egee

Enabling Grids for E-sciencE

gLite future plans

Marco Pappalardo

INFN Catania

On behalf of JRA1

NA4 Generic Applications Meeting – Catania, IT, Jan 9-11 2006

www.eu-egee.org

Information Society

The EGEE-II Project

Enabling Grids for E-sciencE

- **Joint Research Activities of the EGEE-II project concern**
  - the support and incremental development of Grid middleware deployed by the project
  - quality assurance, planning and monitoring for the whole project
- **Project Goal: providing a reference open source implementation of the foundation services**
  - application independent
  - need to be deployed at all sites connected to the infrastructure.

  - On top this foundation, an open-ended set of application specific higher-level services that can be deployed on-demand at specific sites
    - to be provided directly by the project or integrated from other sources and projects

Service	Status	Service	Status
Accounting	Released	File and Replica Catalog	Released
Auditing	Prototype	Information and Monitoring	Released
Authorization	Released	Job Monitoring	Released
Agreement Service	Prototype	Job Provenance	Prototype
Bandwidth Allocation and Reservation	Prototype	Metadata Catalog	Prototype
Computing Element	Released	Network Monitoring	Prototype
Configuration and Instrumentation	Prototype	Package Manager	Prototype
Data Movement: Data Scheduler	Designed	Service Discovery	Released
Data Movement: Data Transfer	Released	Storage Element: SRM	External providers
Dynamic connectivity	Prototype	Storage Element: Data Access	Released
		Workload Management	Released

Enabling Grids for E-sciencE

- **gLite services**
  - were developed to follow a service oriented approach,
    - mostly based on web-services (compliant with WS-I recommendations)
      http://www.ws-i.org/
  - integrated and tested by JRA1
    - to form a consistent software stack that still allows individual components to be used independently.

- **These components are now being commissioned and gradually introduced on the production infrastructure.**
  - This stack needs to be supported, maintained, further evolved, in particular in view of emerging standards, and in some parts completed and hardened from their current prototype stage to a production-level release.

Enabling Grids for E-sciencE

- **Expert knowledge still required for setting up a Grid infrastructure, exploiting the infrastructure efficiently, and to integrate it with other systems and tools.**
    - JRA1 will tackle these problems by reinforcing common best-practices, standardisation, usability and deployability of Grid Foundation Middleware and selected Grid Services.
- **Role of JRA1:**
    - ensure the consolidation of the current middleware
        - by reinforcement of quality and security as well as international collaborations and standards involvement.
    - Middleware Re-engineering Activity will take responsibility for three main tasks in EGEE-II:
        - Grid Foundation Middleware,
        - Grid Services   **Execution Plan**
        - Middleware standardisation.
- **Adoption of the reengineered components in the SA3 distribution, their deployment on the SA1 infrastructure and their use (when appropriate) by NA4**

Now distinguishing:

- **application independent "*Grid Foundation Middleware*"**

  all services that need to be deployed on a production Grid infrastructure in order to provide a consistent, dependable service. It can be regarded as the "*Middleware Infrastructure*".

- **higher level "*Grid Services*".**

  comprise higher level services that certain, but not all, VOs require.

  EGEE-II focus will be on Grid Foundation Middleware,

  Key Grid Services need to be provided in order to offer the EGEE-II applications with end-to-end solutions and to validate the effectiveness of the Grid Foundation. It is expected that many of the Grid Services will be provided by other sources and projects as well.

- **A Service Oriented Architecture (SOA) approach is essential to allow different Grid foundation services to interface with services not developed inside EGEE-II.**
    - Adherence to widely accepted international standards (WSRF)

- **Components**

    Security model and infrastructure
    - define and enforce policies within VOs as well as those needed by the resource provider (i.e. between VOs).
        - *resource access control, resource access auditing and VO membership management.*
        - *policy definition and enforcement, dynamic connectivity, auditing, and interoperability with Shibboleth-based authentication and authorisation infrastructures.*

    Computing Element (CE)
    - set of services that provide homogenous, managed, and secure access to heterogeneous, remote computing resources.
        - *to directly submit jobs or to dynamically deploy VO specific schedulers.*
        - *collaboration w/ Condor and Globus.*

Grid Foundation Middleware (2/2)

Storage Element (SE)

- provides homogeneous access to storage resources, including managed data transfer.
- POSIX-like I/O and FTS.

