



Enabling Grids for E-science

Application activity within EGEE II

*Vincent Breton, on behalf of NA4
All Activity Meeting, July 5-6 2005*

www.eu-egee.org



- **drive the evolution of the grid technology through specific, challenging applications**
 - **pilot applications** (LHC and Biomed) committed to using the large distributed infrastructure EGEE to achieve their scientific goals
- **demonstrate that EGEE-II provides a viable computing infrastructure for research to several scientific communities.**
 - EGEE-II will host a number of scientifically diverse applications with the help of teams of engineers funded by EGEE and the scientific communities concerned

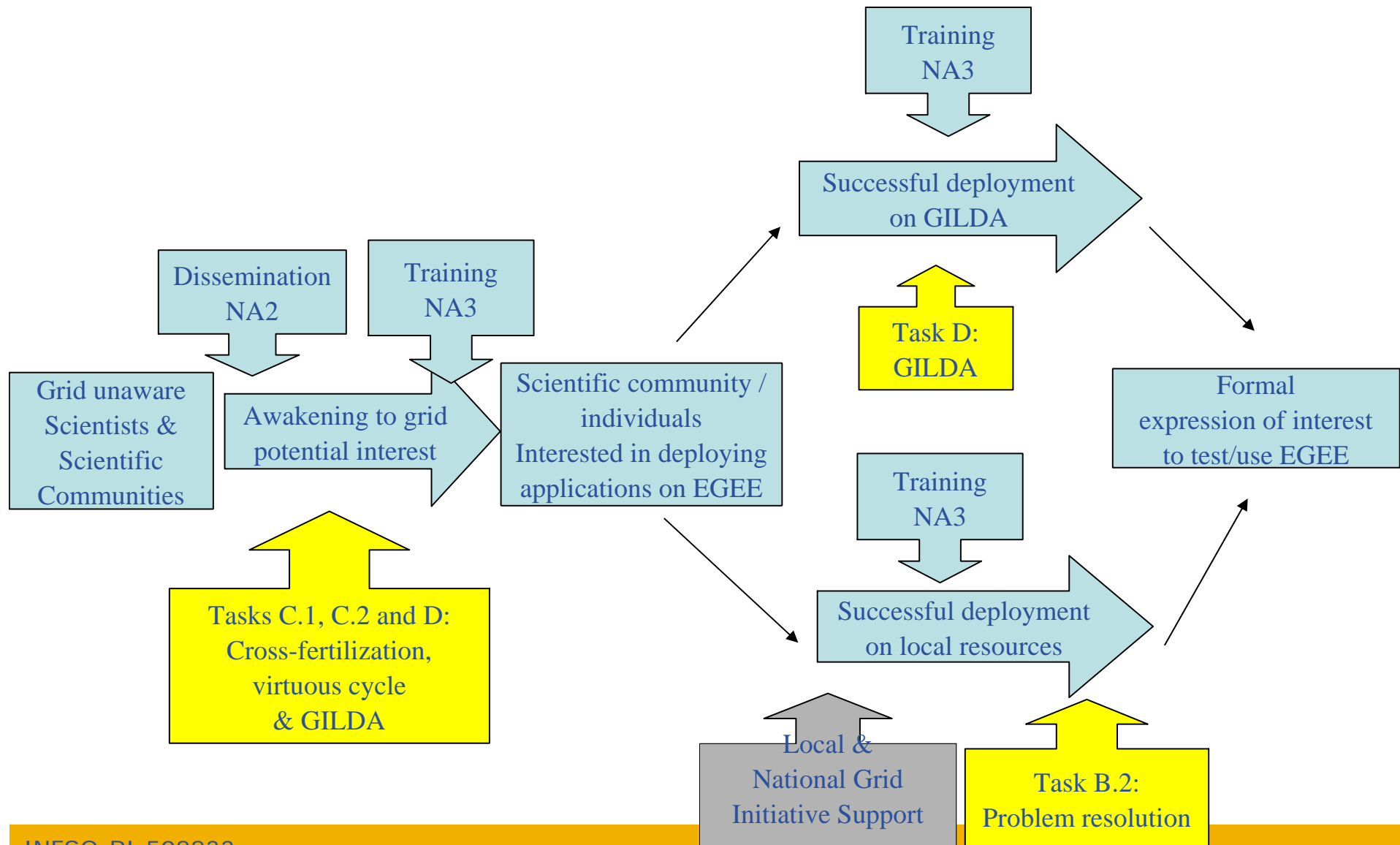
- **Selected applications having an explicit agreement with EGEE and being permitted to use a large number of different resource centres.**
 - EGEE personnel and individuals within the application groups will work together closely to ensure the applications can successfully exploit the EGEE infrastructure.
- **External applications coming from other EU projects having a formal relationship with EGEE**
 - They bring dedicated manpower to ensure an efficient liaison with EGEE
- **Local applications accessing individual (or a small number of) resource centres through EGEE middleware interfaces**
 - The local resource centres are expected to provide the bulk of support for these applications.

