Enabling Grids for E-science in Europe

www.eu-egee.org

Induction to the EGEE Project, 19th May 2004

EGEE Middleware

Erwin Laure
Deputy Middleware Manager

EGEE is a project funded by the European Union under contract IST-2003-508833

Outline

Enabling Grids for E-science in Europe

- Overview
- Design Team & Implementation
- LCG, ARDA and new Middleware
- Integration
- Testing
- Relations with other activities
- EMT & PTF

EGEE Project Structure

Enabling Grids for E-science in Europe

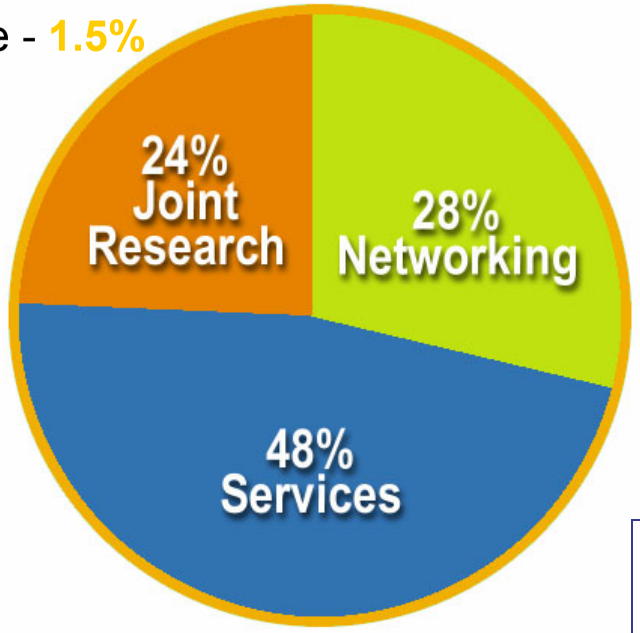
24% Joint Research

JRA1: Middleware Engineering and Integration - 17%

JRA2: Quality Assurance - 1.5%

JRA3: Security - 3%

JRA4: Network Services Development - 2.5%



28% Networking

NA1: Management

NA2: Dissemination and Outreach

NA3: User Training and Education

NA4: Application Identification and Support

NA5: Policy and International Cooperation

48% Services

SA1: Grid Operations, Support and Management

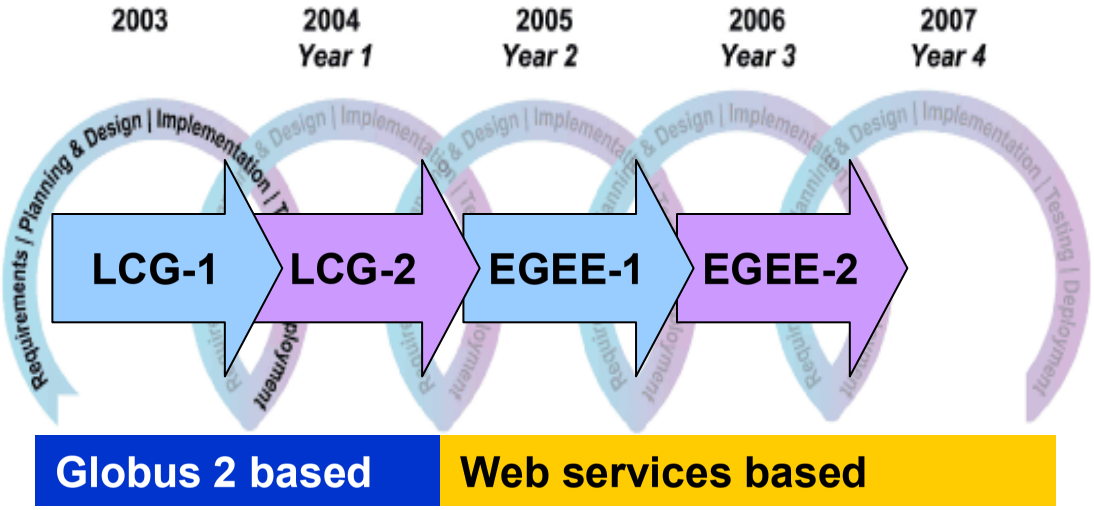
SA2: Network Resource Provision

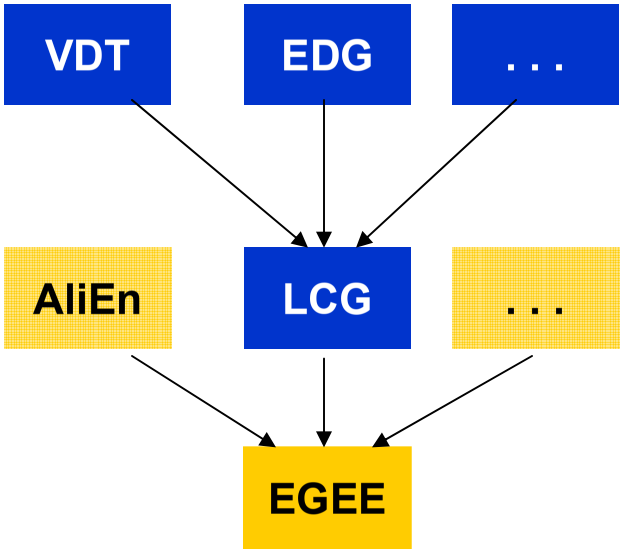
Emphasis in EGEE is on operating a production grid and supporting the end-users

EGEE Implementation

Enabling Grids for E-science in Europe

- From day 1 (1st April 2004)
  - Production grid service based on the LCG infrastructure running LCG-2 grid middleware (*SA*)
