

Minutes of Information System Task Force, 12th May 2016

Local: Maria Alandes (chair, minutes), Maarten Litmaath, Julia Andreeva, Jerome Belleman, Aleksandr Berezhnoi, Alberto Aymar, Marian Babik

Remote: Eygene Ryabinkin, Alessandro di Girolamo, Andrew McNab, Alessandro Paolini, Oxana Smirnova, Florido Paganelli, Marc Caubet, Scott Teige, David Meredith, George Ryall, Stephen Burke, Dave Kelsey, Vincenzo Spinoso, Alessandra Forti, Carles Acosta, Stephan Lammel

Agenda available in Indico

<https://indico.cern.ch/event/517084/>

1. Experience on removing BDII from dedicated WLCG Storage

Marc Caubet presents pic motivation and recipe to remove BDII from their dedicated WLCG dCache servers. Maria asks whether there have been any issues after removing the BDII, Marc confirms nothing seems to be broken in terms of SAM tests. Dave Kelsey adds that security implications of this need to be better understood. It's not clear what happens when removing the EGI tag from GOCDB. Maybe instead of local, another tag could be used. Maria explains the wlcg tag is already in place. There is also a question about REBUS installed capacities if sites start to stop BDII. Maria explains REBUS storage capacity is not used, but it is indeed a good question since it has to be decided what to do with REBUS installed capacities.

Action items:

- *Checkpoint with pic in the next weeks to monitor whether removing the BDII for their storage is still working fine and not breaking anything else (Maria, Marc)*
- *Start discussions on what to do with REBUS installed capacities view (Maria)*

2. EGI input on stopping the BDII

Alessandro presents EGI concerns on stopping the BDII. The main concern is whether removing the BDII could compromise security. This requires further investigation since there are no conclusions yet. Maria asks EGI to come back to the IS TF meeting and confirm whether BDII plays a role in security monitoring. If there are use cases where information is needed from sites or resources, other alternatives than the BDII could be also discussed. Maarten adds that in any case, BDII is a freely accessible public system, so maybe not the best place to publish security related information. EGI also claims that the BDII is used for storage accounting, but Maarten adds that this is a hack since this should be done using the accounting portal. Alessandro di Girolamo asks what happens with security incidents in cloud sites or T3s. Dave replies this is under the experiment

responsibility. Maarten explains that EGI is doing most of the security work for WLCG, therefore we should work together to make sure they have the tools they need to do this work. Eygene explains that the BDII is useful for sites. Maria adds that even if WLCG eventually makes an official statement that it doesn't need the BDII, the sites will be still free to deploy this service if it's useful for them. Oxana adds NDGF lived many years without the BDII until WLCG requested it.

Action items:

- *Checkpoint with EGI in the next meeting to understand what role BDII plays in security monitoring and the implications of stopping BDII publication at the sites (Alessandro P, Vincenzo)*

3. Experience adding computing static information in GOCDB

Maria presents the experience reported by Gareth Roy from Glasgow adding a subset of static computing attributes in GOCDB. This can be easily done as of today. Maria explains that the next step is to consume this information from GOCDB, and work with AGIS developers is ongoing. If this is fine, it will be evaluated how to add more sites and more experiments to plan for a wider deployment. Andrew explains LHCb would be interested on this since they would prefer to consume information from a single place, and they already rely today on GOCDB. Stephan explains that this could be an intermediate step for CMS, but in principle CMS would prefer to move directly to CRIC and consume information from there. Regarding OIM, Maria explains there is also work ongoing with OSG. Scott explains that populating OIM with this data seems like a good idea as long as it is quite static. The origin of the information would probably be the OSG collector which already provides this information, although it's more dynamic. Julia explains that since OIM provides downtime information, it has to be made sure that OSG topology is properly reflected there in any case. Maria suggests to take this offline and report the progress in the next meeting, as in any case, work to consume this from AGIS is in the plans.

Action items:

- *AGIS to consume static computing attributes from GOCDB/OIM (Alessandro dG, Scott, Maria)*

4. Writeable API in GOCDB

David presents the writeable API in GOCDB and the features that could be implemented in a first version. David expresses some concerns for the scalability of the service if the information is going to be very dynamic. Maria and Julia confirm that the information is going to be static or semi-static. The main use case is service discovery. In case experiments need dynamic information, other solutions need to be found. David explains the security model based on API keys. There is a general preference for robot certificates instead. David agrees to use certificates instead. Maarten also requests to have the possibility of having a privileged user. This is already available in GOCDB. The timeline for this due to new developer in the team is ~3 months.

Action items:

- *Report about the progress of the writeable API development in the next meeting (David, George)*

5. BDII dependencies for VOfeed generation

Maria presents existing BDII dependencies in both VOfeed and ETF thanks to input from Marian. The proposal is to rely on CRIC in the future to obtain VO topology, so for the time being, it would

be better if experiments don't put any effort in VOfeed development. Marian explains that ETF has just started and that it would be good to indeed freeze any developments in VOfeeds for the time being. But in order to fully benefit from ETF, at some point queue and storage path information should come from the VOfeed, so it would be good to have a timeline in mind. It has to be taken into account that once queue and storage path information is available directly from the VO topology, ETF will require some changes as well. Progress on defining VO topology specification could be made in any case. Julia explains that CRIC prototype is progressing well and that in one month or so it is expected to have a more concrete plan to be presented at the MB. A checkpoint can be made by then to see where we are and have more concrete plans for VOfeed/ETF.

Florido asks about dynamic information like waiting or running jobs that is currently consumed by ALICE and LHCb and that he has to provide as ARC developer. ALICE explains that they would like to rely on the BDII as much as possible for this since this is the easiest scenario for them. If the BDII was to be eventually stopped, ALICE could get this information directly from Condor. Florido explains that in any case this is also available through the ARC API, it doesn't depend on the BDII. Andrew explains that LHCb is already using this information through ARC commands and not the BDII.

Action items:

- *Checkpoint on CRIC progress and how this affects VOfeed/ETF plans (Julia, Marian)*

6. Validation of the list of WLCG resources per VO: GOCDB/OIM

Maria explains the ongoing work made by Aleksandr to validate VO tags in GOCDB. Since this is a first exercise, several issues have been raised from sites and experiments. It has to be decided whether it is OK to have resources tagged at the sites but not used yet by the experiments. At the moment, tickets are opened for every discrepancy between GOCDB and VOfeed. Maarten explains that there may be exceptions that we should take into account. Maria explains that in particular for LHCb VOfeed not all the production resources were present, and it was believed that VOfeeds were always reflecting experiment topology with accuracy. Andrew explains that this is not the case for LHCb where there were in fact some inconsistencies found. These will be fixed, but still, for LHCb there are some resources provided locally that are not included in the production framework. Julia asks whether in those cases, LHCb may want to be notified of downtimes. It is decided to clarify with LHCb what to expect from their VOfeed and how they want to manage those local private resources. Work with OSG hasn't started yet but this will be also validated using their VO ownership property.

Action items:

- *Confirm with LHCb the status of their VOfeed and what to do with local private resources (Maria, Andrew)*
- *Propose a more refined set of use cases for VO tag validation after the first validation round and decide on the frequency to automate this in the future (Maria)*

7. Next meeting

16th June 2016.