

SAR @ OGF30 (brief and selected summary)

Jon presented some slides.

- EMI has signed up to deliver a storage record definition and implement it.
- Disclaimer: EMI will not make OGF UR2, but will contribute to it.
- Definition needed within January 2011 - making an OGF document available for public hearing should be enough for EMI.
- EMI has sent out requests for storage accounting, got a little feedback - some feedback may be out of scope (per-file access to assess need for new storage space).
- OSG has some extensions to OGF UR (StorageElement, StorageElementRecord), but that may be related to CPU accounting, not SR

Henrik presented some slides.

What to measure:

- We need used space
- Reserved and available space not so clear
 - ◆ Tricky to measure
 - ◆ They're optional anyway
- Should perhaps have total space as well
- Forget about transfer statistics if you want to meet deliverable deadline (nice to have as a future item)
- Focus on what is actually used - SpaceUsed, SpaceAllocated and SpaceReserved, drop SpaceAvailable

Measuring Space

- use is continuous, records are discrete -> need some way of conversion
- should probably replace start+end with timestamp+duration

Measuring Transfers

- somewhat tricky to measure?
- many use cases where you only want to measure storage or only transfers -> should be separate from storage part (storage transfer?)
- the owner of a file and the identity performing a transfer, may not be the same, so the owner of the file should not be accountable for the transfer. Hence a separate record for transfer is probably preferable.
- transfers out of scope for SR

Concepts

Where

- Maybe too complex? StorageElement is a bad word! GLUE uses StorageShare(?)
- Structure of where-block should be flat - hierarchical is messy

What

- Filenames should probably be avoided - privacy issues (filename may contain person id)
- Per-file records should be possible - if someone wants to publish millions of records, they shouldn't be stopped
- NDFG is medium sized WLCG site and has 11M files (not all files will be changed every day though)
- Can always aggregate down the size anyway

Who

- Should adhere to CPU accounting part
- CPU accounting currently has only Global ID and Local ID
- More or less everyone using CPU accounting has extended Global ID with VO data
- SR **needs** VO info, though with better name, like project or group, to appeal to non-grid infrastructures
- Solution: SR includes project data, UR follows up and adds project data

UR2 or separate format

- UR2 will not be ready for EMI deadline
- Make a standard for ID block and use it in SAR

Some points from discussions

EMI deliverable can at best be "putting definition out on public commenting"

SpaceUsed / SpaceReserved / SpaceAvailable looks messy need to adhere to GLUE definition from OGF perspective it is important to have something anyone can use, not just grid people

We cannot do peak measurements at the moment, but maybe something for the future?

Need good argument why storage accounting is taken out of UR

- this is a usage record driven by a storage accounting need. (Does anyone remember the great argument that was proposed?)

If we do need political correctness, we could reuse the top-level element in the UR standard, and have a subelement below it.

The SR format definition is a very nice step forward. Now we need to write it up in a clear and concise way, discuss it on mailing list and put it on public hearing as soon as possible.

-- JonKerrNilsen - 01-Nov-2010

This topic: EMI > EmiDataStorageAccountingSummary20101028

Topic revision: r1 - 2010-11-01 - JonKerrNilsen



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