



Enabling Grids for E-scienceE

SA1 - All Activity Meeting

Ian Bird, Markus Schulz, Nick Thackray
19th May 2005

www.eu-egee.org



- **Status update on the certification of gLite.**
- **Status update on the deployment of the Pre-Production Service**
 - How will the applications be given access to the PPS.
 - Who/which team will be giving support?
- **What's missing in order to deploy gLite services on the production service? What are the steps require to achieve it?**
 - for each gLite service.
 - Explanation of the hybrid (i.e. gLite - LCG-2) deployment strategy, layout and plan.
- **Future LCG-2.x releases and relationships with gLite services.**
- **What are the LCG Service Challenges (goals, which sites are involved, which middleware components are relevant, schedule etc.)**
- **Status and information of the "Baseline Service Group**

- **Tested the deployability, functionality and resilience of gLite version 1.0**
 - Version 1.0 of gLite was not passed as certified because:
 - No data replication mechanism
 - Configuration of the VOMS server was restricted and very unreliable when using the supplied configuration scripts
 - Hard coding of service end-points (particularly in CEs and WMSs) meant that the middleware was not scalable to a production service
 - Failure rate of “hello world” jobs is 5-10% (with zero retries)
- **Now deploying gLite version 1.1 on the certification testbed**
 - Release received on Friday 13th May
 - Contains a fix for the VOMS server (to be tested)
 - Contains a partial solution for data replication (to be tested)

- **Roll-out of the pre-production service is in 2 phases:**
 - Phase I is the installation of a primary set of core services (WMS, VOMS server, L&B, etc.)
 - Sites involved are CERN, CNAF, NIKHEF, PIC and CESGA
 - Phase II is the inclusion of the remaining sites (~10 sites) plus the installation of secondary instances of core services.
- **Phase I is almost complete (using gLite 1.0)**
 - All core services are installed and final testing is under way.
- **When phase II is complete, VOs will be given access through VOMS servers shared between production and pre-production.**
- **Support is initially “self-help” but will move to the same teams which provide production support.**

What is missing?

- **The File Placement Service was not included in gLite 1.1**
 - This is critical because the FPS is responsible for implementation of the security model, without it files can't be replicated AND accessed via gLite-I/O, in addition the traceability is problematic (site boundary)
 - The FPS takes the users replication request, verifies that she is authorized to replicate the file, hands the request to the FTS, which transfers the file using a service certificate. At the end the FPS registers the replica with FIREMAN, including the metadata that controls the access privileges. The service certs are mapped to the gLite-IO service, which owns all the files
 - Note: In the SC3 we use the FTS directly together with the user's proxy. This works, because it is inline with the LCG-2 security model
- **Service discovery via R-GMA (a.k.a. Information System)**
 - Services (CEs) have to be registered with the workload management services, or vice versa (both for push & pull), gLite-IO and SRM are linked statically, WNs are linked statically to a gLite-IO server
 - There is no working information system in gLite (R-GMA works, but the services can't use it!)
- **Manageable, secure distribution of the various service certificates**
 - gLite I/O and the FTS relies on service certificates. These have to be known by the SRMs, FIREMAN, gridFTP and glite/I/O instances.
 - For the full production system we have to manage a few hundred of those certs.
- **Moving the LCG VO/VOMS services to the RFC compliant VOMS version**
 - LCG-2 Vos currently work with an incompatible version of VOMS
- **Monitoring, audit file processing**

What are the steps required to achieve it?

- **The File Placement Service was not included in gLite 1.1**
 - “Solution” is to ignore for the time being that everyone can read any file in the world and place replicas everywhere (even on sites that do not support his VO). Then the user can use the FTS in gLite mode and access them via glite-I/O. The user has to register the replicas explicitly.
 - The correct solution is to get the FPS
- **Service discovery via R-GMA (a.k.a. Information System)**
 - The WLM system can use the BDII
 - Interface the systems to R-GMA and/or the BDII
 - The static configuration of the data management is cumbersome, but can be done, however it is laborious and error prone.
- **Manageable, secure distribution of the various service certificates**
 - A VO like service, with registration services in the regions needs to be setup and some config scripts are needed to keep the accepted DNs active. It is not clear that all services can handle dynamic DN lists....
- **Moving the LCG VO/VOMS services to the RFC compliant VOMS version**
 - The software is available and the servers can be upgraded at any time
- **Monitoring, audit file processing**
 - The SFT has been ported
 - GIIS mon and GridICE has to move
 - Audit file processing needed...

