

Attendees

- Daniela
- Brian
- Dirk
- Andrea
- Gerard
- Graeme
- Maarten
- Wahid
- Philippe
- Giacinto
- Pepe
- Gergely
- Elisa
- Markus

Discussion:

- From Philippe - Inputs/experiences from earlier working groups. Felt like time wasn't well-spent, as the middleware providers sometimes ignored the WLCG reports. What will be produced, who will it be presented to, and will it be listened to? Past experience with SRMv2 especially resulted in little progress with the providers.
 - ◆ Cannot be "requirements on the market" with no follow-up.
- From Graeme: Shares Philippe's concerns. Would help if, from the experiment point of view, what we intend to do and what interfaces we demand. If mwre doesn't comply, we won't use it.
- Dirk: This is a WLCG working group which can clearly document the experiment strategy. This way external projects can judge the impact of a mismatch.
- PC: It must be clear that the WLCG will do the recommendation of the TEG, regardless of what the middleware does. If the middleware is helpful, great; if not, we must not just "forget it". We can't just do "whatever the middleware workplan is".
 - ◆ ML: be careful with "requirements" - they may not materialize, you need a plan B.
 - ◆ We must also make sure the requirements materialize.
 - ◆ PC: We must make sure the WLCG drives, not an abstract community.
- BB: Will try to find commonalities and highlight
- PC: POOL/ROOT/PROOF - this is more about data persistency than data management. AV agrees. AV: thinks we should just have a short, clear statement and be done with it. LHCb has essentially dropped POOL as it no longer needs the ROOT decoupling layer that it was providing, which will leave ATLAS as the only user.
 - ◆ AV: wrt to the previous conversation, POOL is an example where it's external to the experiments but internal to the community and common to many experiments, and this model has been very successful because all developments have been driven from the experiment requirements. This is true both for POOL (even if this is being discontinued as its technical justification of ROOT decoupling no longer holds for most experiments) and for CORAL and COOL (which are still actively supported as discussed in the Database TEG). PC agrees.
 - ◆ Expecting to get a short statement about POOL from ATLAS/LHCb and put it in the TEG's report.
- DD: Suggest do use bi-weekly meetings mainly for core/controversial topics, which need interactive discussion. We will need to do the bulk work via email and should structure input requests (to experiments, sites, s/w providers) with concrete questions.
- About F2F: maybe we target for January to document the "status quo". Concerns about actually being able to provide a coherent strategy document by that time.
 - ◆ Document the status quo, document known commonalities, and "known upcoming changes".
 - ◆ Other TEGs will go definitely go beyond the February deadlines.

- ◆ Noted that there's a massive overlap with storage management. Maybe overlap the F2F with them?
- ◆ It is thought that security TEG will hit data security issues; how to do overlap? ML: Don't want one TEG to come up with ideas that are fairly incompatible with another TEG.
- ◆ DD: The split between storage and data management could be used to collect the experiment data management (user side data access) and the storage service operation (sites) respectively. Suggest to meet (virtually) with storage TEG chairs to prepare split/overlap topics.
- ◆ Will pick exact date in Jan. No interest in Dec. Amsterdam or CERN is acceptable.

From Philippe via email:

All LHC experiments have built their own internal DMS on top of what was existing at the time (storage systems and DM middleware). Indeed the requirement for DM middleware was going through a series of workshops, working group discussions, and despite that the result is not always at the level of the expectations...

After several years during which "the Grid" was essentially considering WMS as important, the emergence of a need for DMS tools was taking place during the Mumbai WLCG workshop in February 2006. This was the first time experiments expressed the need to have different storage classes as well as having a storage abstraction that could be used by DM tools and would include these "new" requirements. SRM was an obvious (the only?) candidate at the time, but it required extensions to its specifications. A working group was formed with all stakeholders (not all showed up), that took several months to come up with specifications for the famous SRM v2.2. The problem was that, at that time, storage systems had already been designed and were in production at sites and were not necessarily taking into account the requirements from SRM v2.2, which is normal as they didn't exist. But even then when they were written down and accepted, rather than adapting the storage systems to the new specs, the new specs were interpreted in the light of the possibilities of the existing storage systems (except for DPM and StoRM). The absence of a unanimous interpretation of SRM spaces for example (is a file in a space, or is it only "put in a space", can one "get a file" in a space?) may look like details, but have a strong influence on the operations of both the sites and the experiments. Another concrete example is the lack of a method for changing files from space, in particular changing service class for a file (e.g. from a T1D1 class to a T1D0 class) without replicating it.

A second attempt was made a few years later with the SRM v2.2 extensions MoU which, although agreed and signed by all parties was never implemented in all systems (e.g. data protection with SRM ACLs).

Concerning DM tools themselves, although it had been agreed that gfal was *the* low level library exposing SRM functionality, not all SRM functionality was exposed, but also not all experiments are using gfal (similarly for lcg_utils). The fact that, although they had been developed within WLCG at CERN these tools became part of the EGEE middleware, and therefore were subject to prior approval by EGEE technical bodies made it very difficult for getting new functionality quickly implemented and more importantly released, even when the implementation was very easy. It seems that with EMI the situation is even worse as new developments are taking place even without consultation with the users' community (gfal2, replacement of lcg_utils). FTS re-engineering has been postponed now for almost one year despite it had been presented and discussed with the experiments.

The expectation for the TEGs is that similar situations will not be reproduced: the TEGs are WLCG entities whose aim is to define the required evolution of systems and tools for the WLCG. It is therefore up to the LCG management to make this evolution happen, independently of whether they are considered of importance by the middleware development consortiums (EMI in particular). It would be a pity that the adaptation to WLCG needs of the middleware provided by third party developers requires more resources than the development itself. It took several years to reach a modus vivendi with the gLite development team in order to have the DM tools/libraries available for testing and even production usage before the middleware was properly certified and released. It was surprising to see that it now requires a lot of effort for achieving the same thing 18 months after the start of the EMI project.

Homework:

- Pick out F2F meeting times. Shooting for January and co-located with Storage.
 - Can people identify what other TEGs they sit on? Would like to make sure there are "ambassadors" to other TEGs to watch for overlaps.
 - Identify some concrete questions to ask experiments to prevent "generic needs" presentations.
 - Try and get at least CMS, and maybe ALICE, to answer questions at the next meeting.
-

This topic: LCG > WlwgTegDmKickoff

Topic revision: r5 - 2011-11-16 - DirkDuellmann



Copyright &© 2008-2020 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback