

gLite: role, objectives and migration plans to EMI

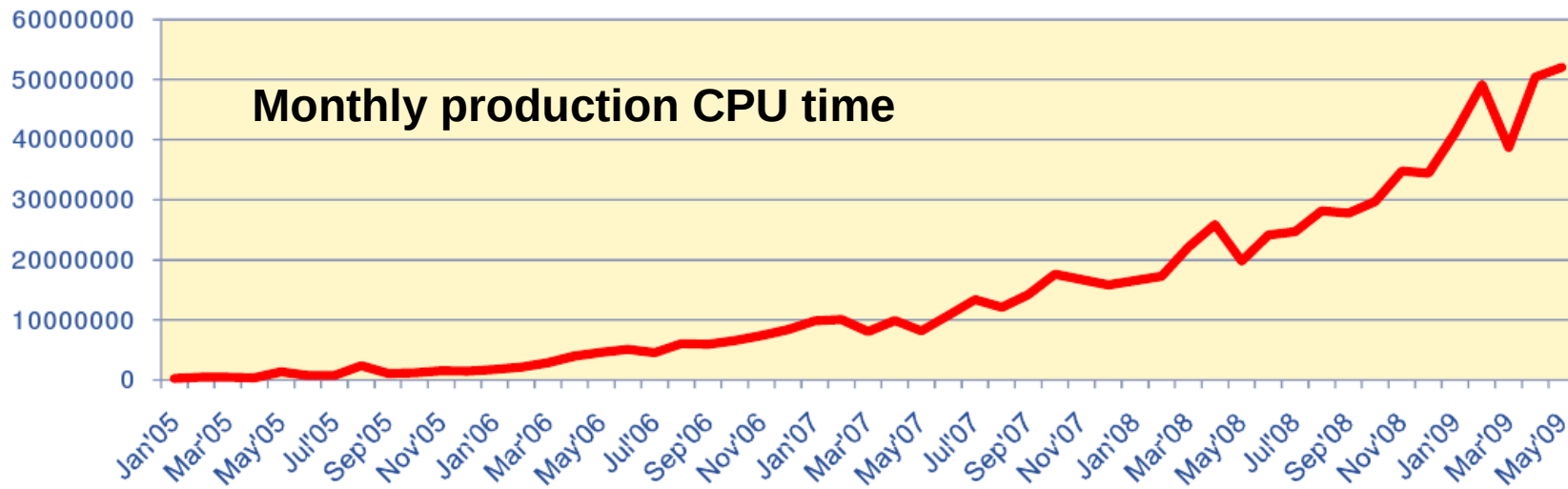
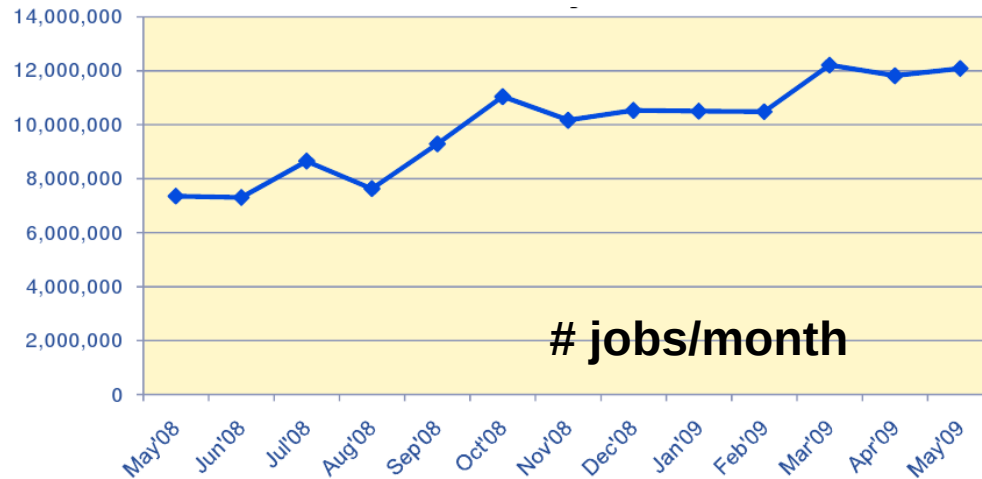
Francesco Giacomini
EGEE-III JRA1

