

Session C: Generic Middleware and Applications

**LCG Workshop
23-24 March 2004**

Erwin Laure (Cern)

Frederic Hemmer, Dirk Düllmann, Massimo Lamanna (Cern)





- **Generic Grid Middleware**

- Middleware development within the EGEE project

Frederic Hemmer

- Generic Middleware Services

Erwin Laure

- **Application Middleware**

- Distributed Analysis on the Grid - the ARDA project

Massimo Lamanna

- POOL - Status and Plans

Dirk Düllmann



Application
specific
software

Application
specific
software

Application
specific
software

*Sometimes no
clear
separation
possible*

Common Application Layer

Generic Middleware

FABRIC

FABRIC

FABRIC

FABRIC

FABRIC

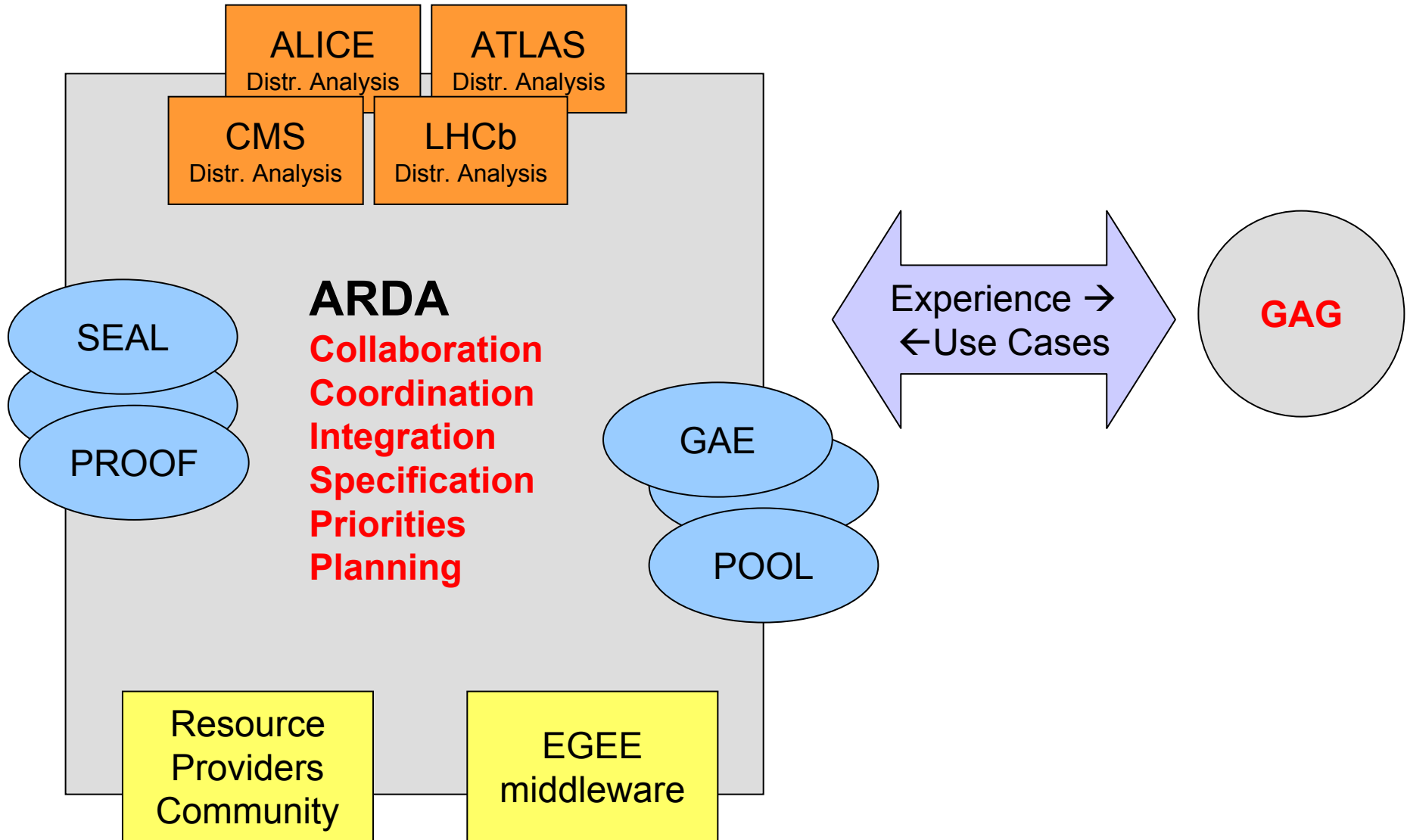


- New service decomposition
 - Strong influence of Alien system
- Role of experience, existing technology...
 - Web service framework