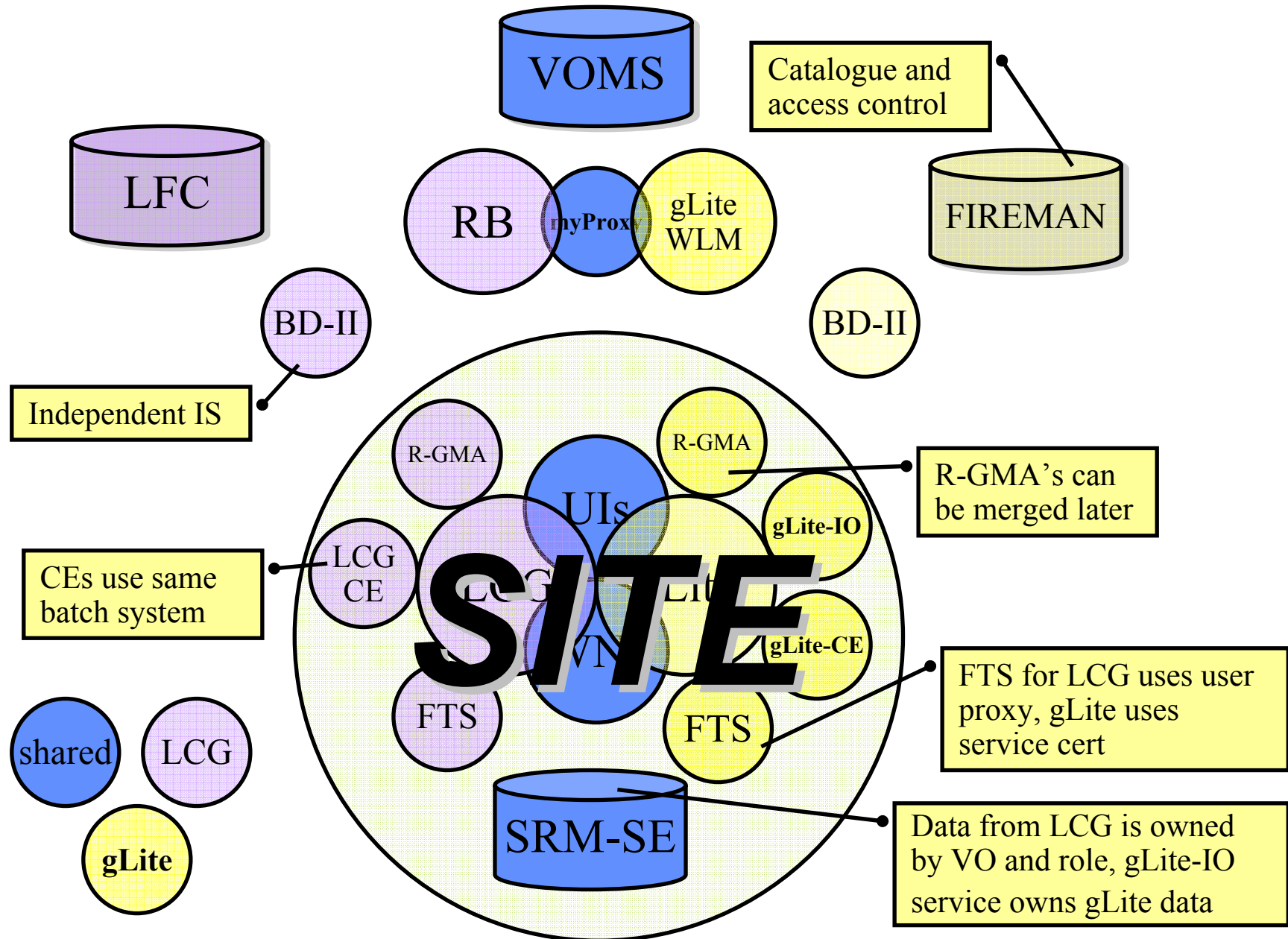
- **Motivation**

- No resources for all activities
 - SC3 + certification + preproduction + production + LCG-2/gLite component mix
- The largest 10 sites provide 3/4 of the resources
- Same scale in terms of sites as the initial preproduction service, but production scale resources.
- Significant overlap with preproduction sites
 - Almost no additional work needed compared to preproduction service