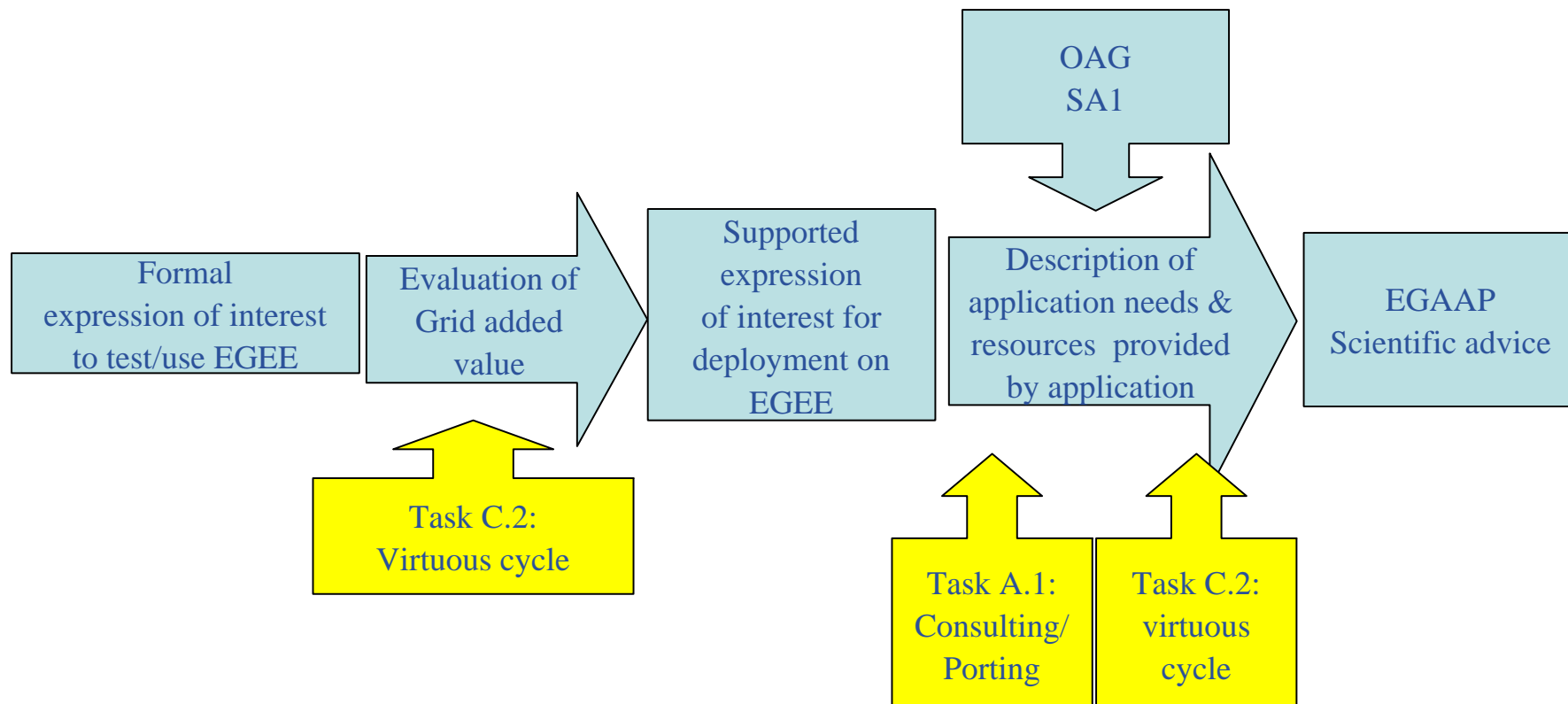
- **Consulting, porting of applications**
 - support for applications to port their software to the grid environment
 - recurrent support for existing applications as new services appear in the grid middleware and the applications evolve
- **Evolution of pilot applications:**
 - to exploit the required grid baseline services deployed in EGEE
 - this activity should be designed together with the applications
- **Validation of the EGEE infrastructure**
 - to guarantee the production and availability of the data for analysis by the large and distributed scientific communities
 - follow carefully functionality and performance issues of the applications on the production service

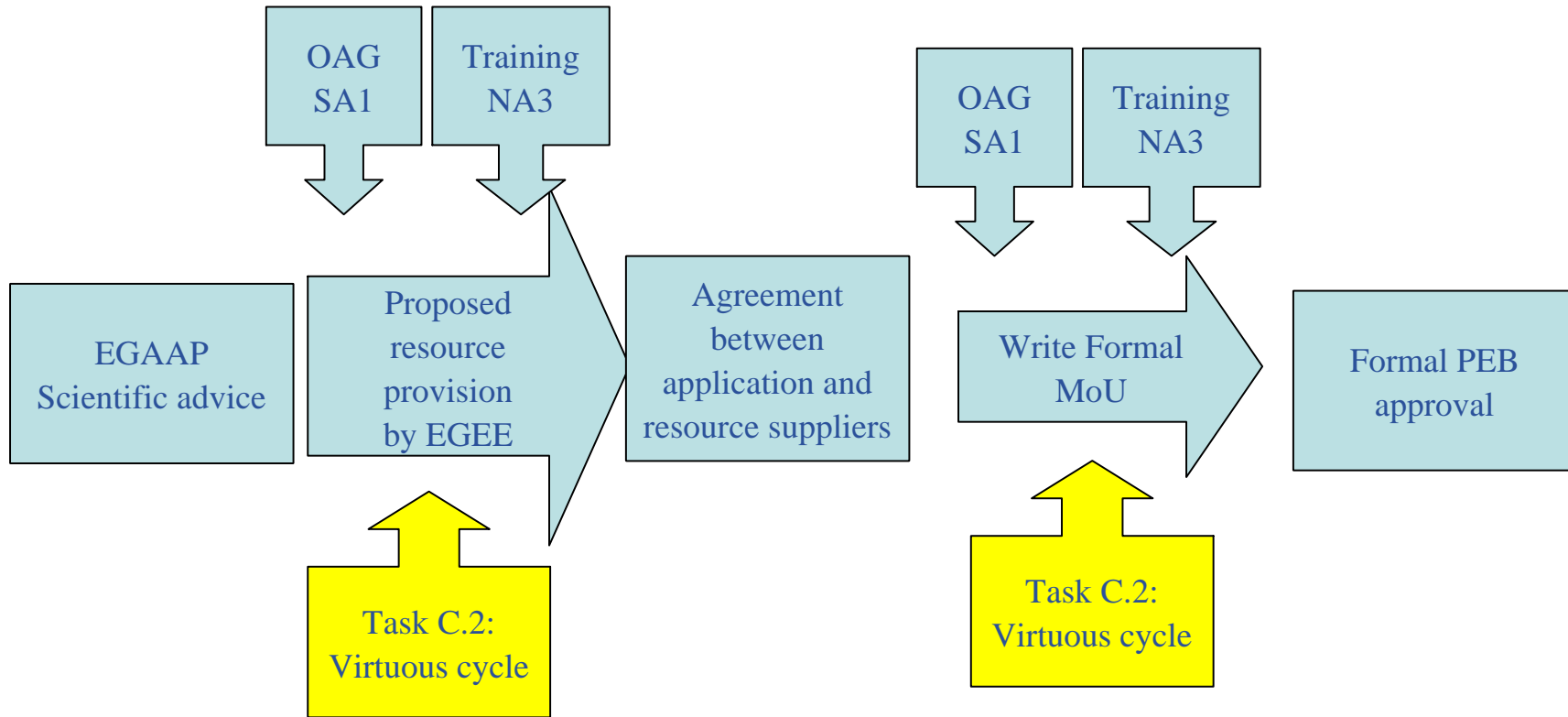
- **Application deployment**
 - act as a liaison between the applications and the operational infrastructure
 - ensure that the minimum services for all supported applications are provided.
- **Troubleshooting/testing**
 - help applications resolve problems with deployment
 - filter problems before they are submitted to operations support (SA1)
 - proactively test the infrastructure (from an application perspective) to identify and resolve problems before they affect production use

