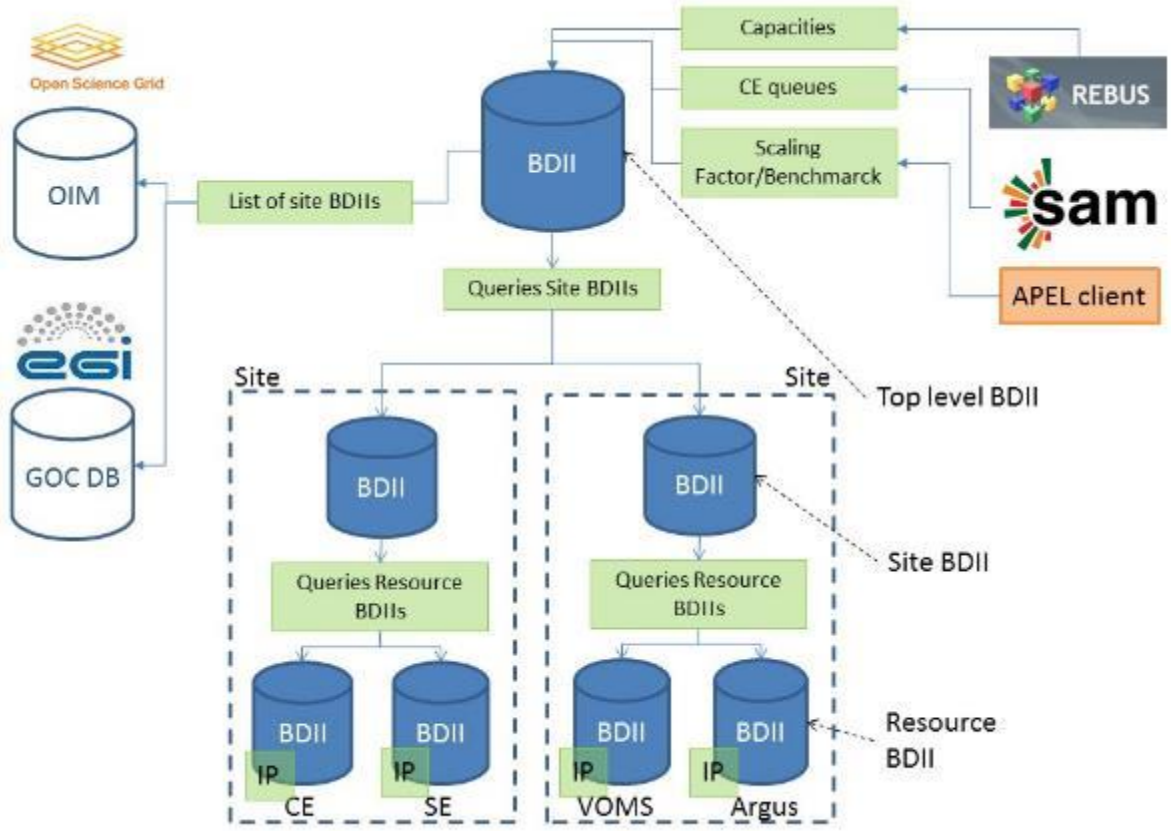
Proposals and next steps

Information System Task Force
July 2016

Current WLCG Information System



WLCG tools dependencies on the IS

- REBUS (https://twiki.cern.ch/twiki/bin/view/EGEE/AllAboutREBUS)
  - It depends on the top BDII to get the information for the capacities view (Logical/Physical CPUs, HS06, Online and Nearline Storage)
- Monitoring
  - SAM3/ETF relies on the SAM BDII to get the list of queue names for CREAM and GRAM CEs
- APEL client
  - It depends on the top BDII to get the scaling factor/Benchmark (GlueHostBenchmarkSI00) to normalise accounting times
- MW upgrade campaigns
  - It uses service type and version information from the BDII

LHC VOs dependencies on the IS (I)

	Attributes vs Information Sources	Resource BDII	Site BDII	Top BDII	GOCDB	OIM	MyOSG	REBUS	Manual modifications?
ALICE	Status of CEs	D/C (CREAM)	D/C (ARC)						NO
	Number of Waiting Jobs								
	Number of Running Jobs								
ATLAS	List of services and associated information: • SEs • CEs • PerfSonars				D/C	D/C			YES
	Queue name	D		C					
	MaxCPUTime	D		C					
	MaxWallClockTime	D		C					
	List of sites and associated information				D/C	D/C			
	Site properties (Lat, Long)	D		C					
	Site downtimes				D/C	D/C			NO
	HS06	D (EGI)					D (OSG)	C	YES
	Logical CPUs	D						C	
CMS	List of CEs	D		C					NO
	Queue name	D		C					YES
	MaxCPUTime	D		C					
	MaxWallClockTime	D		C					
	Logical CPUs	D		C					
	Site downtimes				D/C	D/C			NO
LHCb	List of CEs	D		C					NO
	MaxCPUTime	D		C					
	CPUScalingReference	D		C					
	Site properties (Lat, Long)	D		C					
	Site downtimes				D/C				