- **How?**

- The UIS&WNs are setup with LCG-2 and gLite clients
 - Users environment is setup depending on the CE used
 - Based on LCG-2_4_0 technology
- The myProxy service and the SRM-SE are used by both systems. The service certs are added to the SRM-SEs gridmapfile
- All other services (CE, WLM, BDII, R-GMA) are deployed for both services individually

Extended Pre-Production Model



- **Deployment Strategy**
 - The largest 10 sites stop participating in the current preproduction
 - LCG and gLite released independently
 - gLite with a higher frequency
 - RPMs managed by APT
 - Config management native for gLite and LCG services
 - Config for WN+UI native or tar-ball + YAIM
 - Smaller sites (90%) follow LCG-2_X release(s)
 - Switch to gLite on VO demand
- **Plan**
 - As soon as the gLite passes fundamental tests
 - Best for the June release
 - Release with list of problems (extended pre-production!)
 - BugFix releases minimizing the load on the admins
 - *Clients in user space*
 - *Services via APT update*

- **Why?**
 - Need to deploy components needed in the SC-3
 - gLite not ready for SC-3
- **When?**
 - Early June
 - 1/2 month earlier to allow ramp up of SC3
- **What will be in?**
 - United proxy extension for VOMS
 - New version of the GLUE schema (compatible)
 - File transfer service
 - LFC production service
 - Interoperability with GRID3
 - User level stdio monitoring
 - Bug fixes as always
- **Who has to upgrade?**
 - Tier 1s and the Tier 0 as quickly as possible
 - Tier 2s at their own pace

- **Why?**
 - Give gLite time to mature
 - It is still in use...
- **When?**
 - Every 3 month
- **Relationship with gLite?**
 - Integrate “replacement” services as soon as performance AND stability exceeds current services
 - Give more experience with the service

- **The current small scale pre-production service**
 - Limited access to individual users until the extended preProduction service becomes available
- **Extended Pre-Production**
 - Every member of a VO supported by at least one of the tier1 centres



LCG Service Challenges - Overview

- LHC will enter production (physics) in April 2007
 - Will generate an enormous volume of data
 - Will require huge amount of processing power

- LCG 'solution' is a world-wide Grid
 - Many components understood, deployed, tested..

- But...
 - Unprecedented scale
 - Tremendous challenge of getting large numbers of institutes and individuals, all with existing, sometimes conflicting commitments, to work together