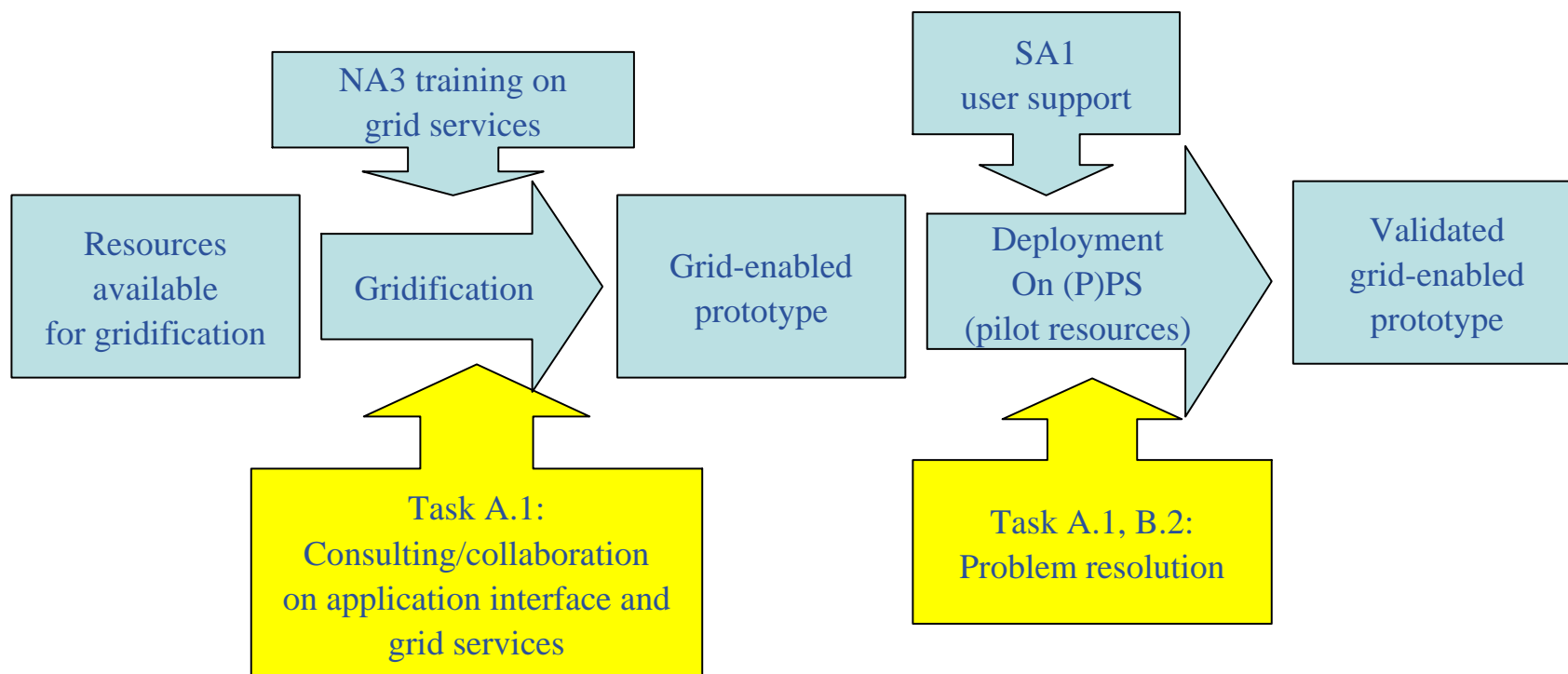
- **Cross-fertilization**
 - ensure regular contact between virtual organizations
 - understand how different applications use the grid
 - provide a means of collaboration between those applications
- **Virtuous cycle**
 - ensure that new applications are identified, properly integrated into the EGEE infrastructure and receive the necessary support,
- **GILDA, a grid laboratory**
 - To attract and progress with new applications, NA4 needs a special environment (separate from the production infrastructure) on which one can experiment with the grid software