EGEE'09, Barcelona, 21-25 September 2009

- **A distribution of middleware components that allows to build a large-scale production-quality Grid infrastructure, with a strong focus on security and interoperability with other infrastructures**
- **Several years of experience in coordinating a large and distributed software development project**

- **Sequence of projects**
 - European Data Grid (EDG) – 2001-2004
 - EGEE – 2004-2006
 - EGEE-II – 2006-2008
 - EGEE-III – 2008-2010
- **The main goal has shifted from developing middleware to building a production-quality infrastructure**
- **Process driven by the real needs of scientific communities**
 - Archeology, Astronomy, Astrophysics, Civil Protection, Computational Chemistry, Earth Sciences, Finance, Fusion, Geophysics, High Energy Physics, Life Sciences, Multimedia, Material Sciences
 - In particular High Energy Physics, whose needs were also (too) urgent due to the start of LHC
 - All projects led by CERN

- ~300 sites
- ~50 countries
- ~150K CPU cores
- ~28 PB disk, ~41PB tape
- ~13000 registered users
- ~200 VOs
- ~400 Kjobs/day



- **gLite is a distribution of software components designed and implemented to enable secure resource sharing on a large scale, i.e. to build a Grid**
- **The distribution is provided by the EGEE project**
- **Most of the components included in gLite are developed within the project itself**
- **Some components are taken from third parties, integrated with the rest and redistributed**
 - e.g. globus, condor, dCache

- Slightly adapted from the OGF PGI WG
- Given a **job description**, a **meta-scheduler** interrogates the **information system**, locates an optimal **execution resource**, which in turn runs it, fetching the necessary **input data** from a **remote storage**. Upon completion, the newly created data is uploaded to a storage resource, registered in the necessary **data indexing catalogs**, and the job record is updated in the **accounting and monitoring system**
- Other use cases are possible
 - e.g. *pilot jobs*, to create an overlay network of agents under the control of the application framework

User Access

General Services



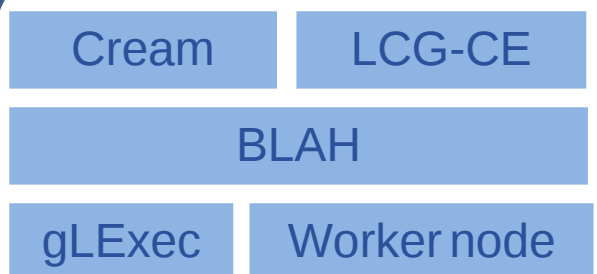
Security Services



BDII

Information Services

Compute Element



Storage Element



- **Finding resources and performing operations on core services by hand is at best inconvenient**
- **Higher-level services are needed**
- **As a general rule a high-level service accepts a logical description of what the user wants to achieve and transforms it in a concrete operation, following it until completion**
- **A high-level service operates on behalf of the user**
 - **Credential delegation/impersonation**

- **The interaction with services is typically asynchronous**
 - An id/token is returned to the user as soon as the service has accepted the request