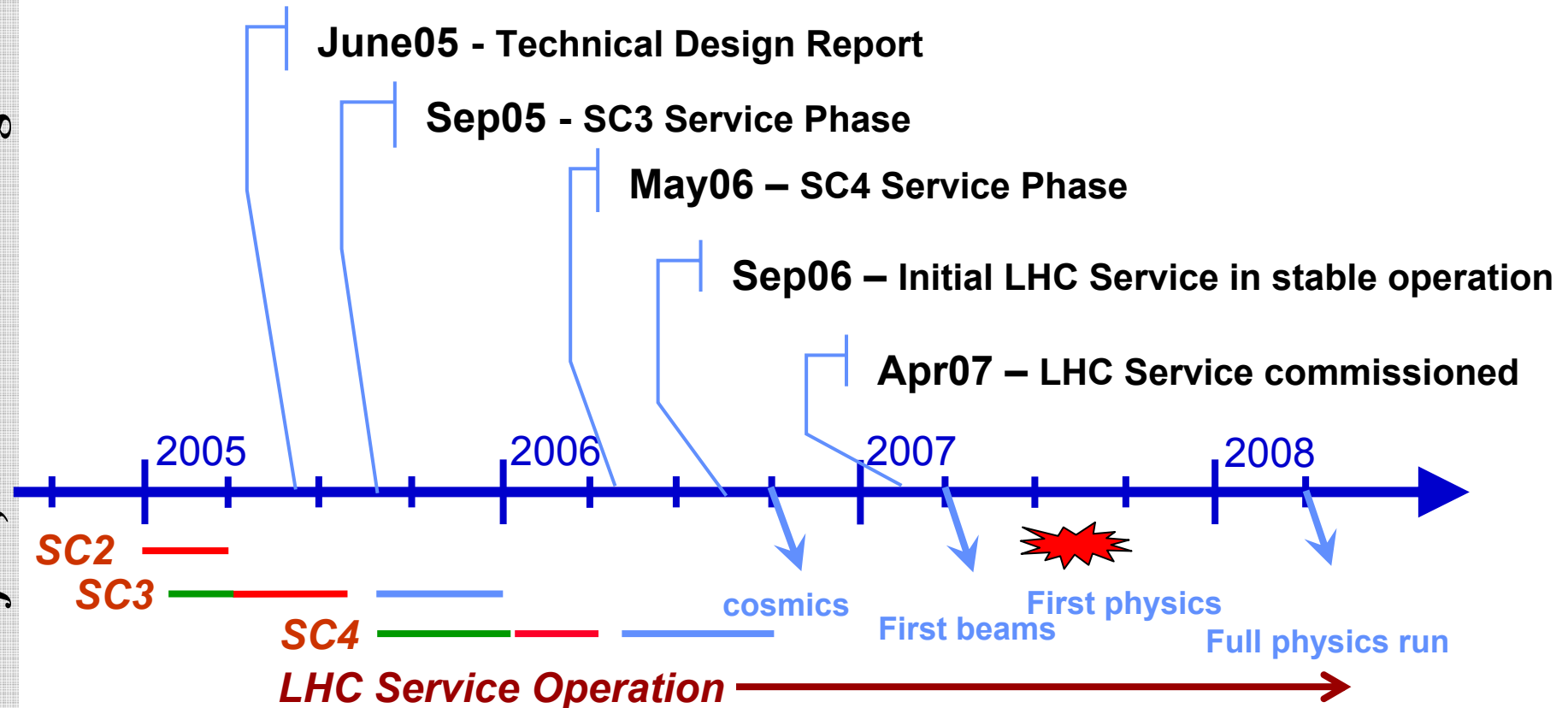
- LCG must be ready at full production capacity, functionality and reliability in less than 2 years from now
 - Issues include h/w acquisition, personnel hiring and training, vendor rollout schedules etc.

- **Should not limit ability of physicist to exploit performance of detectors nor LHC's physics potential**
 - Whilst being stable, reliable and easy to use

Service Challenges - ramp up to LHC start-up service



LCG Project, Service Challenges



- SC2** - Reliable data transfer (disk-network-disk) - 5 Tier-1s, aggregate 500 MB/sec sustained at CERN
- SC3** - Reliable base service - most Tier-1s, some Tier-2s - basic experiment software chain - grid data throughput 500 MB/sec, including mass storage (~25% of the nominal final throughput for the proton period)
- SC4** - All Tier-1s, major Tier-2s - capable of supporting full experiment software chain inc. analysis - sustain nominal final grid data throughput
- LHC Service in Operation** - September 2006 - ramp up to full operational capacity by April 2007 - capable of handling twice the nominal data throughput

Why Service Challenges?