EGEE Middleware

- Interfacing to existing middleware to enable their use in the experiment frameworks
- Early deployment of (a series of) prototypes to ensure functionality and coherence

ARDA project



- One prototype per experiment
 - Formally, the project starts April the 1st
 - Preparation phase already started
 - Same pattern being proposed to each experiments
 - Interfacing to EGEE MW
 - Direct contribution into experiment-specific “Upper Middleware”
 - Focused dedicated effort to be added to the experiment system
 - Not a demonstration system to be added to the experiments plans
 - Mainstream activity

Milestone	Date	Description
1.x.1	May 2004	E2E x prototype definition agreed with the experiment
1.x.2	August? 2004	E2E x prototype using basic EGEE middleware
1.x.3	November? 2004	E2E x prototype improved functionality
1.x	December 2004	E2E prototype for experiment x, capable of analysis
2.x	December 2005	E2E prototype for experiment x, capable of analysis and production



- POOL will be one major client of middleware services
 - POOL will be one of the application mw components for distributed analysis
 - Currently POOL is mainly used in production activities
 - POOL will need to address analysis area this year
 - POOL needs to be able to integrate with middleware (EGEE) provided services
 - Mainly catalogs, file/database access, meta data
- POOL will align with the main ARDA concepts and implementations
 - In particular Collections, Filesets and File
- POOL provides input to ARDA and generic mw
 - Deployment and usage models
 - Assessment of existing components



- **Efficient use of metadata**
 - Is a separation of catalog mapping data from associated meta data useful?
 - Needs focused discussion
- **Higher level interface for bulk insert and bulk query is required**
 - The current use of SOAP RPC call for each individual data entry will not scale to larger productions
- **Transaction concept is required for a maintainable stable production environment**
 - User transactions may span span several services!

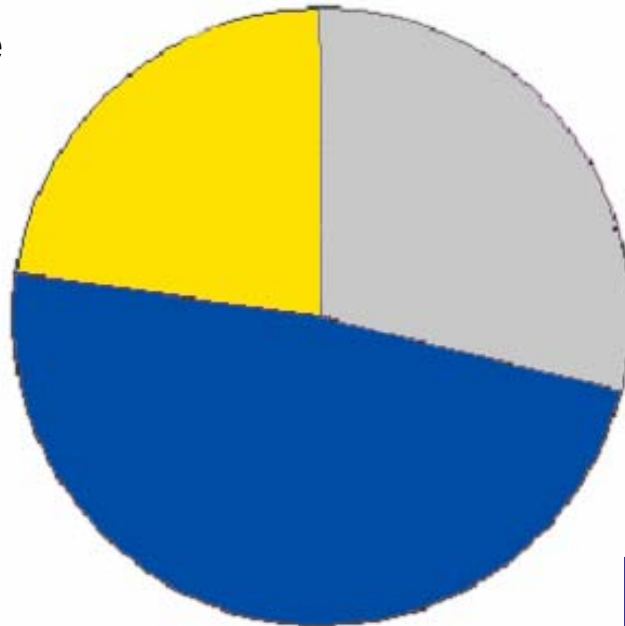
24% Joint Research

JRA1: Middleware Engineering and Integration

JRA2: Quality Assurance

JRA3: Security

JRA4: Network Services Development



48% Services

SA1: Grid Operations, Support and Management

SA2: Network Resource Provision

28% Networking

NA1: Management

NA2: Dissemination and Outreach

NA3: User Training and Education

NA4: Application Identification and Support




NA5: Policy and International Cooperation

Emphasis in EGEE is on operating a production grid and supporting the end-users

Starts 1st April 2004 for 2 years (1st phase) with EU funding of ~32M€

- Hardening and re-engineering of existing middleware functionality, leveraging the experience of partners
- Activity concentrated in few major centers and organized in “Software clusters”
- Key services:
 - Data Management (CERN)
 - Information Collection (UK)
 - Resource Brokering, Accounting (Italy-Czech Republic)
 - Quality Assurance (France)
 - Grid Security (Northern Europe)
 - Middleware Integration (CERN)
 - Middleware Testing (CERN)