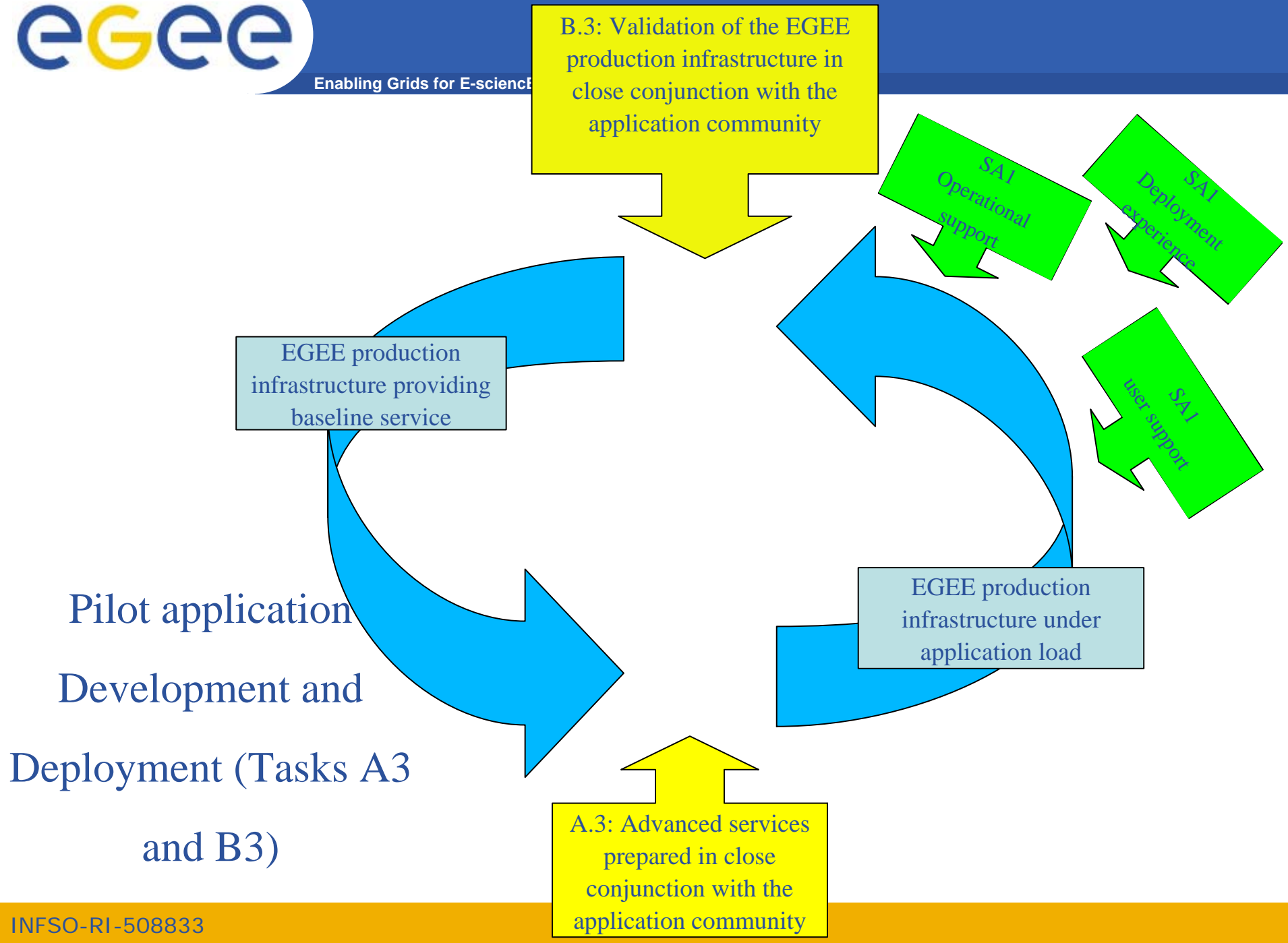
- **Tasks in relation to JRA1**
 - Task A.1: Consulting and porting of applications
 - Task A.2: Testing
 - Task A.3: Evolution of the pilot applications
- **Tasks in relation to SA1**
 - Task B.1: Application deployment
 - Task B.2: Validation of EGEE Infrastructure
- **Tasks in relation to application integration**
 - Task C.1: Cross-fertilization
 - Task C.2: Virtuous cycle
- **Tasks in relation to NA3: GILDA (Grid laboratory; initial porting)**
 - Task D.1: Set-up and central management of the t-Infrastructure
 - Task D.2: Realization of training and GILDA specific dissemination events
 - Task D.3: Initial porting of applications and user support
- **Task E: Specific Application Tasks**
 - Task E.1: Writing of MoUs..
 - Task E.2: User support
 - Task E.3: Demonstrations/evaluations
- **Task F: Management**
 - Task F.1: Task in relation to JRA2.
 - Task F.2: Management/oversight of people.
 - Task F.3: Coordination

