Minutes of Information System Task Force, 12th November 2015

Local: Maria Alandes (chair, minutes), Alexey Anisenkov, Alessandro di Girolamo, Maarten Litmaath, Andrea Sciaba, Andrea Valassi.

Remote: Brian Bockelman, Stephen Burke, Alessandra Forti, Diego Gomes, Balasz Konya, Andrew McNab, Alessandro Paolini, Oxana Smirnova.

Agenda available in Indico

https://indico.cern.ch/event/454975/

1. GLUE 2.0 Validation for WLCG

Maria Alandes presents the ongoing validation activities implemented using the SSB. Only ALICE has expressed interest in performing a validation for the number of waiting jobs. Maria presents the mechanisms in place to automatically validate the WaitingJobs attribute for ALICE and open GGUS tickets to sites if the 444444 values are published after 3h. For the time being the GGUS testing instance is used. Maria also presents the mechanisms that were used in the past to validate information for LHCb (MaxCPU time and Storage information for T1s). Maria explains that the monitoring of this information is in place but no GGUS ticket is opened to the sites at this moment since this part has been disabled until LHCb expresses interest in reviving this. Andrea Valassi mentions that SRM and BDII numbers do not always match in the SSB view. Maria Alandes explains that based on past experience, this mismatch usually comes from misconfigurations in the storage info providers. Maria explains that SRM and BDII storage numbers monitoring for LHCb proves that BDII numbers are as reliable as SRM numbers. Maria Alandes concludes that taken into account the little feedback received so far, it doesn't make sense to define a GLUE 2.0 profile for WLCG. It is also mentioned that it would be interesting if validation of the published information is done at the source to avoid publishing wrong information in the first place. Maria explains that discussion with URT is ongoing to understand how glue-validator could be run before the information is published in the BDII. It is also mentioned that EGI is actively running glue-validator against EGI sites to validate the GLUE 2.0 profile for EGI. Maria will check with OSG whether they implement any sort of validation.

Action items:

- *Check validation mechanisms in OSG (Maria Alandes)*

2. **ALICE plans to move to GLUE 2.0**

Maarten Litmaath explains that ALICE could easily change its code to start querying GLUE 2.0 attributes. However, until there is no clear roadmap to stop publishing GLUE 1.3, ALICE is not going to do any changes, as GLUE 1.3 is working fine. There is a general discussion that in general moving to GLUE 2.0 is a no brainer for the four LHC experiments and that publishing only one schema will simplify the life of sites, which is a reason good enough to stop supporting GLUE 1.3. It is decided that the TF will define a roadmap to stop publishing GLUE 1.3.

Action items:

- *Define a roadmap to stop publishing GLUE 1.3 (Maria Alandes)*

3. **AGIS plans to move to GLUE 2.0 and simplify interactions with IS**

Alexey Anisenkov presents AGIS and the plans to become a more general purpose tool. Maria Alandes asks what type of validation AGIS implements. Alexey explains that there are several mechanisms in place: dropdown menus with limited options to force the users to choose among well-known values (i.e. protocol types), typos validation, and correct hostnames with ports defined. Maarten asks whether for instance, storage numbers validation are implemented and Alexey confirms this is not available. Andrea Sciaba asks what it is the timeline to make AGIS a more generic framework. Alexey explains that there is no timeline defined as this is a background work done by him when he finds some time. Alessandro di Girolamo adds that in any case, if CMS decides to adopt AGIS, this work could be done in 3 or 4 weeks. Right now is not a priority but this could change. Extending the schema is easy, the difficult part comes from understanding the relationships and this needs to be done together with CMS experts.

Alexey also presents a proposal to ask sites to publish only the needed attributes by LHC VOS in JSON or XML format using a REST-full interface. Since this seems to be in line with LHCb's slides, it is agreed to wait for Andrew McNab's presentation before discussing this proposal.

4. **CMS future use cases for the information system**

Andrea summarises CMS future use cases. He explains that CMS is currently asking sites for pledges. This is needed for operations and management. In the past, sites were asked to enter this information manually in siteDB, but not all sites were doing it. If this could be collected by WLCG in a central place, it would be very useful.

5. **LHCb actions to move to GLUE 2.0 and simplify interactions with the IS**

LHCb presents the work done by Andrew McNab to collect information directly from resource BDIIs using the URL field in GOCDB in order to populate Dirac. For the time being this is using GLUE 1.3. Maria asks when LHCb will start querying GLUE 2.0 and Andrew answers that this is not known yet. For the time being he has implemented this proof of concept. Further discussions within LHCb are expected to happen at the Computing week happening the following week.

6. **Vcycle and Vac support for GLUE 2.0 publishing via JSON/HTTPS**

Andrew presents how Vcycle and Vac resources could publish information using GLUE 2.0 via JSON/HTTPS. A similar approach to the one presented in the previous presentation is followed: the Vcycle and Vac resources are published in GOCDB. The URL field in GOCDB is used to declare the HTTPS endpoint from which the JSON file publishing information about these resources could be

obtained. Andrew explains the benefits of using this approach for services that don't run a resource BDII. Maarten asks whether Andrew had a look at GLUE 2.1 where cloud resources are supposed to be better described, but Andrew explains that he has adapted GLUE 2.0 attributes to describe Vcycle and Vac. There is a general discussion about the advantages of asking sites to publish information in JSON through HTTPS. Alessandro likes this approach which is very similar to what ATLAS has proposed, as this could be an evolution of the current model. Maria asks how this approach could get rid of the problems that the current information system is facing. Validation is still an issue. Maarten reminds that EGI sites will still need to support the BDII so this doesn't simplify things since it will be extra work. Maria asks whether the sites will have to write this JSON file manually or whether the developers will have to write an extension of the information provider to provide the JSON file, which will be also more work for them. Alessandro explains that in the current information system there is information which is wrongly published, like storage endpoints and associated protocols that are not published in the same way by all sites. This approach would gather only the needed information giving clear instructions to sites to provide what it is needed. Maria explains that this is what current information providers are supposed to do. Andrea adds that in case this is not happening, it is a bug in the information provider and it should be fixed. Maria asks for concrete examples of sites not publishing this information properly so that we could have a look.

Action items:

- *Give examples of wrongly published information (Alessandro di Girolamo)*

- *Study the proposal of publishing a subset of the current GLUE schema in JSON/HTTPS based on the attributes needed by WLCG (Maria Alandes)*

7. **Next meeting**
   26th November 2015 at 15h30.