

Closeout: Data Access, Organisation and Management

HSF Workshop, Annecy, June 2017

Workshop Discussions

- Tuesday session was the main occasion to meet as a group since San Diego
- Concentrated on data organisation: what is the appropriate granularity to organise the data
 - Current dataset-based organisation has drawbacks for writing and reading
 - Event-based organisation would probably improve the write performance but not a real improvement for analysis
 - Analysis would probably benefit from a “column” based organisation rather than the current partial read approach
 - An area that requires R&D to understand pro and cons of different approaches and the impact on data management
- Other emerging topics: query-based analysis, “virtual datasets”
 - May impact the storage/CPU balance if we reduce the amount of static data

Path Forward

- CWP chapter in a shape good enough to be reviewed
 - https://docs.google.com/document/d/1_YKGs8waBidpS8akmxEVjV00DThINbJAuVelzk2uHg/edit#
 - No major controversial issue remaining, mainly edits
- Missing part is the roadmap: no real time to discuss it during the workshop
 - Concrete R&D actions have to be decided and prioritized
 - Granularity-related R&D has to be coordinated with Analysis WG and Reconstruction WG
 - Object-store technology is another topic identified for the future R&D: potential benefits, impact on cataloguing
- One topic to clarify: possible impact of ML on data access
 - Contradictory assertions that it could increase or decrease data iteration
- Next month should allow to have a good draft of the roadmap and a finalized chapter