Accounting

- collecting the relevant information locally at the resources and making it available at a global or VO level (in a secure manner) for statistical, billing, or scheduling purposes.
- Prototype.

Information and monitoring

- Information on resources must be accessible to other services in a dependable and timely manner.
- Supporting: monitoring of resources and allow user level monitoring (information and monitoring infrastructure, and service discovery).
- Allowing free information flow across different Grid flavours.

Enabling Grids for E-sciencE

- **Grid Services comprise higher level services typically VO specific**
  - exploit the Grid Foundation Middleware for achieving their purpose.
- **Grid Services not only from within EGEE-II**
- **Produce a middleware infrastructure that is general and capable of hosting Grid Services coming from other sources and projects.**
- **Services:**
  - workload management services
    - Grid scheduler like functionality
  - logging, bookkeeping, and job provenance services
    - keep track of actions performed at the Grid level
  - replica management
    - services that reliably schedule data movement and catalog updates,
  - visualization services
    - for various information collected on the Grid,
  - workflow services
    - application specific abstraction of the workflow,
  - Grid economies
    - economy based scheduling, advanced reservation, etc.

Enabling Grids for E-sciencE

- **JRA1 has built strong links with various standardisation efforts**
    - through GGF and also other committees
    - providing input from a production service point of view
    - steering the work in certain areas that need to be continued and expanded.
- **International collaborations**
    - are a necessary pre-requisite to successful standardisation efforts
    - In EGEE, JRA1 built relations with:
        - Condor, Diligent, Globus, NAREGI, NDGF, OSG, and UNICORE.
- **JRA1 has been conservative in adopting emerging, frequently changing standards**
    - as standards mature, the EGEE-II middleware needs to be made compliant

Enabling Grids for E-sciencE

Additional Tasks for all the JRA1 teams:

- **Multi-platform support**
  - Scientific Linux (a RedHat 9 variant)
    - Other versions of Linux and other operating systems will be supported, at least by the service clients.
    - Both 32 and 64 bit platforms need to be supported.
    - The impact of IPv6 needs to be assessed and if necessary appropriate porting activities initiated.
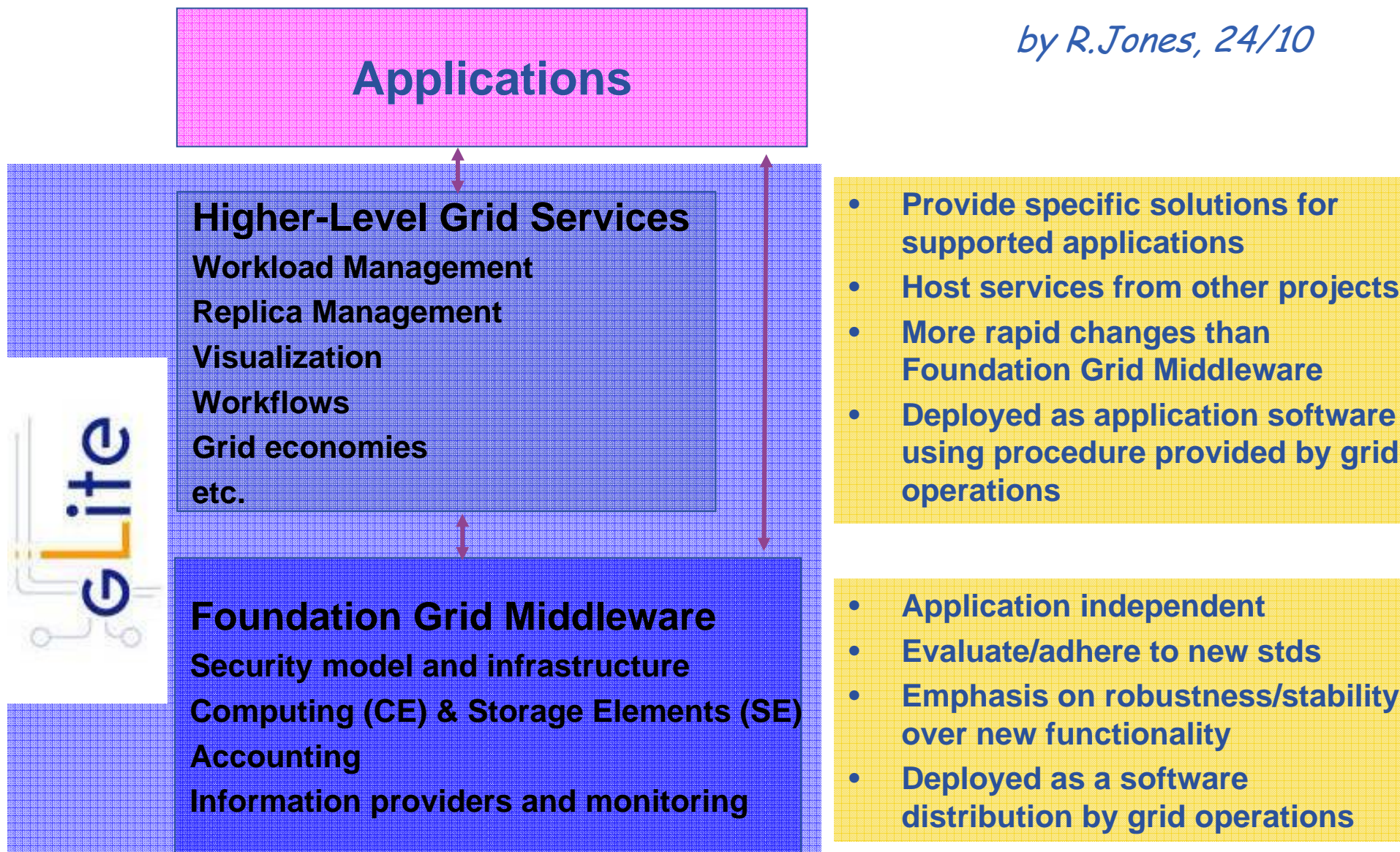
- **Reduced client dependencies**
  - Clients to the Grid foundation services, and the Grid services supported by JRA1, need to be easily installable by non-privileged users