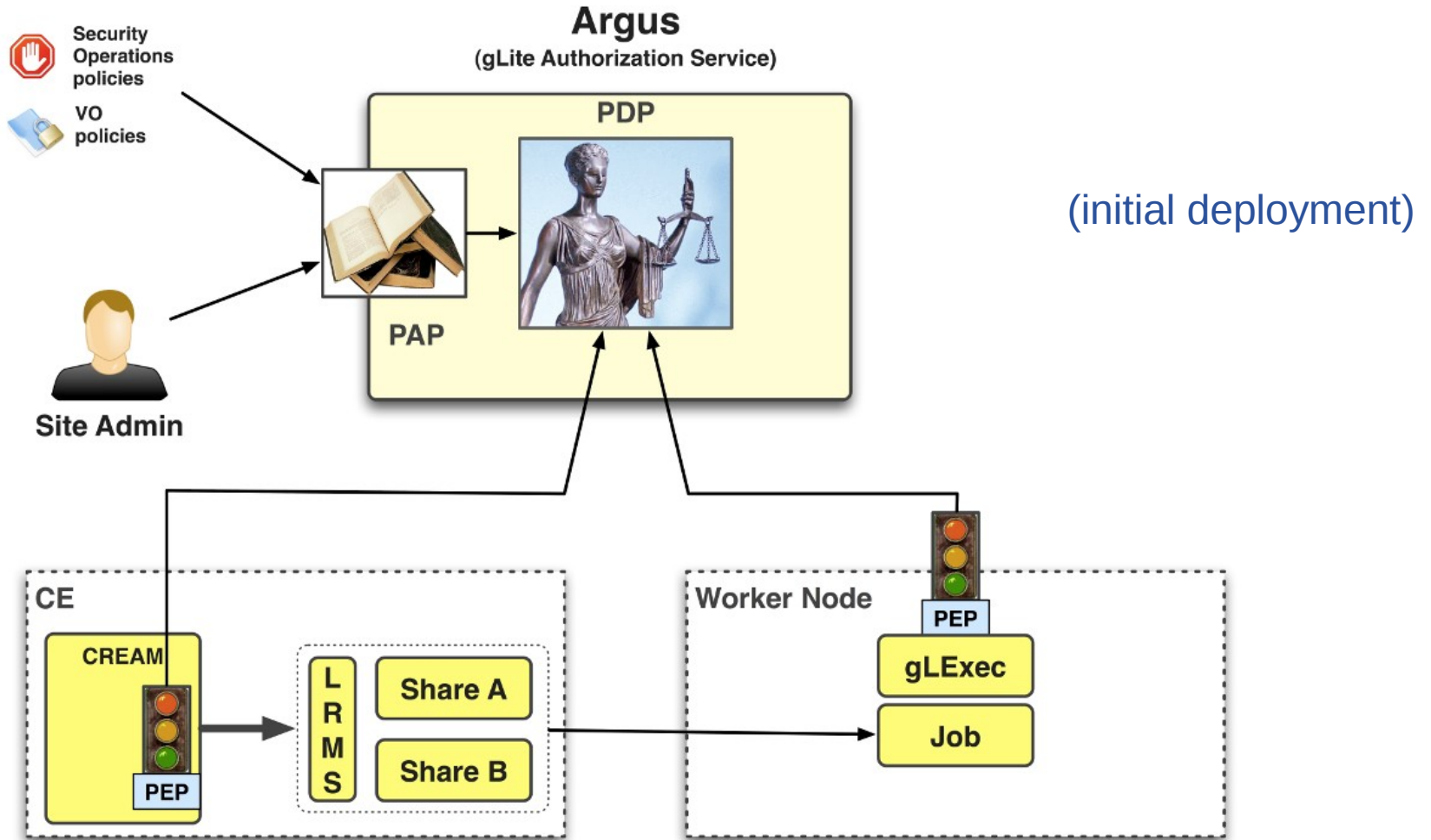
`https://devel12.cnaf.infn.it:9000/zSPX1zpKRvfslWPBnGJxdg`

