

Ideal information system - CMS

Andrea Sciabà

IS meeting with users
January 9, 2013

- **Service discovery**
 - SAM **VO feed generation**: IS used to find CEs and SEs to be tested by SAM
 - Site **visibility**: SSB publishes whether site CEs are visible in BDII
- **Software installation**
 - IS used to publish **installed CMSSW releases**
 - Information on OS, maximum wallclock time, RAM is used by software installation jobs
- **SE capacity**
 - Used, free, total online/nearline space at sites shown in dedicated SSB view
 - **Not used in production** as the information **cannot be trusted**
- **CE status**
 - CE status, (max) total jobs, max CPU time, close SE, estimated response time, etc. used for **gLite WMS CE matchmaking** (soon to become irrelevant)

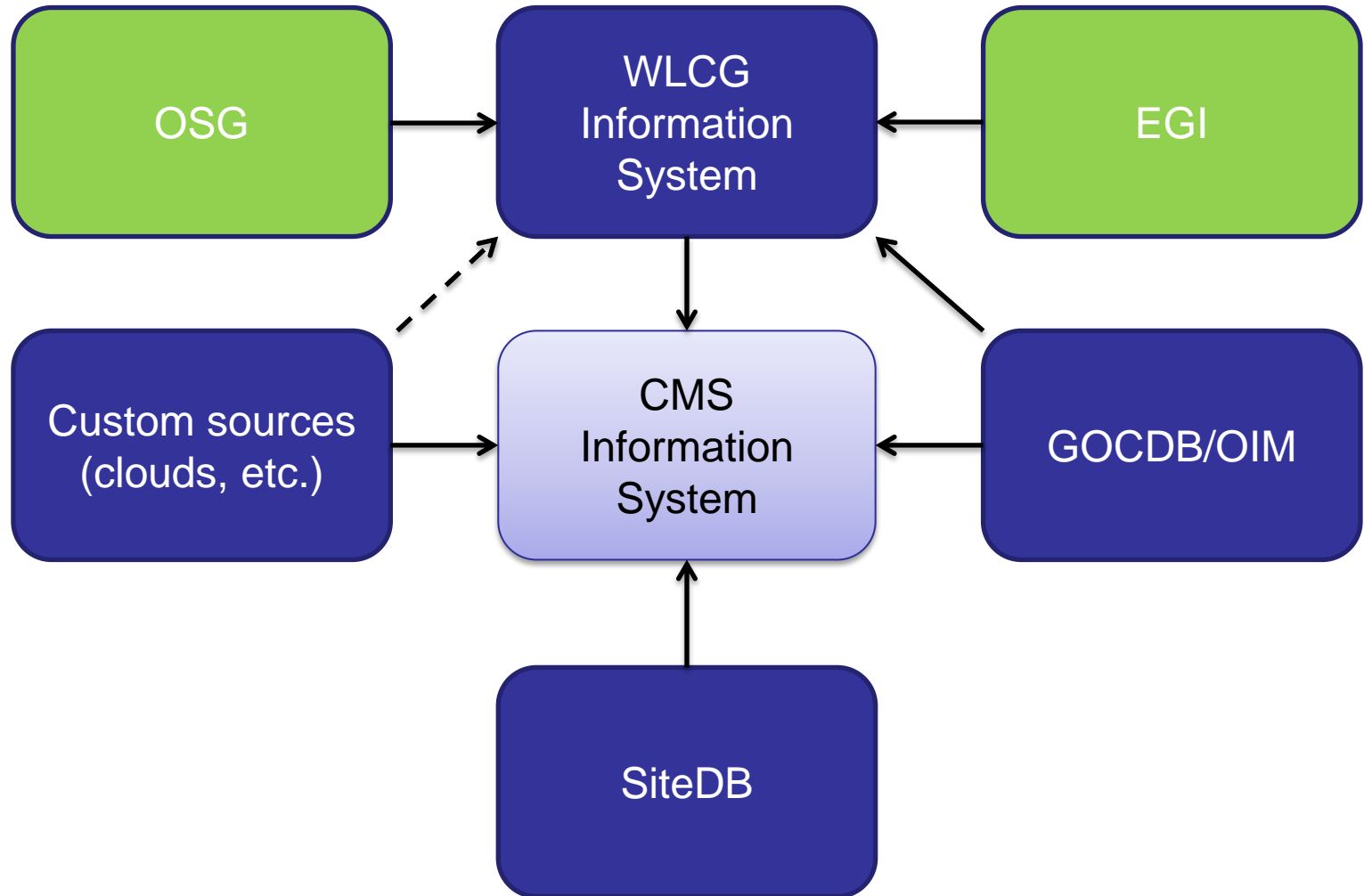
- Publish information for new services
 - E.g. squid for Frontier and CVMFS
 - not an experiment-specific solution if possible!
- Publish information for opportunistic resources
 - E.g. HPC, clouds, non-WLCG sites
 - In other words, plug-in custom sources of information
- Publish information useful for multicore jobs
 - Need to know if resources offered by sites are compatible with the jobs

- Software installation
 - With the full adoption of CVMFS, software tags will become unnecessary
- CE status
 - With the decommissioning of the gLite WMS, matchmaking against dynamic CE attributes will not be done anymore



- Resource discovery must be the main purpose
- Existence of resources
 - List sites and resources at sites
 - Never changes between service commission and decommission
 - Must always be available
- Resource properties
 - Everything which is needed to use a service
 - Generated by the service itself
 - Resilient to glitches and short downtimes
 - Guaranteed to be correct
- Aggregates information from EGI, OSG and NG

- Other requirements
 - REST interface with standard output formats
 - A simple query language (hide LDAP!)
 - A good client
- Experiment-specific information
 - VO site names, custom services, experiment contacts, etc.
 - Stored in VO-specific databases (e.g. SiteDB)
- GOCDB/OIM
 - Keep using for administrative information, downtimes and basic service discovery
 - The service discovery could be enhanced with more properties



- The schema is an implementation detail
 - CMS does not require GLUE 2.0 but it is natural to assume that a common schema helps to provide a uniform description of WLCG resources across federated grid projects
- A correct status of storage and compute resources is “nice to have”
 - Validation efforts for dynamic information are welcome even if not critical
 - Adopt a more “honest” approach: publish information only if it is accurate
 - SE capacity however is particularly important and may profit from storage accounting developments

- A lack of an information system would cause a suboptimal usage of resources (including operational effort)
- The information system (together with GOC/OIM) should provide resource discovery and publish properties for WLCG sites and services
 - If the IS is infrastructure-specific, WLCG should provide an aggregator
 - It must not be VO-specific