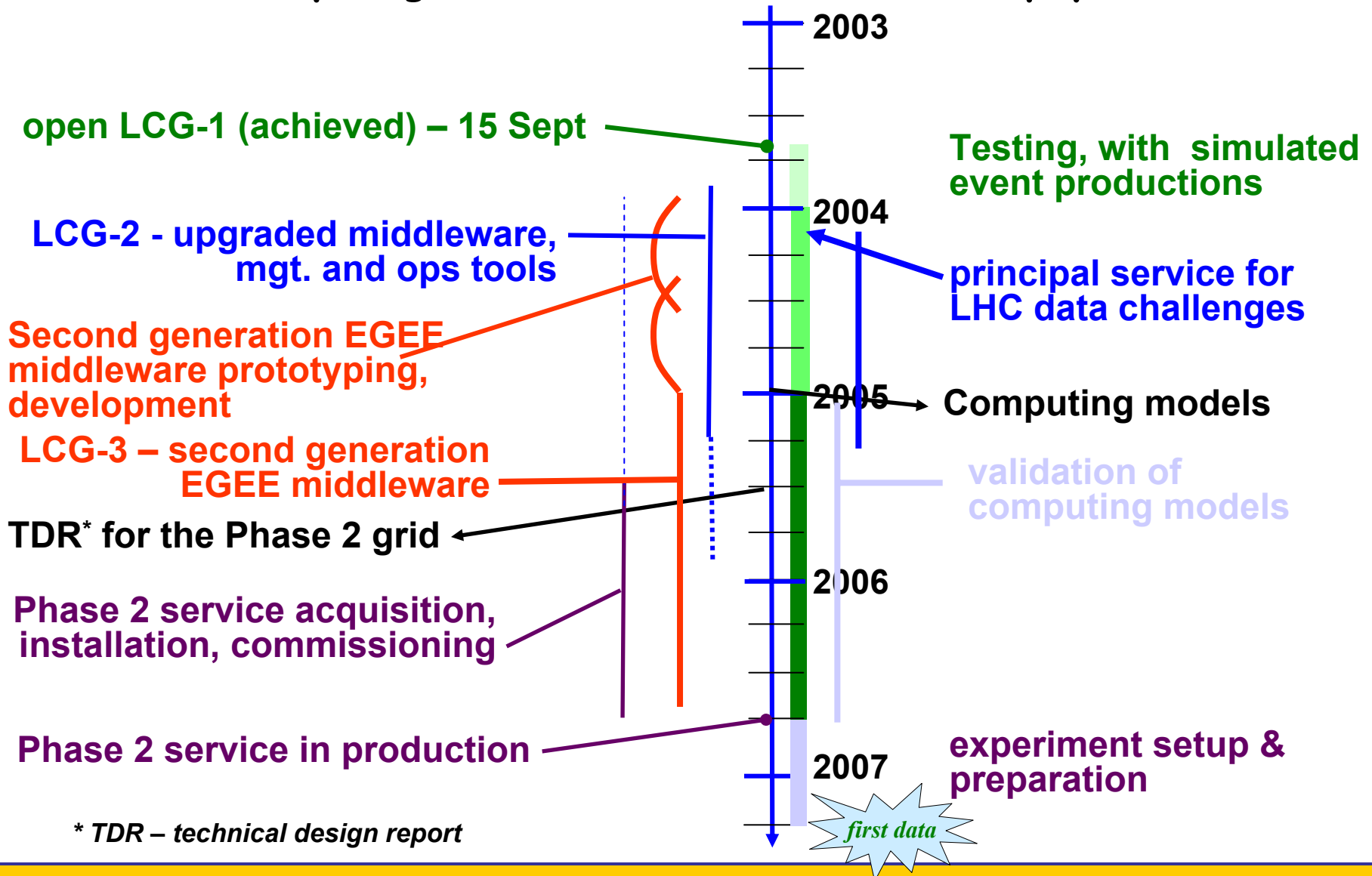
-  Middleware Integration and Testing Centre
-  Middleware Re-engineering Centre
-  Quality and Security Centres