  - LCG-2 will be maintained until the new generation has proven itself (fallback solution)
- In parallel develop a "next generation" grid facility
  - Produce a new set of grid services according to evolving standards (Web Services)
  - Run a development service providing early access for evaluation purposes





Objectives of the EGEE Middleware activity

Enabling Grids for E-science in Europe

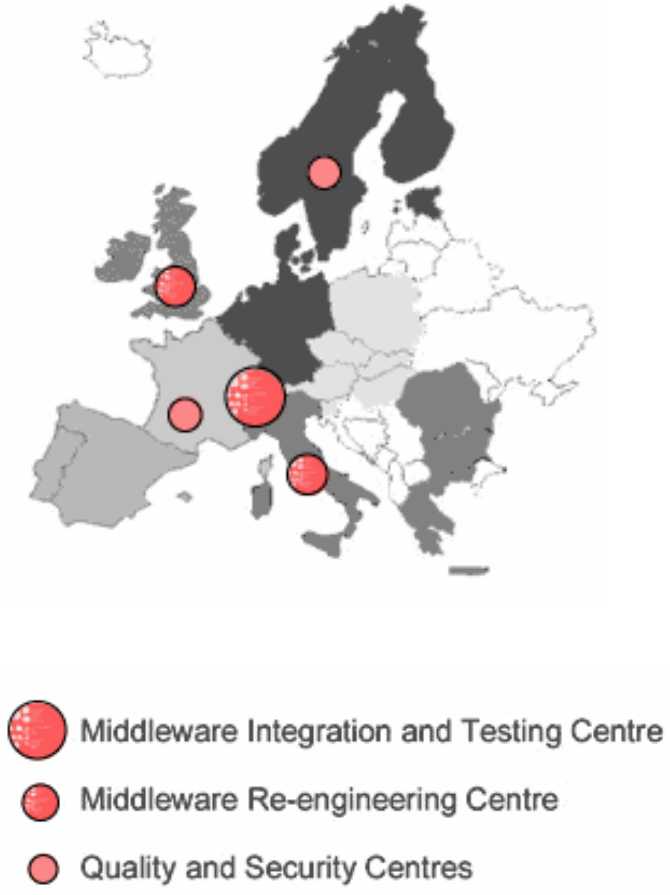
- Provide robust, supportable middleware components
  - Select, re-engineer, integrate identified Grid Services
  - Evolve towards Services Oriented Architecture
  - Adopt emerging OGSI standards*
  - Multiple platforms
- Selection of Middleware based on requirements of
  - The Applications (Bio & HEP)
    - In particular requirements are expected from LCG's ARDA & HepCALII
  - The Operations
    - E.g. deployment, updates, packaging, etc..
- Support and evolve of the middleware components
  - Evolution towards OGSI*
  - Define a re-engineering process
  - Address multiplatform, multiple implementations and interoperability issues
  - Define defect handling processes and responsibilities

*: Now questioned given the WSRF announcement on January 20, 2004. The strategy is to use plain Web Services and review the situation towards the end of the year (GT4).