	Dynamic: it changes very frequently
	Static or semi-static information: it changes very rarely
D	Defined: where the information is defined by the sites
C	Consumed: where the information is consumed by the experiments

LHC VOs dependencies on the IS (II)

- All LHC VOs dependencies on the IS are for computing related information
  - Mostly static
    - Except ALICE that relies on a few dynamic attributes
- No LHC VO relies on the IS for storage information

BDII in the WLCG IS

- It's the main building block
- However…
  - OSG will stop publishing in the BDII on 31.03.2017
    - They will provide the information needed by their stakeholders through other means (OIM, condor collectors, etc)
  - EGI will keep on relying on the BDII
    - Used by EGI monitoring
    - Used to give site support
    - Used to carry out MW Upgrade campaigns
    - Used to report on total capacity of the infrastructure

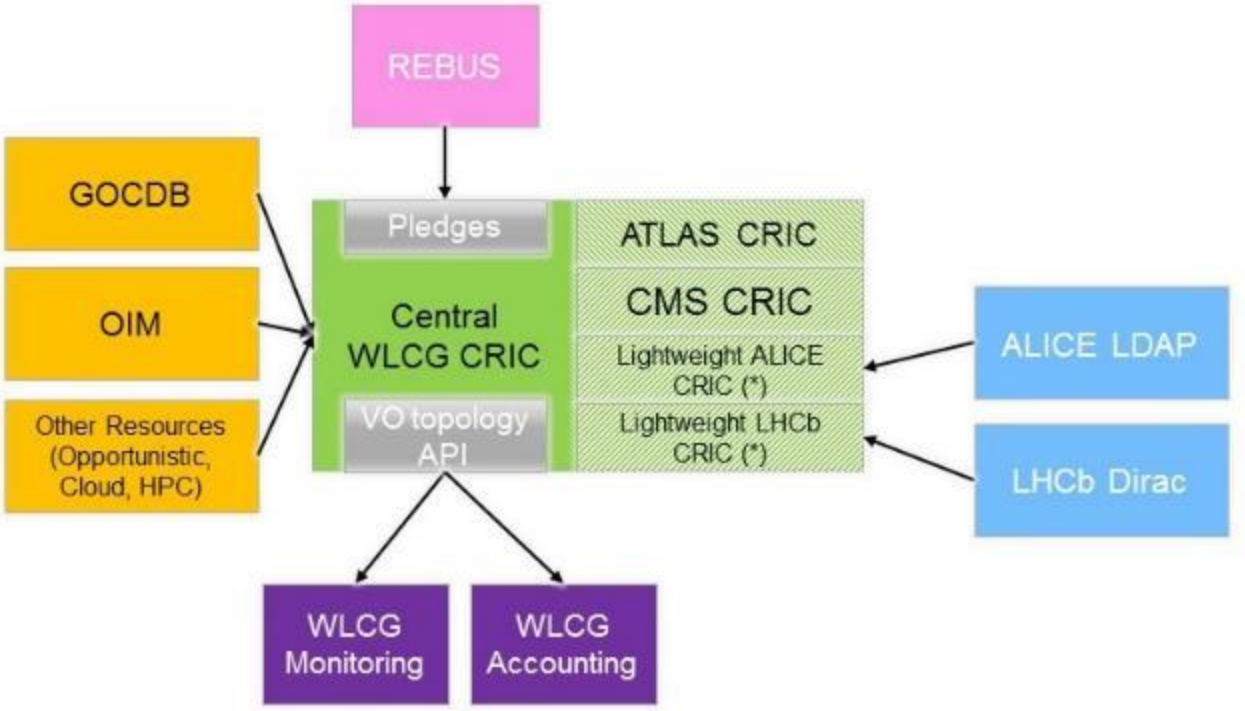
BDII pros and cons

	Pros	Cons
Effort to maintain the service	• Easy to maintain	• Three levels (resource, site and top) needed • Load balanced service for top BDIIs needed • Security patching needed sometimes
Effort to configure information providers	• Automatic in most cases	• Too complex for WLCG use cases (mainly service discovery) • Many attributes have to be published • Only few attributes are consumed
Information quality	• Only place to gather information about site resources (SW versions, capacities, etc)	• Published information is not validated • Validation after publication requires effort • LHC VOs don't trust it
Ease of use	• Existing clients for GLUE 1 schema covering most common queries	• Difficult to learn LDAP queries and GLUE schema • No REST API

Current WLCG IS issues

- It's not flexible enough to adapt to changes
  - Tightly coupled with BDII
    - Once OSG moves out of BDII, we won't have a complete overview of WLCG
      - This is already the case as few OSG sites are not publishing in BDII
- Proliferation of VO home made IS due to lack of trust
  - Quadruplicated effort!
  - Sites effort publishing information in the BDII is worthless as the information is not consumed!

