

**NOTE: the working area for the joint DM/SM-TEG report can be found here:**

- <https://twiki.cern.ch/twiki/bin/view/LCG/ReportDataStorageTEG>

Please note that it is a DRAFT, being edited still. Different versions can be found in the "attachments" part of the twiki. Make sure you pick the latest one. Check the NEWS section on top.\*

---

This twiki contains a list of topics across Data and Storage Management TEGs listed separately but with an indication of overlaps.

## Grouped topics with links to recommendation documents

### **SM.1 Experiment I/O usage; SM.5 LAN protocols and SM.2 Requirements and evolution of storage systems**

Documents at SMIOlanStorageEvolution.

### **SM.3. Archive / Disk Seperation**

Documents at SMArchiveDisk

### **SM.4. Storage Interfaces: SRM and Clouds**

Documents at SMSrmClouds

### **SM.7 "Site-Run Services"**

Documents at SMSiteRunServices

### **SM.6 Security**

Written up with Security TEG: Info at AAIONStorageSystems

## Data Management

### **DM.1 Review of the Data Management demonstrators from summer 2010.**

DM/SM OVERLAP.

### **DM.2 Dataset management and Data placement (policy-- based or dynamic)**

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.3 Data federaion strategies**

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.

**DM.4 Transfers and WAN access protocols(HTTP, xrootd, gsiftp)**

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.

**DM.5 Data transfer management (FTS)**

FTS is again a workhorse for most of the experiments. How do we recommend it evolve in the future? Note: FTS developers could come and present at one of our meetings.

DM/SM OVERLAP.

**DM.6 Understanding data accessibility and security requirements/needs**

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.

**DM.7 POOL**

To my knowledge, ATLAS is the remaining user of POOL. Is it possible to relabel it an experiment-specific piece of software?

DM. But maybe that's an internal ATLAS / CERN-IT discussion.

**DM.8 ROOT, Proof**

How do these lower-level frameworks intersect with the WLCG, if anywhere?

DM/SM OVERLAP. We can discuss in the joint meeting.

**DM.9 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. Could be asked in the joint questions above.

**DM.10 Management of catalogues (LFC, future direcions)**

Future directions of the LFC. How is it deployed, what features are used? What are experimental needs in the future?

DM "ONLY"

# Storage Management

## **SM.1 Experiment I/O usage patterns**

And so performance requirements for storage. I/O Scalability limits.

Overlaps with DM8.

## **SM.2 Requirements and evolution of storage systems**

What is needed by experiments from storage systems and how that will evolve; together with how storage will evolve independent of us

SM Only

## **SM.3 Separation of archives and disk pools/caches**

SM. We should agree on the archive/disk split as a strategy, maybe that's a DM kind of questions to experiments. But once we have the reply, and probably it's a yes, it's a SM item.

## **SM.4 Storage system interfaces to Grid**

Future of SRM.

Usage of "Cloud" storage.

Interoperation

DM/SM OVERLAP. We need a list of what we need as from SRM functions. Joint discussion. Maybe also encourage all exps to have a team working on cloud tech.

## **SM.5 Filesystems/protocols (standards?)**

SM

## **SM.6 Security/access controls**

same comment as for DM.6 above.

## **SM.7 Site-run services.**

Storage management interfaces, performance measurements, monitoring, manageability.

End user experience.

Is there also something here (?) on management - storage accounting; roadmaps/ communication etc.

SM.

-- WahidBhimji - 11-Jan-2012

---

This topic: LCG > TopicsDataStorageTEG

Topic revision: r8 - 2012-03-28 - DanieleBonacorsi



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