EGEE Middleware Software Clusters

Enabling Grids for E-science in Europe

- 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)

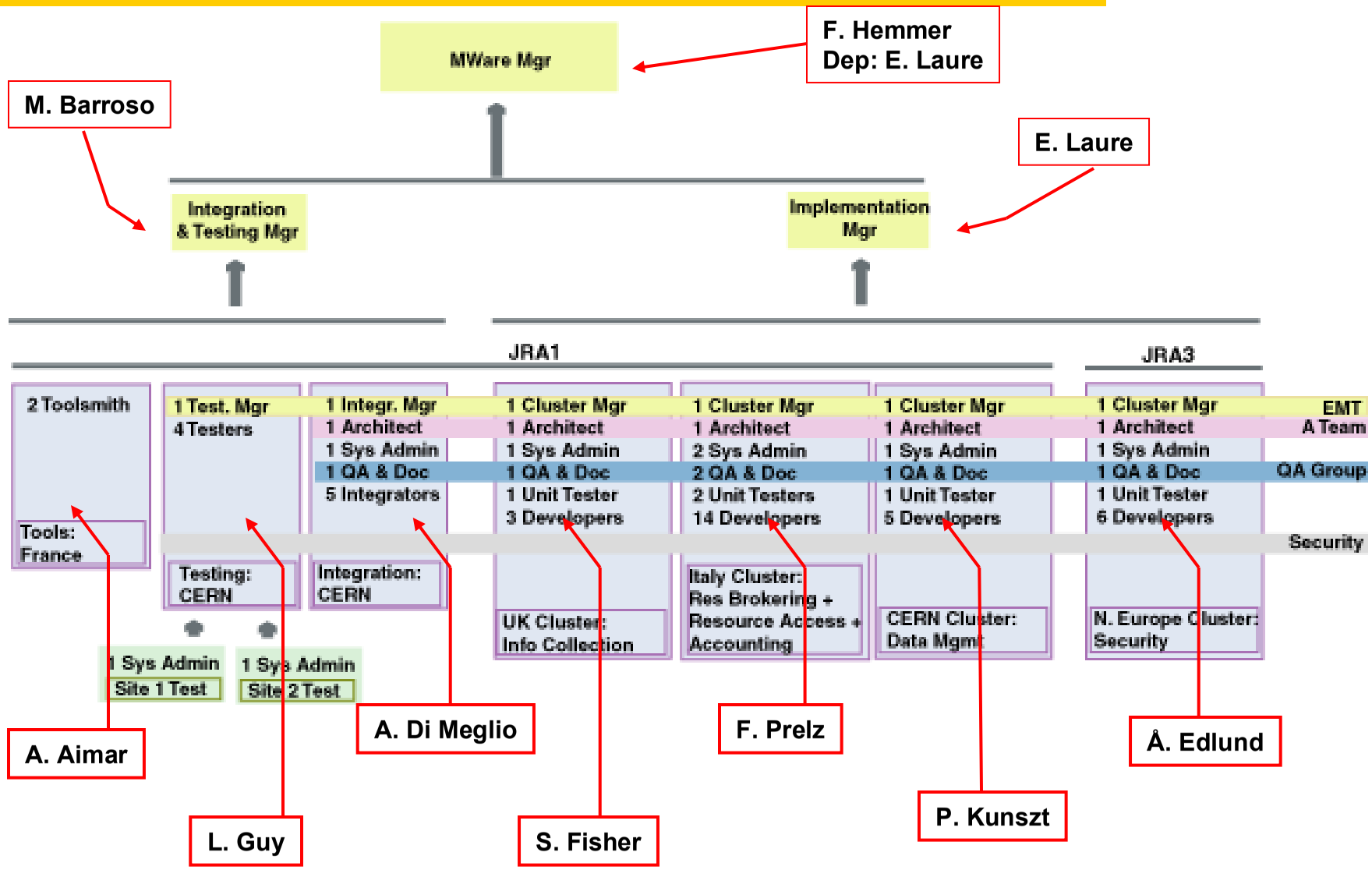


EGEE Middleware Partners

Enabling Grids for E-science in Europe

Location	Activity JRA1	Total Effort (FTE)	Total Effort (PM)	1st Year Effort (PM)
CERN	CERN	32	768	384
Italy	INFN	16	384	192
Italy	Datamat S.p.A.	6	144	72
Czech Republic	CESNET	4	96	48
United Kingdom	CCLRC	8	192	120
France	CNRS	2	48	24
USA	UChicago	N/A	N/A	N/A
USA	USC	N/A	N/A	N/A
USA	UW-Madison	N/A	N/A	N/A
	Total	68	1632	840

JRA1 Organization



Milestones and Deliverables for 2004

Enabling Grids for E-science in Europe

Month	Deliverables & Milestones	Item	Lead Partner	Status
M03	MJRA1.1	Tools for middleware engineering and integration deployed	CERN	Most tools identified (SCM) Issue: CNRS person tbi
M03	DJRA1.1	(Document) Architecture and Planning (Release 1)	CERN	Design document as starting point; writing organized for May
M03	MJRA1.2	Software cluster development and testing infrastructure available	CERN	**Done for CERN/UK/IT/CZ** Nordic to be finalized
M05	MJRA1.3	Integration and testing infrastructure in place including test plans (Release 1)	CERN	Sites identified; teams established;
M05	DJRA1.2	(Document) Design of grid services (Release 1)	CERN	Design document as starting point
M09	MJRA1.4	Software for the Release Candidate 1	CERN	Work on prototype started; will evolve into 1st release candidate

M08 – Amsterdam conference: Tech preview of release candidate 1 available

Characteristics of the new middleware

Enabling Grids for E-science in Europe