LCG Service Time-line



computing service

physics





- LCG Project Leader in EGEE Project Management Board
- EGEE Project Director in LCG POB
- Cross technical management relations between LCG and EGEE established (Middleware and Operations Managers in common between LCG and EGEE PEBs)
- EGEE Technical director in LCG PEB
- EGEE HEP applications hosted at CERN and core resource of the LCG ARDA prototype

Mandate to generic Grid middleware



- Provide a prototype of generic Grid middleware quickly, which experiments can interface to
- Use a service oriented approach
 - OGSI is not applicable anymore – use plain web services
 - Follow WSRF
 - Migration to WSRF should be “easy” once it is settled

A note on OGSI/WSRF/WS/....

- Big hype – nothing has settled yet
- Need to take a step back
 - Focus on the service decomposition, semantics, interplay
 - rather than on the envelope
- WS seems to provide a useful abstraction
 - Widely used in industry, Grid projects, Internet computing (google, Amazon)
 - Need to follow standardization efforts to be able to adopt them once settled

Middleware approach



- Formed a design team with members from
 - AliEn
 - VDT
 - EDG

all members will be part of EGEE as of April 1st
- Monthly meetings
- Started intense technical discussion to
 - Break down the proposed architecture to real components
 - Identify critical components (and what existing software to use for the first instance of a prototype)
 - Define semantics and interfaces of these component
 - Coordinate with LCG AA (e.g. POOL)

- **Data management**
 - Storage Element
 - SRM based; allow POSIX-like access
- **Workload management**
 - Computing Element
 - Allow pull and push mode
- **More discussions needed**
 - Information and monitoring
 - Security
- **Guiding principles:**
 - **Lightweight services**
 - Easily and quickly deployable
 - **Interoperability**
 - Allow for multiple implementations
 - Being based on WS should help
 - **Co-existence with deployed infrastructure**
 - Run as an application
- **Security:**
 - Need to integrate components with quite different security models
 - Start with a minimalist approach based on VOMS and myProxy

*Mid- to
long-term
goals*

Towards a prototype



- Focus on key services discussed in the previous presentations
 - Initially an ad-hoc installation
 - Aim to have first instance ready
 - Open only to a small user group
 - Expect frequent changes (and user feedback and integration of further services)
 - Enter a rapid feedback cycle
 - Continue with the design of remaining services
 - Enrich/harden existing services based on early user-feedback
- Initial prototype
components
for April'04
To be extended/
changed***
- **Access service:**
 - AliEn shell, APIs
 - **Information & Monitoring:**
 - R-GMA (tbd)
 - **CE:**
 - AliEn CE, Globus gatekeeper, CondorG, LCAS/LCMAPS
 - **Security (tbd):**
 - VOMS, myProxy
 - **Workload mgmt:**
 - AliEn task queue;
 - EDG-WMS (tbd)
 - **SE:**
 - SRM (Castor), GridFTP, GFAL, aoid
 - **File Transfer Service:**
 - AliEn FTD
 - **File and Replica Catalog:**
 - AliEn File Catalog, RLS

Summary



- End-to-end systems are needed (from fabric to **users**)
- Quick prototypes of E2E systems to enter a frequent feedback cycle
 - Needs to interface to existing experiments' data
- Close collaboration between experiments, LCG-AA, middleware providers (EGEE, VDT, others) via the ARDA project
 - ARDA should organise a set of regular meetings
 - More detailed planning in ARDA is ongoing
- Generic middleware development done in the context of the EGEE project
 - One set of middleware
 - No competing ARDA, LCG, EGEE flavors
 - Requirements from HEP, Biomed, and other applications
- Experiences of current DC is essential