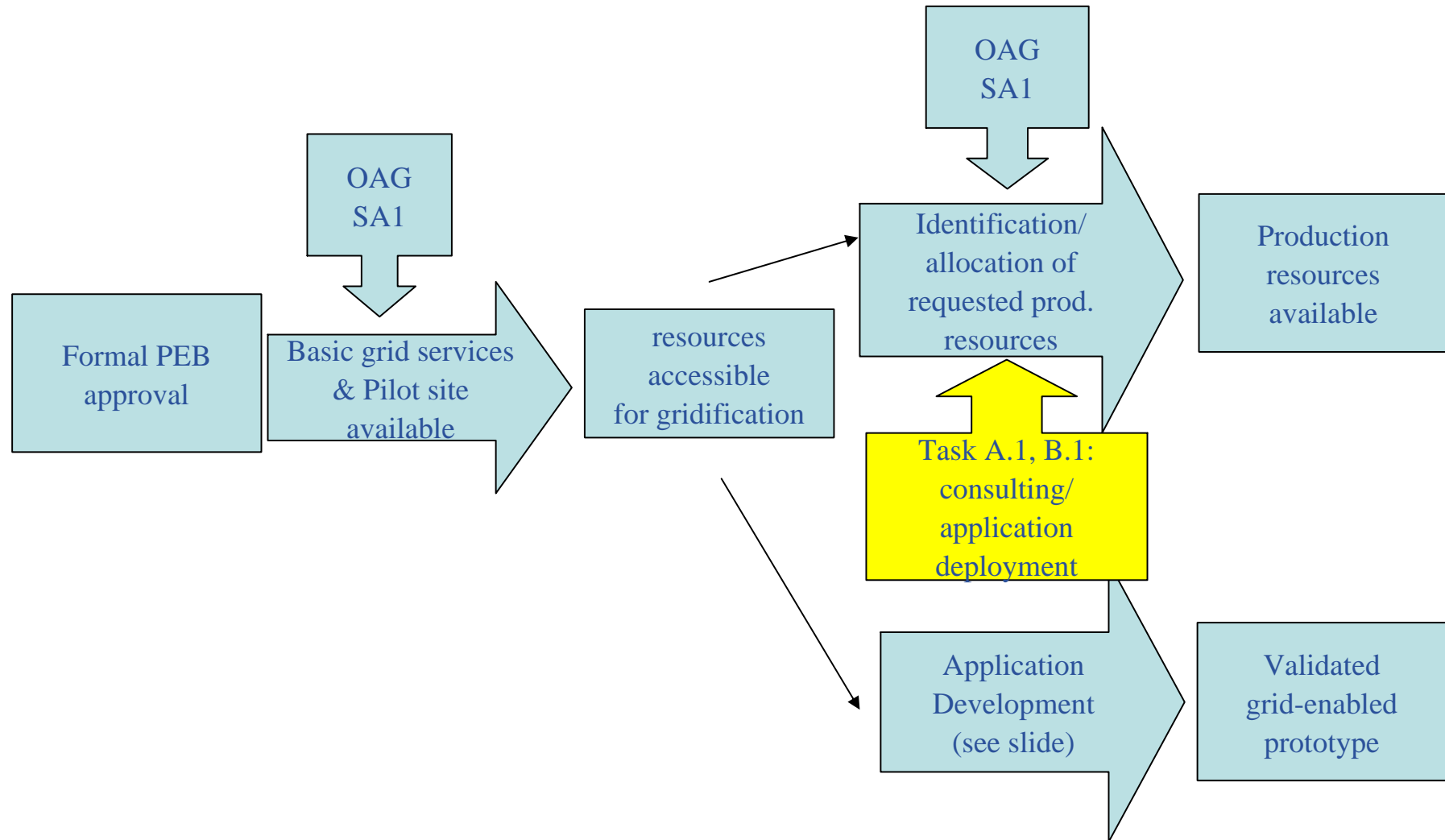


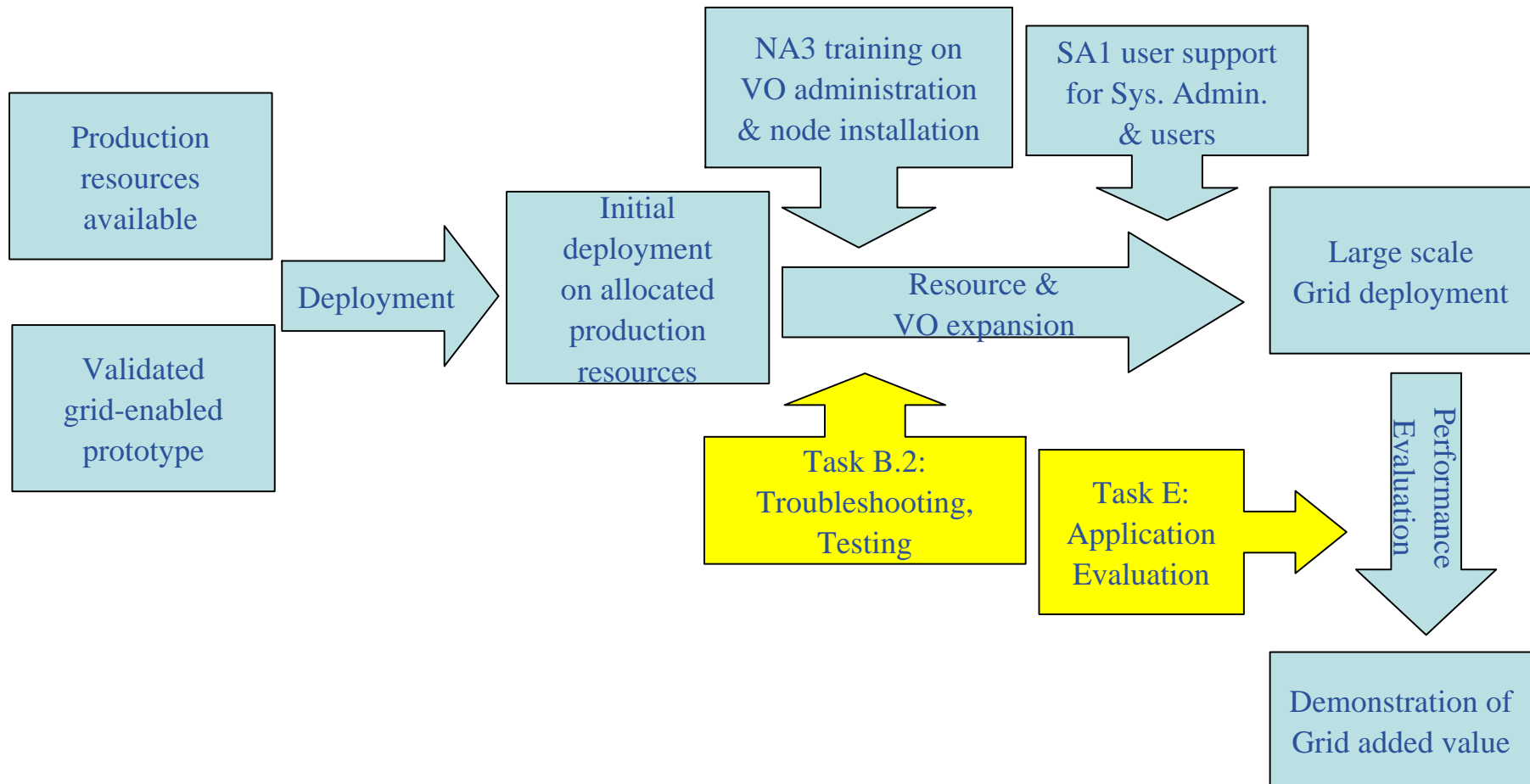










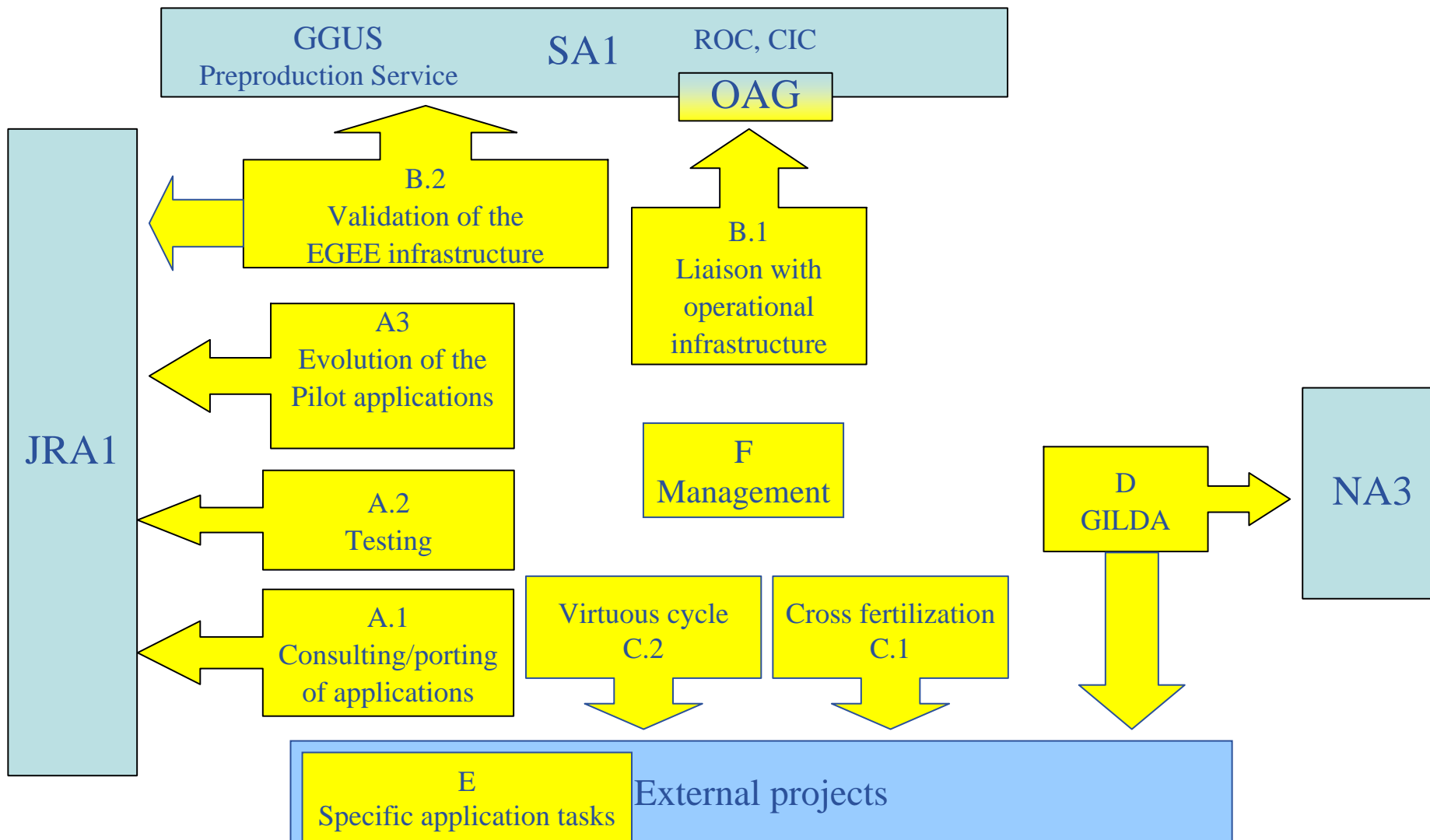


- **Priorities**



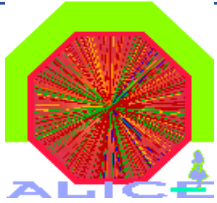
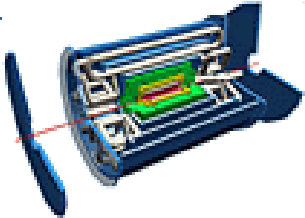

- Strengthening of existing successful teams (ARDA, Biomed task force, Gilda team)
 - Increase ARDA and Biomed task force manpower to address infrastructure validation
 - Increase Gilda team manpower to strengthen outreach
- Coalescence of new technical teams in collaboration with external projects
 - Increase resources to integrate external projects