- Develop a lightweight stack of generic middleware useful to LHC experiments and BioMedicals based upon existing components
  - Biomedical applications have important security requirements (e.g. confidentiality) that need to be addressed.
- Focus is on re-engineering and hardening
- Early prototype and fast feedback turnaround envisaged
- Use a service oriented approach

A note on OGSI/WSRF/WS/….

- Still discussing – nothing has settled yet
- Need to take a step back
  - Focus on the service decomposition, semantics, interplay rather than 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

Design Team

Enabling Grids for E-science in Europe

- Formed in December 2003
- Current members:
  - UK: Steve Fisher
  - IT/CZ: Francesco Prelz
  - Nordic: David Groep
  - VDT: Miron Livny
  - CERN: Predrag Buncic, Peter Kunszt, Frédéric Hemmer, Erwin Laure
- Started service design based on component breakdown defined by the LCG ARDA RTAG
- Leverage experiences and existing components from AliEn, VDT, and EDG.
- A *working* document
  - Overall design & API's
  - **https://edms.cern.ch/document/458972**
- Basis for architecture (DJRA1.1) and design (DJRA1.2) document

Design Team Approach

Enabling Grids for E-science in Europe

- 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
- Focus on key services discussed; exploit existing components
- Initially an ad-hoc prototype installation at CERN and Wisconsin
- Aim to have first instance ready by end of April
  - Open only to a small user community
  - Expect frequent changes (also API changes) based on 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
- Prototype will be used as a vehicle to evolve the software