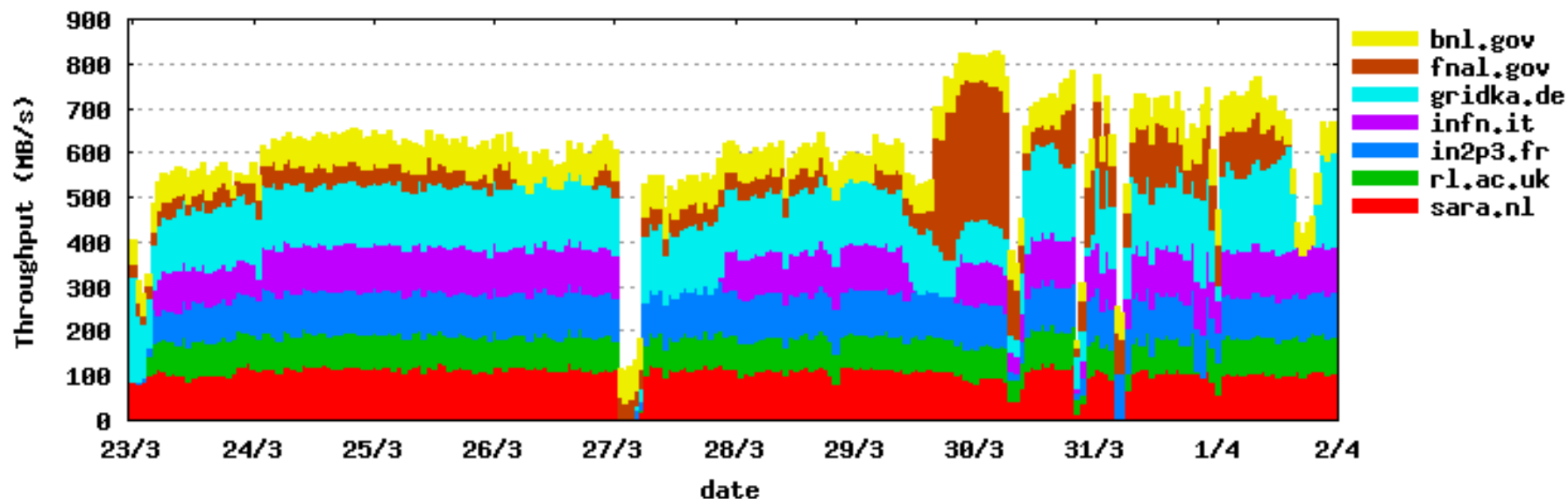
To test Tier-0 ↔ Tier-1 ↔ Tier-2 services

- **Network service**
 - Sufficient bandwidth: ~10 Gbit/sec
 - Backup path
 - Quality of service: security, help desk, error reporting, bug fixing, ..
 - **Robust file transfer service**
 - File servers
 - File Transfer Software (GridFTP)
 - Data Management software (SRM, dCache)
 - Archiving service: tapeservers, taperobots, tapes, tapedrives, ..
 - **Sustainability**
 - Weeks in a row un-interrupted 24/7 operation
 - Manpower implications: ~7 fte/site
 - Quality of service: helpdesk, error reporting, bug fixing, ..
- **Towards a stable production environment for experiments**



SC2 met its throughput targets

- >600MB/s daily average for 10 days was achieved -
Midday 23rd March to Midday 2nd April
 - Not without outages, but system showed it could recover rate again from outages
 - Load reasonable evenly divided over sites (give network bandwidth constraints of Tier-1 sites)





Baseline Services Group: Goals

- Experiments and regional centres agree on baseline services
 - Support the computing models for the initial period of LHC
 - Thus must be in operation by September 2006.
 - The services concerned are those that
 - supplement the basic services
- Not a middleware group - focus on what the experiments need & how to provide it
 - What is provided by the project, what by experiments?
 - Where relevant an agreed fall-back solution should be specified -
 - But fall backs must be available for the SC3 service in 2005.
- scalability/performance metrics.
 - Feasible within next 12 months → for post SC4 (May 2006), & fall-back solutions where not feasible
 - When the report is available the project must negotiate, where necessary, work programmes with the software providers.
 - Expose experiment plans and ideas



Baseline services

- We have reached the following initial understanding on what should be regarded as baseline services

- Storage management services
 - Based on SRM as the interface
- gridftp
- Reliable file transfer service
- ✗ File placement service - perhaps later
- Database services
- Grid catalogue services
- Workload management
 - CE and batch systems seen as essential baseline services,
 - ? WMS not necessarily by all

- Grid monitoring tools and services
 - Focussed on job monitoring - basic level in common, WLM dependent part
- VO management services
 - Clear need for VOMS - limited set of roles, subgroups
- Applications software installation service
- From discussions added:
 - Posix-like I/O service → local files, and include links to catalogues
 - VO agent framework