



WLCG Data Steering Group

Report and discussion

Oliver Keeble, CERN

From the WLCG MB

last GDB about data management activities within WLCG. The following comments were noted:

- Ian Bird observed that there are many activities ongoing within WLCG in the data management area. There are also opportunities for EU funding that are coming up in the near future and it would be useful to revisit activities like the Zephyr proposal in this context. He stated that he would strongly support having a group of people to coordinate these activities and suggested that it would be useful to prepare a draft mandate for this before the pre-GDB in September. Oliver Keeble agreed and stated that he would follow up. Ian Collier expressed his support for this idea and suggested that something like a twice-yearly pre-GDB on data management could be very useful. Torre Wenaus agreed that this is a critical area where communication and keeping an eye on open opportunities are very important. Stefan Roiser noted that this would especially be useful to follow up on the commonalities between the experiments, the move to SRM-less sites being one example. Frank Wuerthwein noted that there is one site in the US (Nebraska) that has no SRM anymore. Ian Bird concluded that the proposal was endorsed by the WLCG MB and should now be implemented.

Action: Oliver Keeble to prepare a draft mandate for a WLCG data management coordination team before the pre-GDB in September.

5. Information System

Original Proposal

- A data steering group is created
 - One person per experiment, storage provider, 2 or 3 for infrastructure
 - No limit on group lifetime
- This steering group
 - Defines long term WLCG strategy for data management
 - Tracks relevant technical themes
 - Meets regularly to define priorities
- And on that basis oversees
 - pre-GDB slots
 - Triggering of task forces if necessary
 - WLCG data discussions in other forums, e.g. the WLCG Workshop data session
 - Reporting to WLCG ops meeting where requested.
 - Escalation and reporting to MB
- The steering group is not itself a decision-making body

The result

- The group has
 - Met 4 times now
 - Collected input from the stakeholders
 - Aggregated the input, categorising under a set of 7 strategic aims
 - Made a first prioritisation of these aims
 - Made a first recommendation about what to do about them
- Full details, agendas, mandate etc
 - <https://twiki.cern.ch/twiki/bin/view/LCG/WLCGDataSteeringGroup>

Representatives

- Alice - Latchezar Betev
- Atlas - Martin Barisits
- CMS – Brian Bockelman
- LHCb- Philippe Charpentier, Christophe Haen
- Infrastructure
 - Alastair Dewhurst
 - Michel Jouvin
 - Pepe Flix
 - Frank Wuerthwein
- Operations
 - Julia Andreeva
- Storage
 - Castor, dCache, DPM, EOS, StorRM, XrootD

More input please...

- Everyone is encouraged to give input to their representative, or right now...
- Priorities for topics to address at the community level

Strategic Areas

Infrastructure diversity

Promote the growth of the infrastructure through resource diversity.

- Cloud storage, object stores
- HPC Data management
- Colocation with other communities
- Adapting workflow to resources
 - -> QoS

Common global strategy

Articulate a common vision to guide strategy

- Can a common vision be described?
 - How could we get this? CWP?
 - A reference model under iterative refinement?
 - Identify which higher-level services are candidates for common projects
- Sustainability and collaboration
 - SKA
 - Globus project

Quality of Service

Exploit resources in the most effective way through “QoS”, understanding how different classes of service (and therefore cost) are appropriate for different workflows

- Parameter space – where are the gains?
- Role of tape beyond archive
- Archives backed other than by tape
- Cache hierarchies
- Tuneable reliability

Infrastructure Architecture

Understand how higher level architectural changes can be exploited

- Multi-site storage
- Regional federations
- Caches
- Diskless sites
- -> Networking

Operations

Reduce costs through making storage systems and data management operations easier and more transparent

- Storage accounting, space reporting
- Monitoring
- Topology system, CRIC

Analysis

Enable future analysis models by providing the relevant services

- Access paradigms
 - Interfaces to storage systems
 - Queries, hints.
 - Event level access
- Data formats
- -> CWP analysis group

AAI, Federated Identity

Ensure that data management systems, and storage in particular, is ready for any new authentication and authorisation infrastructure

- Edugain
- OpenID Connect
- Web/non-web workflows
- GSI authentication

Priorities

- Infrastructure diversity
 - In particular, object stores (Atlas, CMS, xrootd)
- Archival storage implementations (disk/tape) and their optimal exploitation (Atlas, Alice, Castor/EOS)
- Caching storage (Atlas, CMS)
- post-X509 AAI, Federated identity (longer term)

Next steps

- Dedicated meetings on the priorities
 - Next data pre-GDB – Sept 12th (tbc)
 - Resource reporting
 - Tape/archive exploitation
 - Object stores
- Cross-cutting discussions
 - AAI
 - Networking?
- Exchange forum?
 - Periodic f2f to let prototype teams share experience