Initial Services

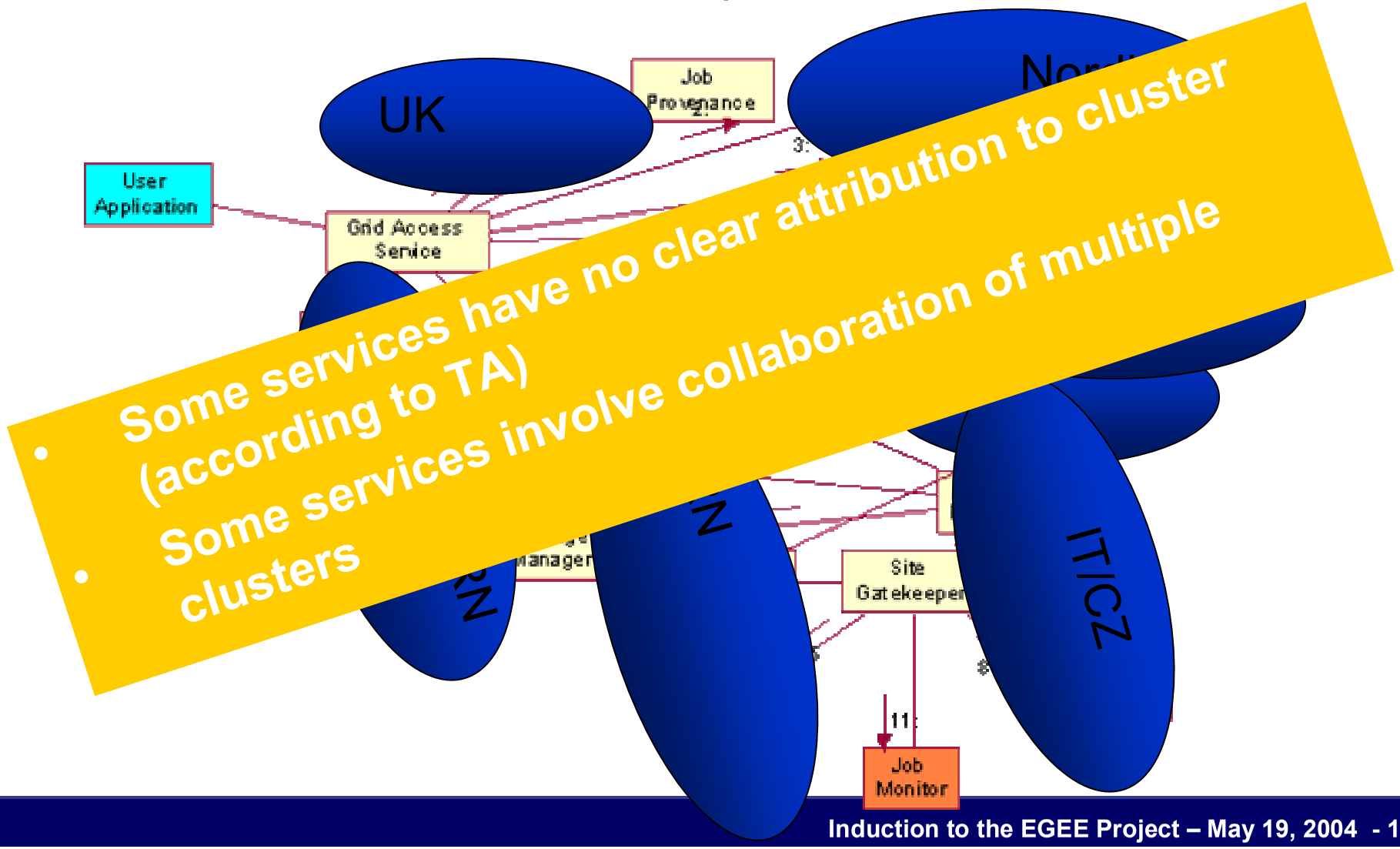
Enabling Grids for E-science in Europe

- 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 (medium/long term)
      - 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