Future WLCG IS



(*) Maintained by WLCG to store very simple experiment topology information (i.e. experiment names)

Computing Resource Information Catalogue

- WLCG Central CRIC
  - Describes WLCG topology
  - Single entry point for WLCG topology consuming information from different sources (GOCDB, OIM, REBUS)
- Experiment CRIC
  - Describes experiment topology
  - It uses the information from central CRIC including additional attributes and configuration needed by experiment operations and workflows
- Lightweight CRIC
  - Maintains a relationship between experiment site names and WLCG site names
    - Needed for monitoring and accounting

Plan to move to the new WLCG IS

- Implement WLCG CRIC
- As a first step consume needed information from BDII (and also GOCDB/OIM)
- Then define GOCDB and OIM as the unique information sources for CRIC (For EGI and OSG resources)
  - If needed, move static information for service discovery to GOCDB/OIM
  - This will allow WLCG stop depending on BDII
    - We need to understand what it is needed to continue the collaboration with EGI to get site support, etc
      - BDII seems to be right now a requisite, we need to find alternatives

Stopping WLCG dependencies on BDII

- Study alternatives for the existing use cases
  - REBUS
    - Add capacities to CRIC
    - Maintain these values manually at least once per year
      - Identified use cases only need an estimation
      - No need for 100% accurate values
  - Monitoring
    - SAM3/ETF should get CE queue names from CRIC
  - APEL client
    - Understand better how normalisation/scaling has to be performed (part of the Accounting TF)
      - Hopefully no information needed in the IS!
  - MW upgrade campaigns
    - Explore other alternatives like Pakiti or MW Package Reporter

Next steps for the IS TF

- Document central CRIC features and what it should provide
- Then plan transition to stop BDII dependencies
  - In the meantime, the IS TF can explore the different alternatives and define a plan to be implemented when CRIC is ready
  - 5 working areas
    - REBUS capacities
    - Monitoring
    - APEL client
    - MW upgrade campaigns
    - EGI support
- No changes in REBUS, BDII, GOCDB/OIM for the time being

Additional material

Why EGI keeps relying on BDII

- EGI and NGIs needs BDII on storage resources to assess the total installed storage capacity
  - to justify national and European investments in the infrastructure
- EGI helpdesk uses the information system and EGI monitoring for investigating on problems affecting the resources
  - In many cases the problems affects interacting resources (CE and SE), knowing the status of all the component is fundamental
  - SE not published are not monitored
- Taking services out of the EGI scope will **stop applying any operational procedure** to these services
  - For not published resources the EGI scope tag is removed for preventing monitoring failures
  - For example MW/SW upgrade campaigns
- For these reasons EGI is actively working on the information system
  - GLUE2.1 cloud extensions
  - VAPOR (web information system browser)
- Most of the NGIs will suggest sites to continue publishing SE information nevertheless

Information Sources overview

GOCDB

	Input	Update Frequency	Technology	Clients	Output	GLUE Schema	Validation
GOCDB http://gocdb.egi.eu/	Manual	Depends on each service	Oracle Apache	REST API (read only)	GLUE2 XML GLUE2 JSON	2	NO

- It contains static information about the following services relevant for WLCG
  - ARC-CE, CREAM-CE, HTCondor CE
  - FTS, VOMS, MyProxy, site and top BDII
  - SRM, Classic SE, xrootd, xrootd redirector
  - perfSONAR

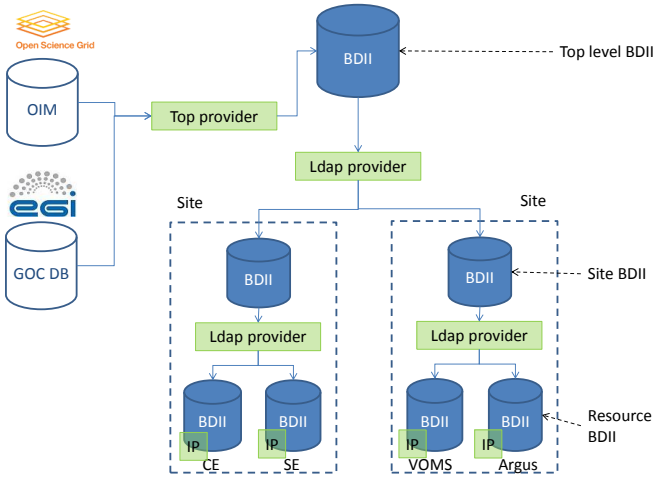
OIM

	Input	Update Frequency	Technology	Clients	Output	GLUE Schema	Validation
OIM https://oim.grid.iu.edu	Manual	Depends on each service	???	None	MyOSG XML	NA	Basic syntax validation

