



Information System Use Cases

WLCG Management Board

15th September 2015

Maria Alandes on behalf of the Information
System Task Force



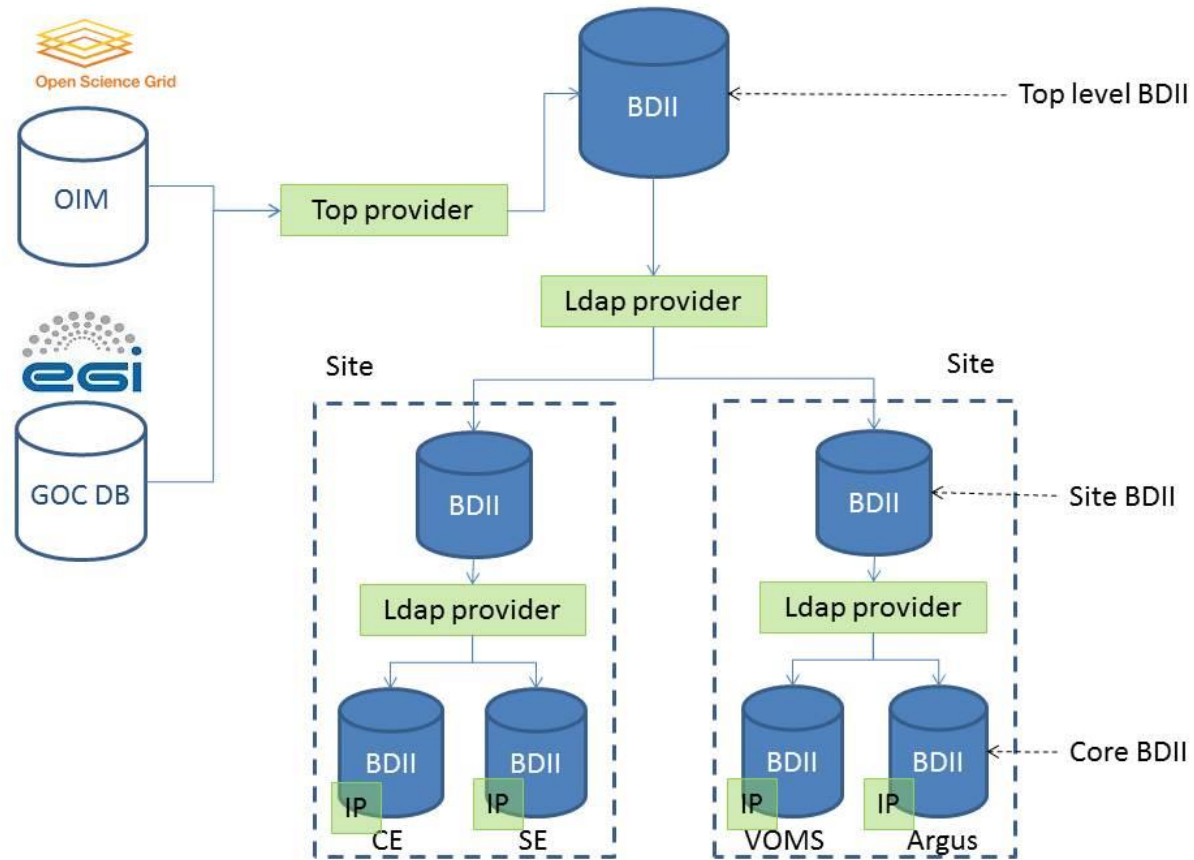


Introduction

- New Task Force created on July 2015
 - <https://twiki.cern.ch/twiki/bin/view/EGEE/WLCGISEvolution>
- One of the main goals is to identify the existing use cases for the WLCG Information System
- A document has been written with input from TF members representing LHC VOs and WLCG activities currently relying on the IS
 - Document can be found attached in the Indico Agenda
 - A summary will be presented in the following slides



WLCG Information System Overview





Uses Cases Summary

- The Information System is currently being used by all LHC VOs
- Many activities in WLCG have also a dependency on it
- Uses cases are many and there is no common pattern
 - Some experiments use it for service discovery, others don't
 - Some experiments integrate the information in their workload management workflows, others don't
- GLUE 1.3 is the GLUE version used by all use cases
- As far as clients are concerned, *Idapsearch* is widely used
 - *lcg-infosites* is also used to a greater extent by sys admins



