



A human grid initiative.

Life Science Virtual Research Community Requirements

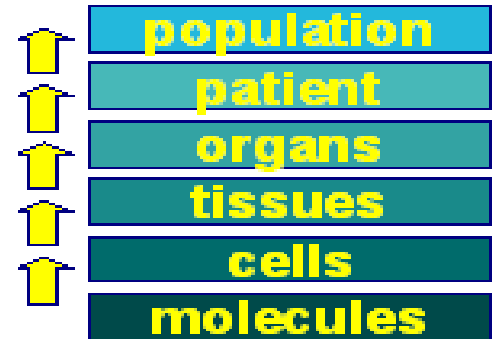
Session requirements, EGI technical Forum,
September 16, 2010

Johan Montagnat, CNRS

(on behalf of Samuel Keuchkerian, HealthGrid)

» Life Sciences covers a broad range of activities, ranging from biology to patient healthcare

» The multi-scales view (HealthGrid'03)



» Grid-empowered Life Sciences activity focuses on computational Life Sciences *and* distributed healthcare-related informations systems and databanks

» LS data is distributed, and highly sensitive

» LS community is widely spread (both geographically and in terms of research themes)

Grid adoption in Life Sciences

- » Very large potential
 - » Genome sequencing projects and databanks
 - » Distributed populations data sets involved in epidemiology
 - » Computational neurosciences
- » Grid adoption visible world-wide. Still a slow take up
 - » Real communities scattered and not used to real collaborations
 - » Medical data is sensitive
 - » Healthcare is a business
 - » Grids are difficult to use
- » ***The first critical requirement is a transparent grid***
 - » **Reliable**
 - » **Usable by non-expert end users**
 - » **With simplified access procedures**

Critical requirements

- » Reliable
 - » Concerns all services (special problems with LFC, VOMS and storage)
 - » Infrastructure monitoring to anticipate difficulties
 - » At VO-level
 - » End-to-end (across services): a job is not done as long as the result is not accessible to the user client
- » Usable by non-expert end users
 - » Adapted interfaces
 - » Reuse of generic tools
- » With simplified access procedures
 - » Lightweight governance model
 - » Lightened tooling
 - » Accompanying new users (self-adoption is rare)

VRC model being implemented

- » Grouping several VOs
 - » biomed, enmr, lsgrid, vlemmed
 - » including the generic “biomed” VO that enables international collaboration and large-scale computations
- » International Support Centre for Life Science grid users
 - » To act as a single, representative contact point for EGI
 - » To accompany users and deliver best experience
 - » At the interface between VOs and operations
- » Other benefits
 - » Reduce operational burden
 - » Prevents community scattering in a decentralized EGI nebula
- » Supported by several NGIs and projects
 - » Dutch, French, German, Italian, Spanish, Swiss
 - » EGI-Inspire, Lifewatch ESFRI
- » See https://dav.healthgrid.org/lsvrc/LSVRC_proposition_09-08-2010-final.pdf

Governance model

- » One coordinator (renewed every 6 months)
- » Sub-domain scientific leaders
 - » Sub-groups defined in biomed VO (as VOMS groups)
 - » Sub-groups may require specific resources (e.g. data access control)
- » Open to new users
 - » Registration into groups (sub-domain, projects, country)
- » Model to be refined
 - » Depending on Organisational Level Agreements
 - » OLAs between NGIs and VRCs should ensure adherence from NGIs with dedicated infrastructure and support for VRCs
 - » MoUs to be made with major grid infrastructures (EGI) and ESFRIs

- » Need to formalize agreements with resource providers
 - » Currently years-old informal agreements without guarantee and frequent decommissioning without notification
- » Work on-going on the definition of a users database
 - » To manage scientific affiliation and authorization
 - » To define user lists (per sub-domain, per project...) and mailing lists
 - » To maintain consistency (remove non-active / expired users)
 - » To follow-up users activity
- » Link users database with VOMS service and applications database
 - » Requires VOMS and application database interface

VO/VRC monitoring tools

- » VO/VRC monitoring tools required
 - » Monitor grid services operation at the VO-level (VOMS, LFC, CEs, SEs)
 - » Monitor VO production (CPU/data efficiency, number of failed jobs, data placement, etc)
- » Work in-progress to set up a Nagios server
 - » Specific monitoring probes to develop
- » Ad-hoc technical tools put in place by the LSVRC technical team
 - » Technical team on duty with periodic shifts reports
 - » Hudson integration server for infrastructure monitoring
- » Dashboard should be developed and used across the VRC

Technical team requirements

- » Missing / flawed information
 - » Available storage space on SEs (for monitoring)
 - » Number of CPUs available per site
 - » Decommissioning announcements
 - » Remove site in maintenance from BDII
 - » SE tags (OK vs faulty)
- » LFC and VOMS are single-points of failure (they do fail)
- » Missing critical data management services
 - » Data migration
 - » Improved LFC clients for VO management
- » Improving monitoring and error detection
 - » How to monitor transfers from SEs to CEs?
 - » Job submission (test WMS/CE pairs)

VO-specific services

- » Many services are developed to cover the community needs
 - » Workflow systems: Taverna, MOTEUR...
 - » Pilot job systems: WISDOM, DIANE...
 - » Data protection: Hydra encryption service...
- » Most tools have a much broader usability
- » An integrated, accessible user environment is still needed

Conclusions

- » Grid technologies adoption requires easy accessibility and reliability
- » The governance model is based on the coordination inside the VRC of a catch-all large-scale international VOs and more dedicated VOs
- » User management and infrastructure monitoring tools required
- » Most VO/VRC management needs are not specific to the LS VRC
 - » NA3 should help mutualizing the efforts between VRCs
 - » VO-level monitoring is needed ... for all VOs

Thank you