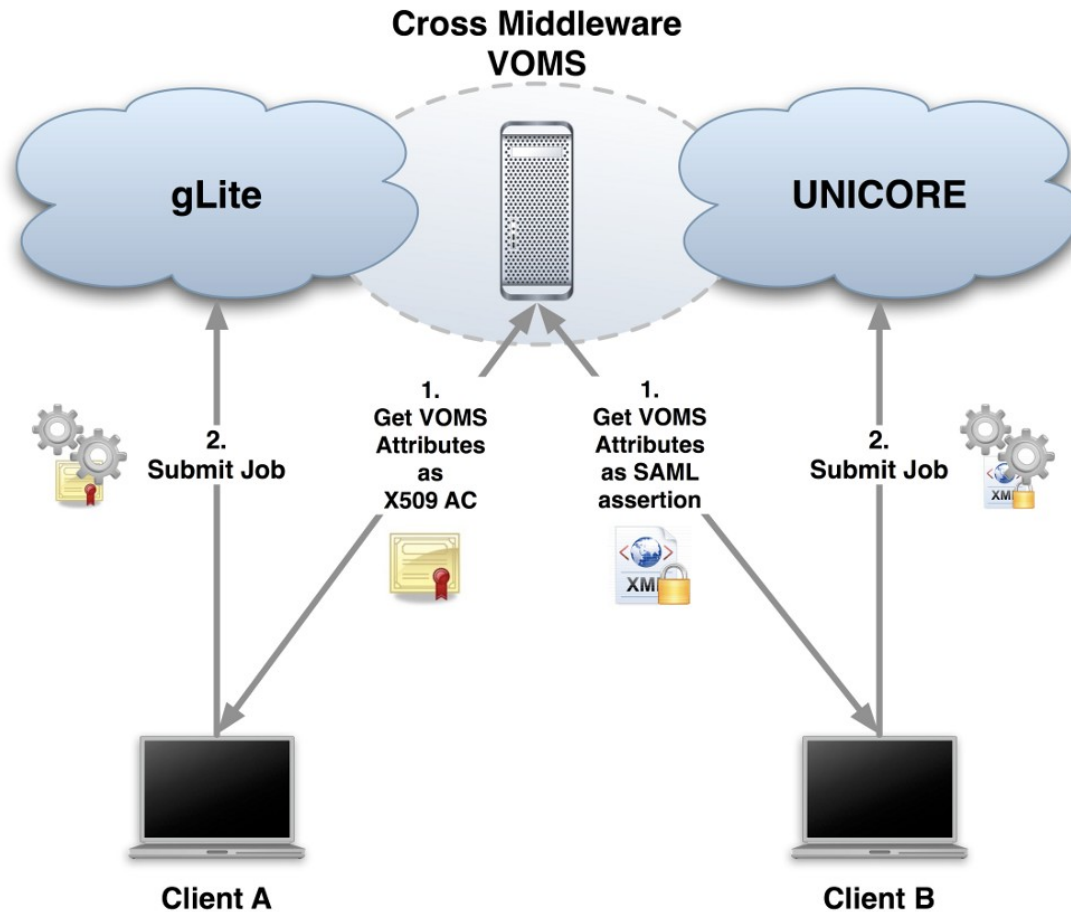
- The user can then use the id to check the progress
- **Optimizations are easier to achieve**
 - Collective operations
 - Global (or at least non-local) knowledge
- **Management of errors**

- **Other (scientific) Grids exist**
 - NorduGrid, DEISA, OSG, ...
 - Often serving the same community
- **Interoperability through a five-step plan**
 - Requirements collection
 - Identify a clear need to interoperate with another Grid
 - Involve the users
 - Analysis
 - understand the similarities and differences between the infrastructures
 - Development
 - Find and implement a solution that just works: parallel deployment, adaptors and translators, gateways, ...
 - Support
 - Maintain a production-quality service
 - Standardization

- **The long term goal is to achieve interoperability via the use of standards**
- **Interoperability is hard, standardization is harder**
 - Hope in PGI WG
- **Almost irrelevant for the functionality**
 - Users don't care
- **EMI should aim at adopting established standards or work towards their specification**
 - WS-*; BES/JSDL for job Management; SRM and GridFTP for Data Management; GLUE for Information Systems; UR/RUS for Accounting; X509, SAML/XACML for AuthN/AuthZ; ...
 - By OGF, W3C, OASIS, IETF, ...