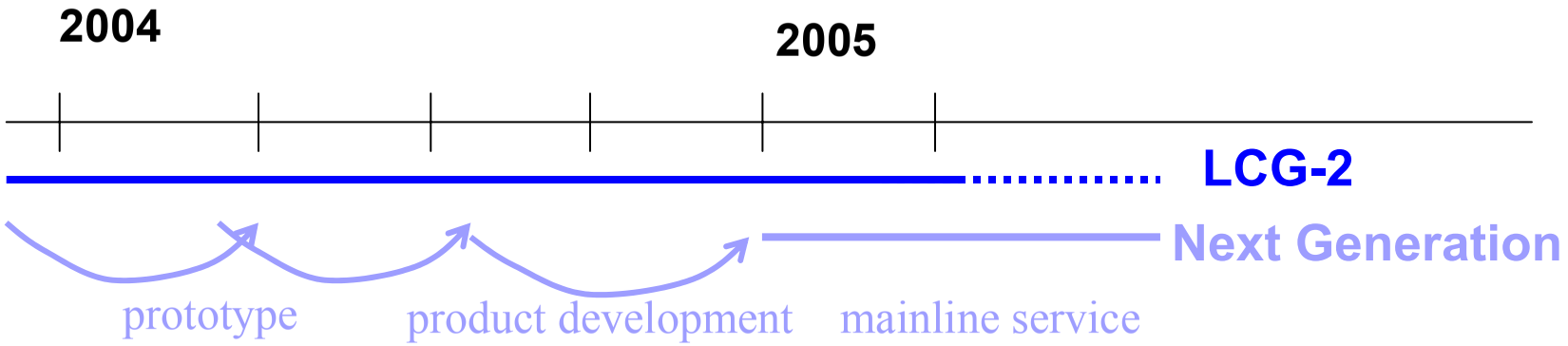
High Level Service Decomposition

- Taken from the ARDA blueprint



LCG and Next Generation Middleware

Enabling Grids for E-science in Europe



- LCG-2 will be the main service for the 2004 data challenges
- This will provide essential experience on operating and managing a global grid service – and will be supported and **developed**
- Target is to establish a base (fallback) solution for early LHC years

- LCG-2 will be maintained until the new generation has proven itself

Middleware & ARDA

Enabling Grids for E-science in Europe

- ARDA RTAG has influenced considerably the EGEE Middleware activity
  - Reference included in the Technical Annex
  - Group of Middleware providers met as of December 2003
    - Monthly meetings (design & implementation)
  - Goal to define and provide Middleware components as described in the ARDA RTAG
  - Participants from AliEn, EDG, VDT
- ARDA Project has been established
  - It is a distinct project, focused on the usage of the Middleware within the experiments
  - Providing resources to HEP to help delivering end to end analysis prototypes
  - Providing an organization to discuss and agree on Middleware components

