Storage accounting in EMI and OGF

Jon Kerr Nilsen
EMI JRA1 deputy
Dept. of Physics, Univ. of Oslo

EMI StAR

- EMI has created a proposal for a storage accounting record (StAR)

- Finalised in June this year

- Final version: http://cdsweb.cern.ch/record/1352472?ln=en

- Will be implemented by EMI storage providers

What StAR is not

- To be able to arrive to a useful record in time we had to limit the scope

- Many interesting storage-related features to record
  - I/O (like in Amazon S3)
  - Metadata access (stat,ls)
  - Available space
  - File name
  - …

- Scope of StAR is limited to consumption of storage space

DILBERT by Scott Adams

What StAR is

- Definition of usage record for storage

- Inspired by OGF UR1.0

- Summarises used space

- XML-based schema

- Non-overlapping/non-contiguous records means no space used (ValidDuration configured by sys-admin)

StAR structure

- Resource
  - Fields describing the system the resource was consumed on
  - `StorageSystem`, `StorageShare`
- Resource consumption
  - How much of the described resource has been used
  - `ResourceCapacityUsed`, `LogicalCapacityUsed`
- Consumption details
  - Fields describing what is consumed
  - `StorageMedia`, `StorageClass`, `DirectoryPath`, …
- Identity
  - Describes the person or group accountable for the consumption
  - `SubjectIdentity` block (`UserIdentity`, `Group`, `GroupAttribute`, …)
- Record identity and duration

EMI StAR timeline

- May 2011: StAR definition agreement
- Oct. 2011: Each SE to provide an implementation plan
- May 2012: Each SE to implement accounting sensors
- June 2012: Accounting record publishers (APEL and/ or DGAS) to publish storage records
- December 2012: Each SE to test and deploy their accounting sensors (for EMI 3 release)
- April 2013: Accounting sensors released in EMI 3

EMI INFSO-RI-261611

Each SE to provide an implementation plan

- Need to figure out what to publish
  - Bad approach: Publish only required fields and trivially available fields
    - Might be good as a starting point, but probably not very useful
  - Better approach: Publish the fields which are interesting for end-users
    - Who are the end-users?
  - Best approach: Ask EGI (nicely) which fields are interesting and publish those
    - EGI knows the user communities
    - Accounting is about infrastructure, so is EGI
    - If EGI was to come with an EGI profile of StAR, we would probably jump right on it
- Need to figure out how/where to publish
  - SEs only responsible to make records available – accounting sensors
  - Accounting clients (APEL/DGAS) responsible for publishing
- Need to synch what the SEs will publish

OGF33 accounting session

- "Next Generation Usage Records" OGF session yesterday
- A bit of technical discussions around storage accounting
  - Available space
    - Not covered by StAR, but clearly use-cases for it
    - Not really usage - but closely related semantically
    - Capacity record might be needed
  - TimeDuration vs ValidDuration
    - ValidDuration similar to TTL, says for how long record can be trusted
    - TimeDuration gives time since last measurement
    - Both are related to accounting, not to usage

OGF33 accounting session

- Discussed how to move on towards UR2.0
  - Currently definitions for CPU (UR1.0) and storage (StAR)
  - Work ongoing in EGI for accounting VMs, DaaS, application use and MPI
  - Decided to go for one definition for all resources
  - Important that XML schema can be easily extended when new resources are added
  - Could have one empty schema just stating that "this is a usage record"
  - then extend it with resource record and make sure similar elements have the same name

- How to move on
  - Clearly more discussions needed
  - Aim to arrange biweekly telcons to keep process roling

Thank you!