdii now? L. Field: The site BDII is an advertise service, nobody is supposed to query the site-bdii.

T. Ferrari: if the performance problem reported could be solved, the service infrastructure requested at sites would be greatly simplified.

## NGI\_NDGF

Top-BDII and ARC info system are not currently interoperable (the ARC info system that is deployed is based on the Nordugrid schema). A custom solution to translate ARC info into Glue1.3 has been developed to publish ARC resource information into the CERN top-BDII (NDGF has no own top-BDII currently). This workaround is currently serving only sites operated by NGI\_NDGF (not external ARC sites that currently are not published in BDII). A site BDII works as advertise service for all NGI\_NDGF resources.

ARC does not provide a cache capability, only a registry. An ARC-CE pushes data to the index server.

For NDGF glue 1.3 is not acceptable, because of the distributed Tier-1, while GLUE2 can easily represent geographically distributed sites (across many countries).

B. Konya: ARC CE can publish GLUE 2.0, it’s a matter of configuration (documentation will be provided). By publishing GLUE 2.0, information can be pulled by top-bdii without need of format conversions.

**Action (B. Konya) to provide documentation on how to enable the publishing of GLUE 2.0 information in ARC CE in order to publish GLUE 2.0 information**.

In order to publish GLUE 2.0 however a more recent version of ARC CE is needed. Old instances have to be upgraded.

**Action (V. Hansper): to start an upgrade campaign of ARC CEs in NGI\_NDGF**.

**Action (U. Tigerstedt): to report to the group results with GLUE 2.0 publication enabled on ARC CE**

The switching to GLUE 2.0 however requests that also the operational tools can consume GLUE 2.0 information.

Action (D. Cesini): to discuss plans for usage of GLUE 2.0 information with the operations tool PTs.

Given the dependency of some of the tools on gstat, it is also important that gstat can support GLUE 2.0. People attending the workshop are not aware of gstat development plans addressing this.

Action (T. Ferrari): to contact E. Yen to gather information about gstat plans

## NGI\_PL

D. Nilsen/KIT on behalf of NGI\_PL operations manager

In NGI\_PL every production site is accessible by exposing both gLite and UNICORE. NGI\_PL supports three middleware stacks: gLite, UNICORE and QosCosGrid. Because the same set of compute resources is exposed by different service end-points, tools for the computation of installed capacity should avoid double counting.

The UNICORE user communities in NGI\_PL are using products shipped with EMI, and they need a unified information system for both computing and storage. UNICORE-based applications use the UNICORE registry for discovery of UNICORE compute resources.

**Requirement: a client able to query the UNICORE registry (or a generalization of it) as well as the top-BDII.**

**Requirement**: **The state of a service and site (downtime, suspended, in production etc.) should be advertised consistently and taken into account by clients. The service itself, the information system and GOC DB should agree on the state. Clients should make use of downtime information to help prevent unnecessary failures (and tickets). Currently end-points that are in downtime or are uncertified are published in the central cache regardless of their status.**

## NIG\_DE

D. Nielsen/NGI\_DE. The German NGI supports three middleware stacks: gLite, GLOBUS and UNICORE.

For what concerns GLOBUS, currently v. 4 is used and MDS is the reference information system (transition to GLOBUS 5 is still not planned). NGI\_DE requires the deployment of GLUE 2.0 for clear separation of resource and service end-point information, and requires (as stated above for NGI\_NL) consistency of status information between GOCDB and the information system about downtimes, outages etc. Clients should be able to extract this information consistently.

Discussion of publishing of downtime information: can downtime information be handled in the EMIR service? L. Field: this is possible. S. Burke: it is important that downtime information is published at both site-level and service end-point level.

## GPGPU support and generic resource support

J. Walsh/NGI\_IE. Usage of GPGPU is getting more popular for the support of highly parallelized applications. These resources need to be advertised to be made accessible to grid jobs. GPU information is similar to CPU information: processor model, benchmark, number of GPU cores, GPU clock rate, etc. The type of API (e.g. CUDA, OpenCL) has to be advertised as well.

Currently one site publishes “CUDA” as a software tag, however this information is NOT sufficient to express how to access the resource. Support of GPUs in batch systems is still in its infancy. The most advanced support is in Torque (versions greater than 2.5.5), which supports at a basic level of GPU resources.

TCD have a separate queue for GPU resources, and have defined a partition and standing reservation to ensure that only jobs destined for GPU queue go to one of the nodes. Process and GPU separation are needed to prevent that user jobs on multi core nodes try to access GPU at the same time.

L. Field: there are no major problems to include this in the information system. The first point to understand is if GLUE 1.3 is needed or GLUE 2.0.