ARDA Working Group Recommendations

Enabling Grids for E-science in Europe

- 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

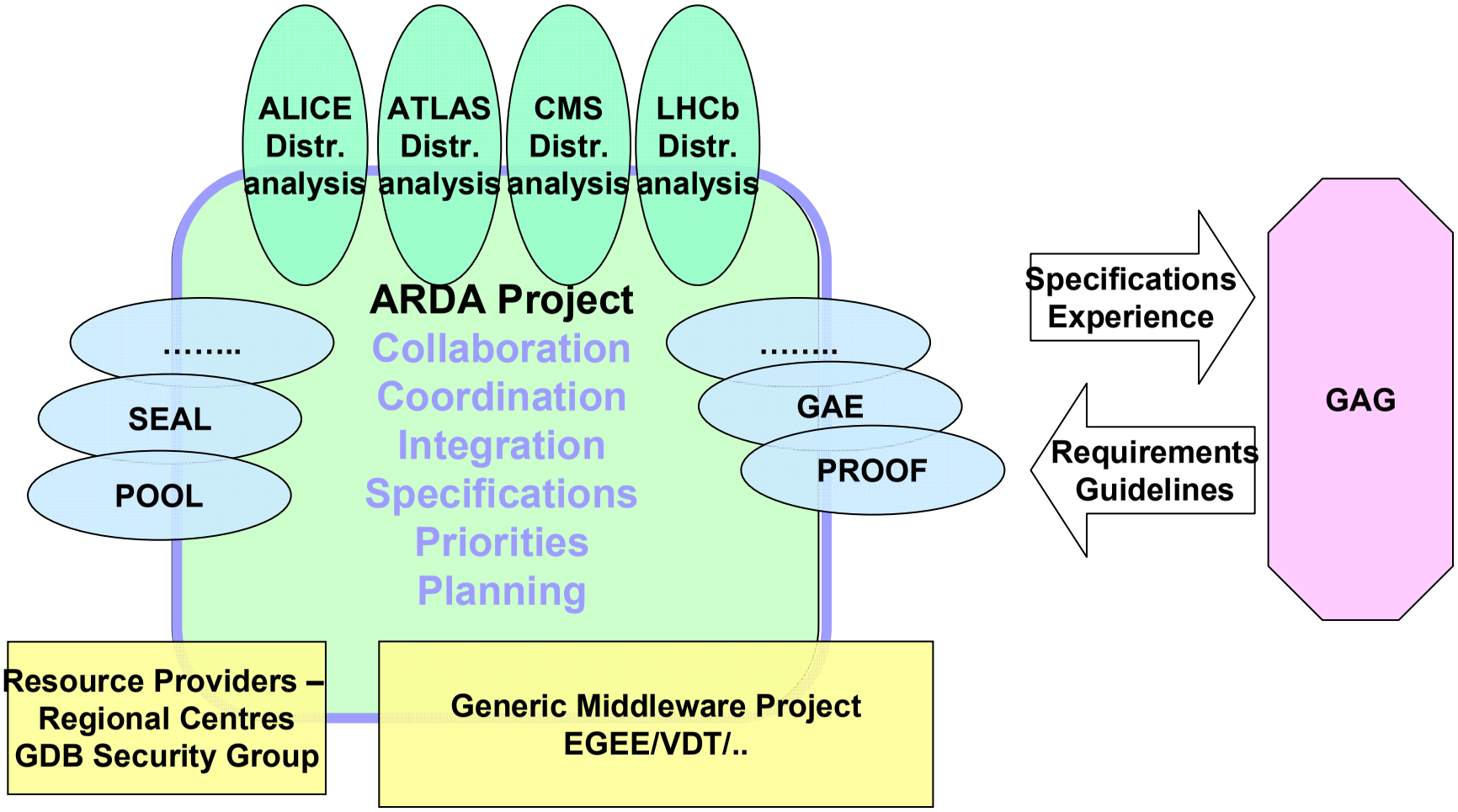
ARDA Project

ARDA End-to-end prototypes

Enabling Grids for E-science in Europe

- Provide a fast feedback to the EGEE MW development team
  - Avoid uncoordinated evolution of the middleware
  - Coherence between users expectations and final product
- Guarantee the experiments are ready to benefit from the new Middleware as soon it becomes available
  - Expose the experiments (and the community in charge of the deployment) to the current evolution of the whole system, to be prepared to use it in the best and quickest way
- Move forward towards new-generation real systems
  - Prototypes should be exercised with realistic workload and conditions (experiments absolutely required for that!)
    - No academic exercises or synthetic demonstrations
  - A lot of work (and useful software) is involved in current experiments data challenges: this will be used as a starting point
    - Adapt/complete/refactorise the existing: we do not need another system!

The ARDA Project



ARDA Project plan

Enabling Grids for E-science in Europe

- One prototype per experiment
  - 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
- Important Note: we do not have an "ARDA" Mechanism for BioMedical & Other applications
  - We do not need other projects, but the need same spirit

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

EGEE, LCG & ARDA
High-Level Strategy for Middleware

Enabling Grids for E-science in Europe

- LCG-2 middleware package strongly supported and evolved
  - Demonstrating a base solution for LHC start-up
  - Supported until overtaken by EGEE Middleware
- EGEE Middleware –
  - Re-engineered *generic* middleware package
  - Incorporating experience from AliEn, EDG, …., VDT
  - Architected for scale and performance requirements of LCG
  - "batch" and "analysis"
- Fast prototyping approach – with clear end-to-end goals
  - Short update cycles to give LHC experiments the chance to influence and give feedback
- Important Note: we do not have an "ARDA" Mechanism for BioMedical & Other applications

We do not need other projects, but the need same spirit

Integration

Enabling Grids for E-science in Europe

- A master Software Configuration Plan is being finalized now
- Compliant with internationally agreed standards (ISO 10007-2003 E, IEEE SCM Guidelines series)
- Most EGEE stakeholders have already been involved in the process to make sure everybody is aware of, contributes to and uses the plan
- An EGEE JRA1 Developer's Guide will follow shortly in collaboration with JRA2 (Quality Assurance) based on the SCM Plan
- SCM Contents
  - Configuration and Version Control
  - Build Systems
  - Release Process
  - Other Configuration and Change Control Procedures