Improved systems diagnostics
  - Troubleshooting of integrated systems for Grid computing
    - Better error distribution, easier access to relevant diagnostics and more human-readable errors should be pursued to ease system operations.

egee

Enabling Grids for E-sciencE

by R.Jones, 24/10

Applications

Higher-Level Grid Services

Workload Management

Replica Management

Visualization

Workflows

Grid economies

etc.

gLite

Foundation Grid Middleware

Security model and infrastructure

Computing (CE) & Storage Elements (SE)

Accounting

Information providers and monitoring

- **Provide specific solutions for supported applications**
- **Host services from other projects**
- **More rapid changes than Foundation Grid Middleware**
- **Deployed as application software using procedure provided by grid operations**

- **Application independent**
- **Evaluate/adhere to new stds**
- **Emphasis on robustness/stability over new functionality**
- **Deployed as a software distribution by grid operations**

egee

Enabling Grids for E-sciencE

- **Converge from LCG and gLite to a single middleware stack called gLite. The first version will be gLite 3.0.0**
- **gLite 1.5.0 and LCG 2.7.0 will be the last independent releases (expected in January)**
- **gLite 3.0.0 will contain the following components:**
  - All components already in LCG 2.7.0 plus upgrades
    - this already includes new versions of VOMS, R-GMA and FTS
  - The Workload Management System and the DGAS accounting system of gLite 1.5.0
  - The PPS will have also the other gLite 1.5.0 components
- **After the release the missing components of gLite 1.5.0 will be included in a new release, or the same functionality will be added to the existing components**
  - Will start from the data management system (Fireman, gLiteI/0, Hydra, AMGA)

- **TCG proposed timeline:**

- **Integration: During January**

  – On component level (separate build systems)

  – Merging the service configuration tools

    ▪ Taking into account outcome of the site manager's survey

- **Testing and Certification: During February**

  – Preproduction service as deployment test

  – Based on existing test suites

    ▪ Very little time for merging the test frames

- **Public Release: End of February**

  – Deploying WLM in parallel on large sites

  – Small sites can afford just one gatekeeper

- **Comments:**

- **This timeline is extremely aggressive**
  - Interference between LCG-2_7_0 and gLite-3.0.0
  - January and February are "short months"
    - We can't ignore CERN's Christmas break and CHEP
- **But:**
- **gLite-3.0.0 and LCG-2_7_0**
  - Components have been in a release already
  - Should work as they are
    - coexistence has been tested on the preproduction service