Drawbacks of the Information System

- Known issues identified for a long time
 - Already reported in 2012 at TEG reports
<https://twiki.cern.ch/twiki/bin/view/EGEE/ISTEG2012Summay>
 - Discussed in Infosys meeting
<https://indico.cern.ch/category/4475/>
 - GSR was an attempt to address these issues
<https://twiki.cern.ch/twiki/bin/view/EGEE/WLCGGISR>
- Some issues addressed in the past years
 - Information accuracy and validation → glue-validator
 - Information volatility → top BDII cache
- Some issues inherent to the existing architecture
 - Mix of static and dynamic information
- Some issues recently discussed
 - Lack of flexibility to aggregate across multiple types of inputs and formats
 - Manual editing of certain classes of information is a desired functionality, while other classes will normally be collected automatically through information providers



Next steps

- The collected use cases demonstrate the need to maintain a WLCG information system
- A review of the current implementation is in any case needed since OSG will stop publishing in the Information System
 - 100% consensus and participation is needed to move forward and ensure interoperation
 - EGI, OSG and NDGF will present their future plans at the next TF meeting <https://indico.cern.ch/event/441747/>
- The main goal of the TF is to work on a new information system
 - To address the identified drawbacks of the existing IS
 - To address the collected use cases
 - Next step would be to define the architecture of the new IS
 - New doesn't necessarily imply a new service written from scratch!
 - Example: Extension of OIM and GOCDB to cover existing use cases
 - GOCDB new features to be presented also at the next TF meeting
 - Example: GSR
 - ... other possibilities to be explored
 - All this will be discussed within the TF



REBUS REQUIREMENTS





New requirements for REBUS

- During the discussion within the TF, the following requirements affecting REBUS have been collected:
 - Experiments are using pledges per site and not per federation
 - Sites are contacted by experiments to get this information
 - Experiments would like to know what it is supposed to be available at a site
 - It would be extremely useful if this could be provided by REBUS per site
 - Even if it is an indicative value and not an actual pledge
 - Definition of installed capacity (or a better name)
 - An official definition of installed capacity is needed
 - Some input already provided in the TF
 - Experiments would like to know what it is actually available at a site → available capacity?



ADDITIONAL MATERIAL



Use of the Information System by the LHC VOs



	BDII Dependency	Main Interface	GLUE version	BDII type	BDII query frequency
ALICE	YES	ldapsearch	GLUE 1.3	Resource & site BDII	Every min
ATLAS				Top BDII	Every 2h
CMS					<ul style="list-style-type: none"> • Bootstrap (*) • Several times per hour (**)
LHCb				Every 12h	

(*) Pilot infrastructure uses the Information System the first time a site is added

(**) Vofeed creation



Use Cases by the LHC VOs

	Use Cases	GLUE 1 Attributes
ALICE	<ul style="list-style-type: none">• CE status	<i>GlueCEStateStatus</i> <i>GlueCEStateWaitingJobs</i> <i>GlueCEStateRunningJobs</i>
ATLAS	<ul style="list-style-type: none">• CE attributes• HEP-SPEC06 and Logical CPUs	<i>GlueSite and GlueCE attributes</i>
CMS	<ul style="list-style-type: none">• CE service discovery• CE attributes• Logical CPUs	<i>GlueCEUniqueID</i> <i>GlueCEPolicyMaxCPUTime</i> <i>GlueCEPolicyMaxWallClockTime</i>
LHCb	<ul style="list-style-type: none">• CE service discovery• CE attributes	<i>GlueCE, GlueHost and GlueSE attributes</i>



Use Cases by other activities in WLCG

	Use Cases	GLUE 1 Attributes
SAM	<ul style="list-style-type: none">Queue name	<i>GlueCEUniqueID</i>
REBUS	<ul style="list-style-type: none">Site Capacities	<i>GlueCE and GlueSE attributes</i>
GFAL2	<ul style="list-style-type: none">SE endpoints	<i>GlueService attributes</i>
IT-SDC C5 reports	<ul style="list-style-type: none">Site Capacity per OS, batch system, CEs and SEs	<i>GlueCE, GlueSE, GlueHost and GlueSubcluster attributes</i>
WLCG Google Earth Dashboard	<ul style="list-style-type: none">Site coordinates	<i>GlueSiteUniqueID GlueSiteLatitude GlueSiteLongitude</i>
APEL accounting	<ul style="list-style-type: none">HEP-SPEC06Message broker discovery	<i>CPUScalingReferenceSI00 GlueHostBenchmarkSI00 GlueService attributes</i>



Feedback from sites

- GridPP sites have also provided feedback on the way the Information System is used
- More feedback from other sites could be collected in future versions of the document
- Highlights
 - Useful for service discovery and migration campaigns
 - Contains too much information that is not used
 - Mix of static and dynamic information
 - GLUE 1 and GLUE 2 schema support is a burden