Q: Do we have user communities demanding this? J. Walsh: Computational Chemistry

Action (T. Ferrari): to assess the interest in publication of GPU information, and set-up a task force with glue experts to understand how to use the schema.

# EGI requirements for the implementation of a federated cloud

M. Drescher/EGI.eu on behalf of the Federated Cloud TF

**Requirement. A top-level cache is needed for publication of dynamic information about the cloud. Examples of static/quasi-static information are: hypervisor, images available, authz authn mechanisms, network, available virtual images, the possibility to upload own images and the corresponding format, the data interface supported and its format, the resource capacity (number of cores, storage, memory per node). Dynamic information is needed to select where to submit the virtual machines at a given point in time. Virtual machines themselves do not necessarily need to advertise themselves but the services running on them may publish themselves**.

The most suitable model is the one where a registry can be queried for static information on federated clouds, and the top-level cache is queried for dynamic information (or the end-points of the services themselves are queried).

The task force is currently looking at OCCI to understand what of OCCI can be published in glue2, and to compare the OCCI and GLUE renderings in order to understand the most suitable abstract resource model.

S. Memon: OCCI is currently providing a subset of GLUE information. What is the position of OGF?

Q: do you see any architectural problem, based on the EMI/IGE architectures to host cloud information?

A: not easy to assess it now. The community forum offers a good chance the contact users and develop this topic further.

# Discussion

## Registries (GOCDB, EMIR, IIS)

S. Memon: EMIR is not storing historic data, is there use case for historic data?

A: historic information is needed in GOCDB to trace service/site lifetime (creation, certification, decommissioning) after a service/site reach end of life, and to keep track of association between services and sites for accounting purposes and security purposes (in case logs are needed after a service/site are canceled).

L. Field: the service discovery use cases of GOCDB and EMIR should be analyzed to explore commonalities (restful interfaces, data models etc.).

The publishing of GOCDB information in GLUE 2.0 schema should be considered.

I. Saverchenko: users typically need information about applications, libraries, the hosting environment all these are dynamic information, and should be exposed.

S. Memon: EMIR exposes a small subset of GLUE 2.0 information, it can have downtime information in the record.

L. Field: everything that is a service should be able to publish itself using GLUE2. We should agree on the interfaces.

**Requirement. GOCDB, EMIR and IIS are all registries and these should agree on the interface and the data format.**

L. Field. Shall we suggest a registry interface to OGF? Both EMIR and Globus are exposing a RESTful interface. People agree that this would be beneficial.

## Top-level cache for ARC, gLite, GLOBUS, dCache, UNICORE

L. Field. The EMI proposal for EMI-2 is the following:

TopBDII is a middleware-agnostic top-level cache which

* pulls information for glite resources from site-bdii
* pulls ARC and UNICORE information from the resource information provider (these resources are discovered from GOCDB). Resources will have to publish GLUE 2.0 information. This will require the capability of top-BDII to generate a list of LDAP end-points for each site (instead of a single one for the site-BDII).

If IIS can also expose a LDAP interface at a resource level, GOCDB GLOBUS resources will be pulled as well (Action, I. Saverchenko).

B. Konya: the operational tools will have to be capable of consuming GLUE 2.0 through LDAP.

S. Burke: EGI should define a top-BDII deploying scenario, because right now every site is free to use an arbitrary top-BDII instance.

T. Ferrari: correct. As starting point, EGI defined the list of authoritative top-BDIIs (one per NGI) for which a minimum availability is guaranteed (99%).

I. Saverchenko: according to this scenario, the most critical thing for globus services is the LDAP interface to GLUE2.0, as the IGE data model is already GLUE 2.0 the data model is not a show stopper and GLUE 2.0 is already in the development plans. There should be an EGI profile that says what is needed to be published from EGI services (Action on T. Ferrari)

**Requirement. To support this scenario, GOCDB should be able to handle multiple GIIS endpoints for every service that needs to be polled by the top-BDII cache to let top-BDII query these services**.

## Information validation

T. Ferrari: EMI has plans for validation of information before it is published. Can you provide more information?

L. Field: there will be a GLUE validator – a generic component – that checks the content of the glue2 schema. The mechanisms on how to integrate it, has to be discussed. If we do a validation, what should we do when we find a problem? Should the service be stopped, or just the publishing of information, or would a validation failure just generate a log information?

**Requirement. The proposal for GLUE 2.0 information is the following: (1) validation should be run before service bootstrapping (2) in case of errors, the service should not start (3) in case of errors during service operation, the error should be reported in a log.**

## Migration to GLUE 2.0

S. Burke: it is time to start an upgrade plan of site-BDIIs in order to

1. fix a known vulnerability problem

2. allow the publishing of services in GLUE 2.0 format

(Action, T. Ferrari)

A second information system workshop will be organized during the user community forum.