

# Table of Contents

Data Management TEG Topic List.....	1
-------------------------------------	---

# Data Management TEG Topic List

We have been given the following topics by the WLCG MB (some aspects expanded on by Brian Bockelman; later additions from SM annotate what overlap we see between Storage and DM)

- Review of the Data Management demonstrators from summer 2010. (DM/SM OVERLAP. Most topics in Amsterdam actually covered the DM (higher) layer. But many of them have implications in the SM, also, and some are actually infrastructure-related.)
- Dataset management: Currently, the common tools operate at the "file level" (file transfer, file catalog), oblivious to the fact that each experiment has built a custom dataset mechanism on top of them. What commonalities could be extracted? Is it possible/wise/necessary for the WLCG to play some role at the dataset level? (DM)
- Strategies for data federations in the WLCG. How do on-demand / caching architectures (c.f. ARC or Xrootd) fit into the larger WLCG data management ecosystem? (DM/SM OVERLAP. Enormous implications on SM, but DM could probably take a lead here, and in a latter part we could step in, e.g. how would you manage the storage implications. We would encourage the DM TEG to discuss and clarify things earlier on this wrt other topics.)
- Wide area transfer protocols. GridFTP has been the "workhorse", but it has shown significant limitations: the striping mechanism is a nightmare for disks, and it inherits design issues from FTP that cause it to not work well with NATs. Recently, HTTP and Xrootd have been suggested as replacements. (DM/SM OVERLAP. But, again, probably, DM can take the lead here.)
- Future strategies for FTS. FTS is again a workhorse for most of the experiments. How do we recommend it evolve in the future? Note: we can probably ask the FTS developers to come and present at one of our meetings. (DM/SM OVERLAP. FTS devs we agree should be invited ideally to a joint meeting.)
- What are the evolving requirements for data accessibility and security? I believe ATLAS/CMS/LHCb depend on the 75 sites to each individually enforce the correct experiment-internal access policies to their data, while ALICE's model delegates the internal access policies back to the experiment. How pleased/displeased is each experiment, and is there an opportunity for "cross pollination"? (DM/SM OVERLAP. But we should understand what the Security TEG does on this.)
- POOL: To my knowledge, ATLAS is the remaining user of POOL. Is it possible to relabel it an experiment-specific piece of software? (DM)
- ROOT, Proof? How do these lower-level frameworks intersect with the WLCG, if anywhere? (DM/SM OVERLAP. We can discuss in the joint Dec meeting.)
- Namespace management. Each experiment does namespace management very differently; this is often a tripping point in cross-experiment discussions (as an example, CMS does not use GUIDs and LFN<->PFN mappings can be done in constant time without a database-based catalog). Can we outline at the "philosophical" level what each experiment uses? (DM/SM OVERLAP)
- Future directions of the LFC. How is it deployed, what features are used? What are experimental needs in the future? (DM)

From the set of Storage Management drivers from the MB, we found one with possible overlap:

- Storage system interfaces to Grid (SRM future?). Interoperation (DM/SM OVERLAP. We need a list of what we need as from SRM functions. Joint discussion. Maybe also encourage all expts to have a team working on cloud tech.)

Please put any additional suggested topics below, and leave a signature so we know who added it.

---

This topic: LCG > WLCGTEGDataManagementTopics  
Topic revision: r1 - 2012-01-12 - BrianBockelman



Copyright &© 2008-2021 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.  
or Ideas, requests, problems regarding TWiki? use [Discourse](#) or [Send feedback](#)