- **Request: 70 FTEs**

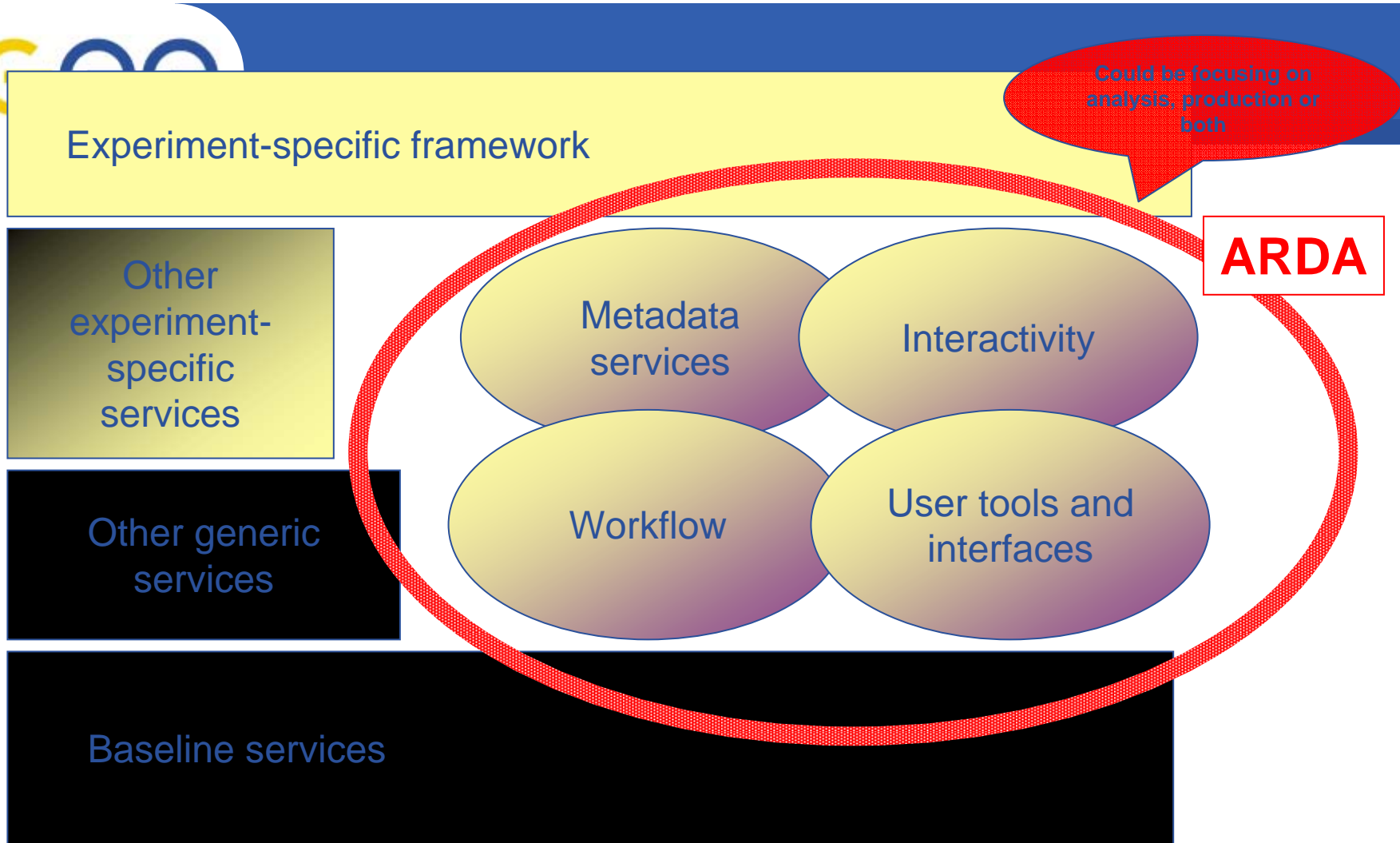
- 21 FTEs for Biomed, including Biomed task force
- 25 FTEs for HEP, including ARDA
- 16 FTEs for Other supported applications, including Gilda
- 8 FTEs for inter-activity coordination, integration of external projects and management issues



- **What should not be changed**
 - Parallel activities together within each experiment
 - Priorities discussed and agreed with each experiment
 - Coherent with their milestones and plans
 - Single team environment (ARDA)
 - Prototype activity instead of top-down approach
 - Deliver a working component of each experiment distributed analysis system
- **What should change**
 - Diminishing low-level testing activities
 - Very useful as a contribution to support gLite development/deployment in 2004/5
 - *Together with other NA4 team (good example of collaboration)*
 - Globally there should be **more** testing
 - Extend activity on production
 - Clear message from all LHC experiments
 - It is happening already now spontaneously
 - Evolve and demonstrate production and analysis environment on top of the baseline service layer

LHC Experiment	Main focus	Basic prototype component /framework	Middleware
	GUI to Grid	GANGA/DaVinci	
	Interactive analysis	PROOF AiiROOT	
	High-level services	DIAL/Athena	
	Explore/exploit native gLite functionality	ORCA	

- **NA4-HEP at CERN**
 - Continue the activity on analysis
 - New line on production systems
 - How do these activities map onto the EGEE II draft proposal?
 - Main activities:
 - Evolution of pilot application
 - *The ARDA activity explores/develop together with the experiments crucial missing components*
 - Validation of the EGEE infrastructure
 - *The delivery of the distributed analysis systems (together with the experiments) provide the validation of the system as a whole*
- **Russian contribution to NA4-HEP**
 - Detailed inventory of the activity still going on
 - Positive feedback so far from the LHC collaboration
 - Need of a better integration for efficiency
 - *Partner vs experiments reporting lines*
 - *Bit more centralised reporting line to enable the LHC collaboration to have a better link with these activities*
 - *NA4-HEP head should collect activity reports (Timesheets, QReports?)*