Coherent, standards-based framework for authZ decisions

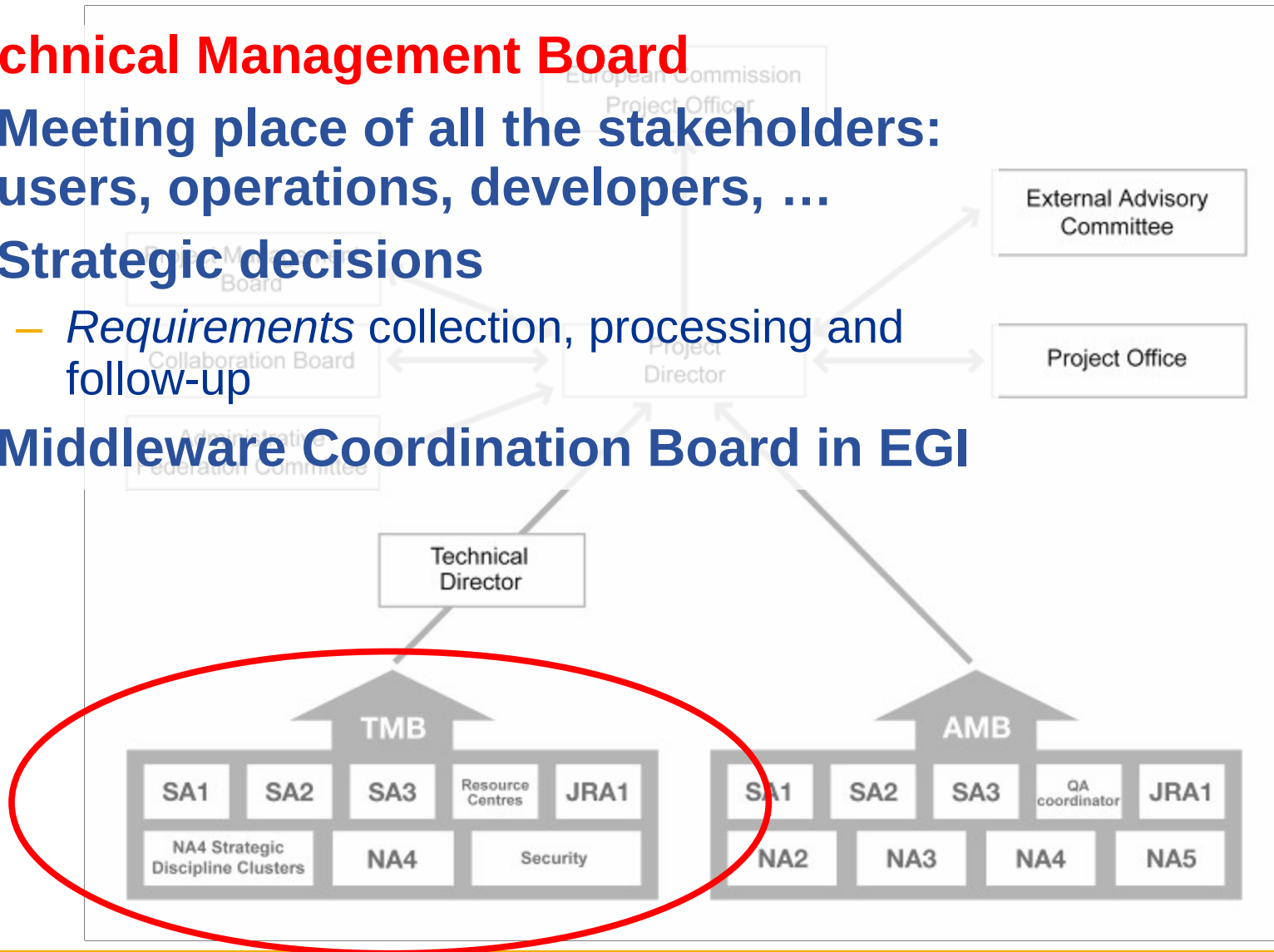




- **Next step (in EMI?): better integration with existing AAls**
 - Short-Lived Credential Service (SLCS)
 - Lower an important entry barrier to use the Grid

Technical Management Board

- Meeting place of all the stakeholders: users, operations, developers, ...
- Strategic decisions
 - Requirements collection, processing and follow-up
- Middleware Coordination Board in EGI



- **Continuous transfer of responsibilities to the actual development team**
- **A product team is responsible for delivering certified, deployable, production-quality software to operations, according to specified criteria**
 - Contract-like relationship between the infrastructure and the software provider
 - Quality Assurance is essential
- **Coordination through**
 - Central team
 - Bi-weekly meetings among team leaders and representatives of the operations
- **Large-scale testing facilities are essential**
 - *Experimental services* (~ alpha testing)
 - *Pilot services* (~ beta testing)

- **gLite provides a distribution of middleware components that allows to build (or be part of) a Grid to enable global science**
 - Focus on production quality, interoperability/standardization and security
 - It makes the EGEE infrastructure a reality
- **The future is EMI**
 - Address the needs of scientific communities in EGI together with ARC, UNICORE and others
- **gLite contribution to EMI is:**
 - Software
 - Lessons learned in organization