- https://edms.cern.ch/document/446241

- Solid steps towards MJRA1.3 (PM5)

Testing

Enabling Grids for E-science in Europe

- The 3 initial testing sites are CERN, NIKHEF and RAL
  - More sites can join the testing activity at a later stage !
  - Must fulfil site requirements
- Testing activities will be driven by the test plan document
- Test plan being developed based on user requirements documents:
  - Application requirements from NA4: HEPCAL I&II, AWG documents, Bio-informatics requirements documents from EDG
  - Deployment requirements being discussed with SA1
  - ARDA working document for core Grid services
  - Security: work with JRA3 to design and plan security testing
- The test plan is a living document: it will evolve to remain consistent with the evolution of the software
- Coordination with NA4 testing and external groups (e.g. Globus) established
- Solid steps towards MJRA1.3 (PM5)

JRA1/SA1 - Process description

Enabling Grids for E-science in Europe

- No official delivery of requirements from SA1 to JRA1 stated in the TA
- The definition, discussion and agreement of the requirements has already started, done through dedicated meetings

- This is an ongoing process:
  - Not all the requirements defined yet
  - Set of requirements agreed, need basic agreement to start working! But can be reviewed at any time there is a valid reason for it

- **Requirement table stored at:**
  https://edms.cern.ch/document/456865
- **Meeting minutes stored at:**
  https://edms.cern.ch/document/451069

JRA1/SA1 - Requirements

Enabling Grids for E-science in Europe

1. Middleware delivery to SA1
2. Release management
3. Deployment scenarios
4. Middleware configuration
   - JRA1 will provide a standard set of configuration files and documentation with examples that SA1 can use to design tools. Format to be agreed between SA1-JRA1
   - It is the responsibility of SA1 to provide configuration tools to the sites
5. Enforcement of the procedures
6. Platforms to support
   - Primary platform: Red Hat Enterprise 3.0, gcc 3.2.3 and icc8 compilers (both 32 and 64-bits) .
   - Secondary platform: Windows (XP/2003), vc++ 7.1 compiler (both 32 and 64-bits)
7. Versions for compilers, libraries, third party software
8. Programming languages
9. Packaging and software distribution
10. Others
    - Sites must be allowed to organize the network as they wish, internal or external connectivity, NAT, firewall, etc, all must be possible, no special constraints. WNs must not require Outgoing IP connectivity; Not inbound connectivity either.

NA4/JRA3/JRA4

Enabling Grids for E-science in Europe

- NA4
  - HEP: ARDA project started; ensures close relations between HEP and middleware
  - Bio: activities with similar spirit needed – focused meeting planned for July
- JRA3
  - Security Group formed, JRA1 members identified
  - First meeting took place on May 5-6, 2004
    - GAP analysis has been performed and draft available
- JRA4
  - SCM plan presented and discussed
  - More discussions on which components of JRA4 will be required in the overall architecture/design need to take place

Engineering Management Team

Enabling Grids for E-science in Europe

- Managing Middleware Engineering Process
  - Requirements Collection
  - Design & Development
  - Integration
  - Testing
  - Maintenance
- Composed of
  - Middleware Manager
  - Middleware Implementation Manager
  - Middleware Integration and Testing Manager
  - Software Clusters managers
  - Security Person
  - QA person
- Weekly Meeting
  - Minutes available at https://edms.cern.ch/document/458097

Project Technical Forum

Enabling Grids for E-science in Europe

- Merge requirements and use cases from the application, security, and operations groups.
- Manage the evolution of the requirements and use cases.
- Prioritize the requirements and verify if (and how) the defined use cases are satisfied by the EGEE architectural design.
- Review the grid service specifications and control changes to those specifications and external interfaces.
- Seek and promote common solutions where commonalities exist in the middleware or application domains.
- Promote convergence with other grid projects by incorporating requirements from external projects using (or hoping to use) EGEE software and by discussing the EGEE design with other projects.
- PTF chairman will represent EGEE in standardisation bodies such as GFF/OASIS

More information

Enabling Grids for E-science in Europe

http://cern.ch/egee-jra1