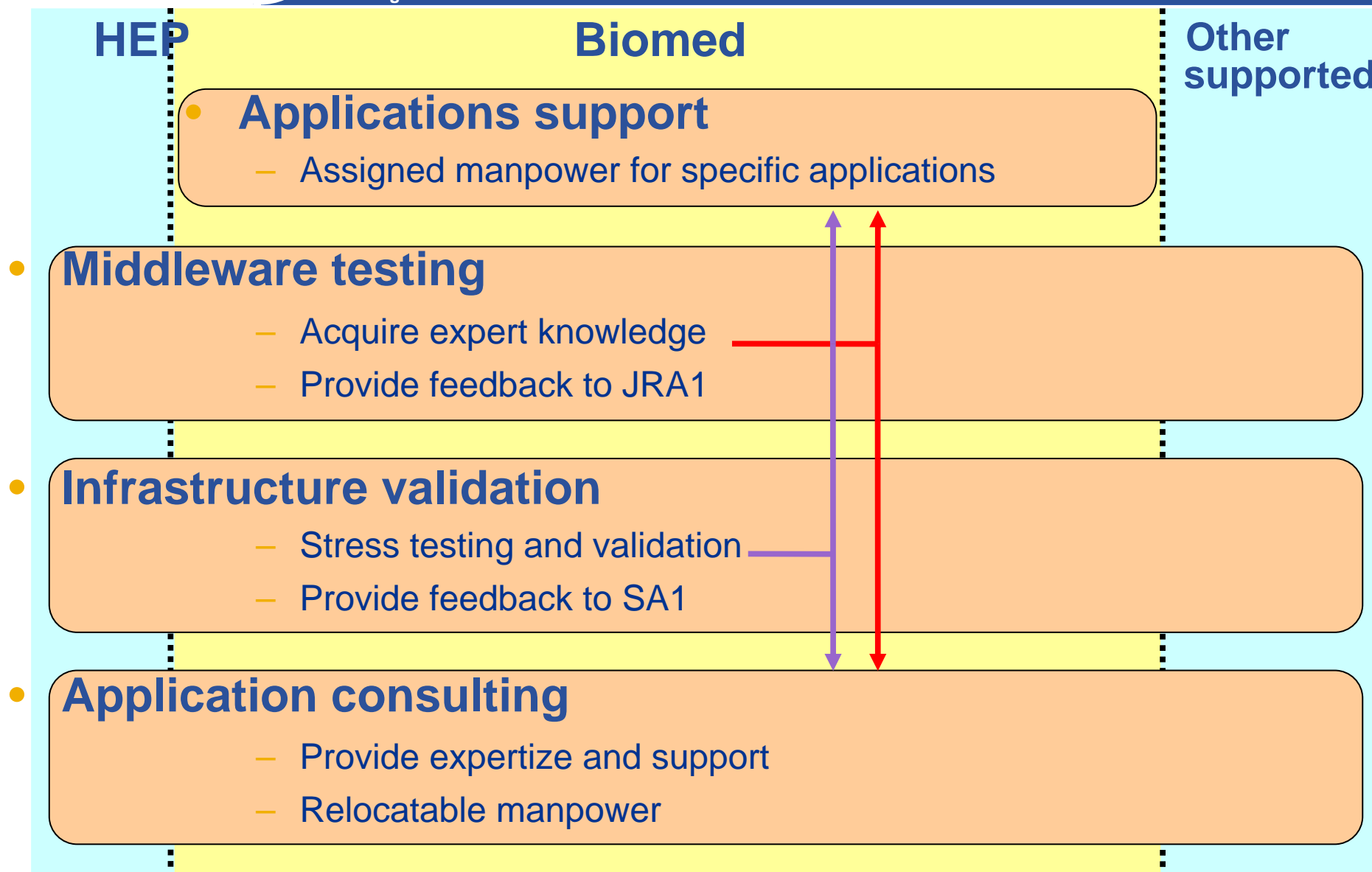
- **Developments in ARDA**
 - integral part of the experiments system
 - developed in conjunction with each experiment
 - The services mentioned here are just examples



Biomedical applications portfolio ready for EGEE2 – J. Montagnat

- **Flagship applications**
 - **Drug discovery**: currently deploying the first biomed Data Challenge
 - Molecular docking for new drug discovery (Malaria target)
 - Infrastructure stress testing last week + DC simulation on-going this week
 - DC starts on July 11th.
 - **Bronze standards**: Taverna workflow managers on top of EGEE
 - medical image registration algorithms assessment
 - EGEE-MyGrid-EMBRACE collaboration
 - **gPTM3D**: first EGEE review
 - Radiological images interactive manipulation
 - application-level scheduling on top of EGEE middleware
- **And a variety of applications representing several user communities**
 - Bioinformatics: **GPS@** web portal, **SPLATCHE** genome evolution
 - Molecular analysis: **xmipp_*** for electron microscopy
 - Medical imaging: **GATE** radiotherapy planning, **SiMRI3D** MRI simulator, **pharmacokinetics**
 - Knowledge based: **CDSS** expert system

EGEE1	EGEE2
<ul style="list-style-type: none"> • Application requirements <ul style="list-style-type: none"> – Collection – Prioritization 	<ul style="list-style-type: none"> – Updates – Broad coverage and testing <ul style="list-style-type: none"> ▪ Security ▪ Interactivity ▪ Metadata ▪ Parallelism...
<ul style="list-style-type: none"> • Application deployment <ul style="list-style-type: none"> – From prototypes – First data challenge 	<ul style="list-style-type: none"> – Regular usage of the grid – Expose biomed services
<ul style="list-style-type: none"> • Demonstrate the relevance of grids <ul style="list-style-type: none"> – Attract users/applications 	<ul style="list-style-type: none"> – Ease access, interfaces
<ul style="list-style-type: none"> • Produce scientific results <ul style="list-style-type: none"> – Demonstration/validation 	<ul style="list-style-type: none"> – long term storage and exploitation



- **All EGEE-1 supported communities should be supported also in EGEE-2 (otherwise the loss of image would be enormous!)**
 - Compchem
 - MAGIC
 - ESR/EGEODE
 - Planck
 - E-GRID
- **EGEE-2 needs to establish good relationships with applications coming from external projects from day 1**
- **More room should be given to other application requirements in EGEE-2 boards and fora**
- **According to what's going on in the first negotiation meetings of the new projects, EU is less and less keen to consider HEP applications as strategic in the rest of FP6 and especially in FP7**
- **In EGEE-2 we must really re-consider our past attitude to ask EU for money for all sciences and spend it in a way that results are useful only for HEP**