- It contains static information about the following services relevant for WLCG
  - CE, Condor Collector
  - FTS, VOMS, BDII
  - SRM, xrootd, xrootd redirector
  - perfSONAR
  - Squid

BDII

	Input	Update Frequency	Technology	Clients	Output	GLUE Schema	Validation
Resource BDII	Mixed: • Manual (configuration file) • Automatic (Information Provider)	10min	Python Openldap	lcg-info lcg-infosites ginfo ldapsearch	LDIF	1 and 2	YES After publication
Site BDII	Automatic	10min					
Top BDII	Automatic (Bootstraped by GOCDB/OIM)	10min					



REBUS (Installed Capacities)

	Input	Update Frequency	Technology	Clients	Output	GLUE Schema	Validation
REBUS https://rebus.cern.ch	Automatic (top BDII and MyOSG)	1h	Django Python	None	JSON, CSV	1	NO

How information sources are currently used

Details taken from Information System Use Cases Document

ALICE

Attributes	Where is the information defined?	Freshness?	Where is the information consumed?	Freshness?	Is it modified manually by the experiment?
Status of CEs	CREAM resource BDII ARC site BDII	10min	CREAM resource BDII ARC site BDII	One per minute	NO
Number of Waiting Jobs					
Number of running jobs					

ATLAS

Attributes	Where is information defined?	Freshness?	Where is information consumed?	Freshness?	Is it modified manually by the experiment?
List of services (SEs, CEs, PerfSonars)	GOCDB/OIM	Depends on the site admin	GOCDB/OIM	Every 6h	YES
Queue name	Resource BDII	10min	Top BDII	Every 2h	
MaxCPUTime					
MaxWallClockTime					
List of sites	GOCDB/OIM	Depends on the site admin	GOCDB/OIM	Every 6h	
Site properties	Resource BDII	10min	Top BDII	Every 2h	
Site downtimes	GOCDB/OIM	Depends on the site admin	GOCDB/OIM	Every 30min	NO
HS06	Resource BDII (EGI) MyOSG (OSG)	10min (EGI) Manual (OSG)	REBUS	Every 24h	YES
Logical CPUs	Resource BDII	10min			

CMS

Attributes	Where is information defined?	Freshness?	Where is information consumed?	Freshness?	Is it modified manually by the experiment?
List of CEs	Resource BDII	10min	Top BDII	Bootstrap glideinWMS factory	YES
Queue Name					
MaxCPUTime					
MaxWallClockTime					
Logical CPUs					
Site downtimes	GOCDB/OIM	Depends on the site admin	GOCDB/OIM	??	NO

LHCb

Attributes	Where is information defined?	Freshness?	Where is information consumed?	Freshness?	Is it modified manually by the experiment?
List of CEs	Resource BDII	10min	Top BDII	Every 12h	NO
MaxCPUTime					
CPUScalingReference					
Site Properties					
Site downtimes	GOCDB	Depends on the site admin	GOCDB	??	

Other Information Sources

Use cases where information is taken from somewhere else

ALICE

- Storage
  - Site admins manually provide storage information in the ALICE LDAP server
  - MonALISA sensors on the SE head nodes and disk servers collect stats on what is actually available
    - Known as “xrootd info”
    - http://alimonitor.cern.ch/stats?page=SE/table

ATLAS

- Storage
  - The site admin manually define SEs endpoints in AGIS

CMS

- Storage
  - Maintained by each site manually in the SITECONF/PhEDEx/storage.xml file
    - Automatically published to central PhEDEx by PhEDEx agents running at the site
    - Also stored in GIT
    - No synchronisation with GOCDB/OIM
- Xrootd Federations
  - Endpoints maintained manually by the Dashboard team
- Topology vs Site downtime information
  - The VOFeed is used to complement site downtime information
  - VOFeed is not synchronised with GOCDB/OIM
  - VOFeed currently uses BDII for CEs and PhEDEx for SEs
- perfSONAR?
  - Not used by CMS at the moment

LHCb

- Storage
  - Endpoints are inserted manually in DIRAC CS
    - Only a few of them are needed
- perfSONAR
  - GOCDB will be used in the future to discover the endpoints and synchronise information in DIRAC CS (under test, not yet part of production release)