Enabling Grids for E-sciencE

- **Driven by the TCG according to the EGEE-II process**

  – other components form gLite 1.5.0

  – adaptation of components to meet user needs

  – Introduction of missing functionality

- **Input from applications Task Forces**
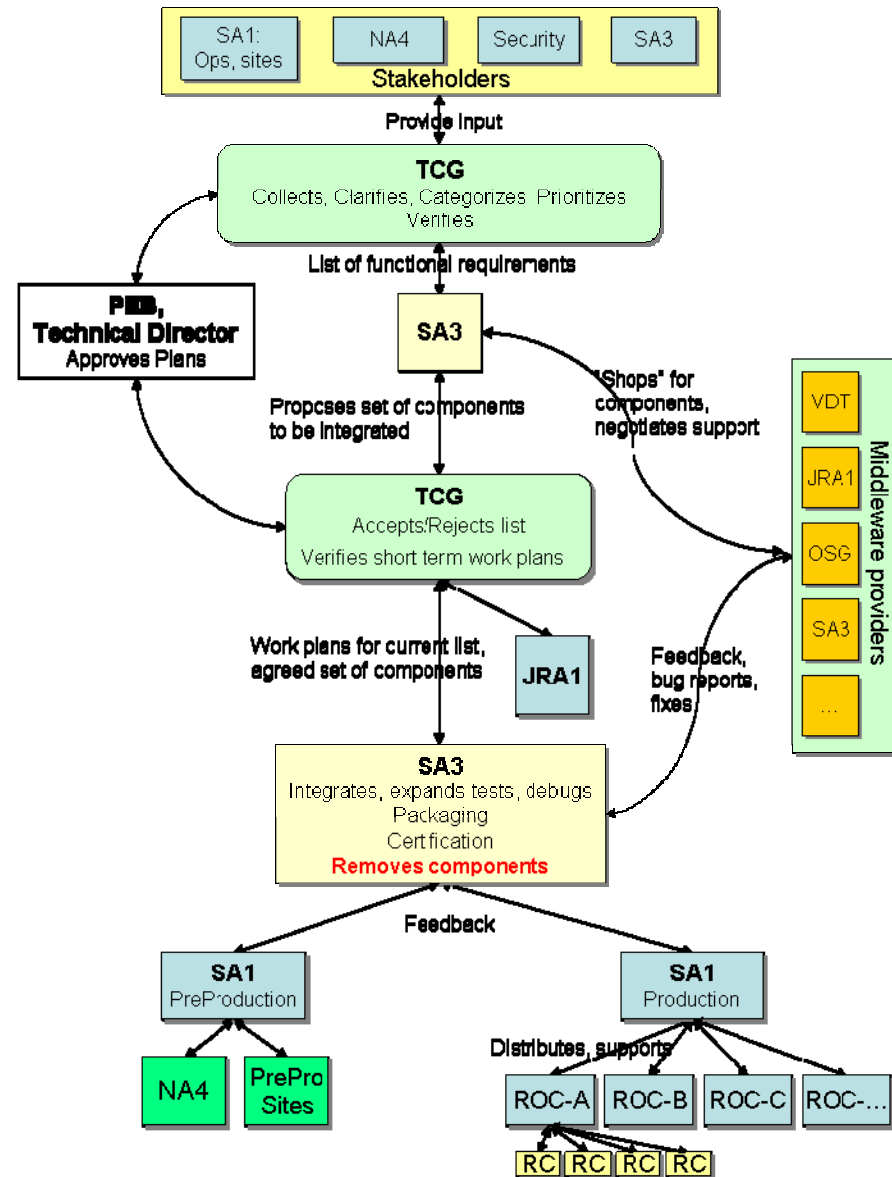
- EGEE-II Enabling Grids for E-sciencE II Technical Annex document, Dec 12 2005

- JRA1 Middleware Reengineering, EGEE 2nd EU Review, Geneva, Dec 6-7 2005

- Thanks for their key support to:
  - **Claudio Grandi**
  - **Francesco Prelz**