

# EGEE

## Enabling Grids for E-science in Europe

**EGEE all-activity meeting:  
status of NA4**

V. Breton on behalf of NA4



# Outline

- Presentation of general NA4 issues (V.B.)
  - Objectives of the activity
  - Partners involved
  - Milestones and deliverables
  - Breakdown of work
  - Risk assessment
  - Relation with other activities
  - Next steps before project conference
- Presentation of HEP specific issues (Frank Harris)

**Merci Guy**

# NA4: Identification and support of early-user and established applications on the EGEE infrastructure

- To identify through the dissemination partners and a well defined integration process a portfolio of early user applications from a broad range of application sectors from academia, industry and commerce.
- To support development and production use of all of these applications on the EGEE infrastructure and thereby establish a strong user base on which to build a broad EGEE user community.
- To initially focus on two well-defined application areas – Particle Physics and Life sciences.

# Different families of applications on EGEE

- "pilot" applications
  - used to test EGEE middleware and to evaluate performances
  - under the responsibility of NA4 funded partners in charge of HEP and biomed
  - starting at project day 0.
- "internal" applications
  - come from within the project in the sense that they involve EGEE partners in collaboration with institutes external to EGEE (ex GPCALMA, Babar, UK e-science projects,...)
  - have already a good middleware experience.
  - should be identified as they are often deployed at a national level and are therefore extremely dependent on interoperability between EGEE and national initiatives.
  - however, it must be clear NA4 can not commit to support them
  - What about VOs and CPU cycles ?
- "external" applications.
  - from collaborations external to EGEE (ex DILIGENT, eHealth FP6 projects, Korean Grid)
  - need training/support from NA3/NA4 for deployment on EGEE.

Activity NA4	Total effort (FTE) during 2 years	Total effort (PM) during 2 years	1 <sup>st</sup> year effort (PM)	2 <sup>nd</sup> year effort (PM)	Activity
<b>CERN</b>	16	192	96	96	HEP
<b>CESNET</b>	3.2	38.4	19.2	19.2	Generic
<b>MTA</b>	0.8	9.6	4.8	4.8	Generic
<b>UEDIN</b>	2	24	12	12	NA3 liaison
<b>CNRS</b>	20	240	120	120	Bio+Coord+Test
<b>CSSI</b>	4	48	24	24	Testing team
<b>CRSA</b>	4	48	24	24	Industry forum
<b>DKRZ</b>	2	24	12	12	Generic
<b>FhG</b>	2	24	12	12	Generic
<b>INFN</b>	8	96	48	48	Generic
<b>FOM</b>	4	48	24	24	Generic
<b>IHEP</b>	3.2	38.4	19.2	19.2	HEP
<b>IMPB RAS</b>	2.6	31.2	15.6	15.6	Bio
<b>ITEP</b>	4	48	24	24	HEP
<b>JINR</b>	3.2	38.4	19.2	19.2	HEP
<b>PNPI</b>	3.2	38.4	19.2	19.2	HEP
<b>RRC KI</b>	4	48	24	24	HEP
<b>SINP-MSU</b>	4	48	24	24	HEP
<b>CSIC</b>	4	48	24	24	Bio
<b>UPV</b>	4	48	24	24	Bio
<b>TOTAL</b>	98.2	1178.4	589.2	589.2	

# Roles and staffing

<u>Federation</u>	<u>Role</u>	<u>FTE Requested</u>	<u>FTE Unfunded</u>	<u>Financing Requested</u>
CERN	HEP Applications (coord.)	4	4	800
UK+Ireland	NA3 Liaison	0,5	0,5	94
Italy	Generic app (coord)	2	2	400
France	General coord., BioMed, Test team, Industry diss.	7	7	1392
Northern Europe	Generic applications	1	1	194
Germany + Switzerland	Generic applications	1	1	199
Central Europe	Generic applications	1	1	60
South West Europe	BioMed	2	2	348
Russia	HEP, BioMed	3	3	151
<b>Totals</b>		<b>21,5</b>	<b>21,5</b>	<b>3608 k€</b>

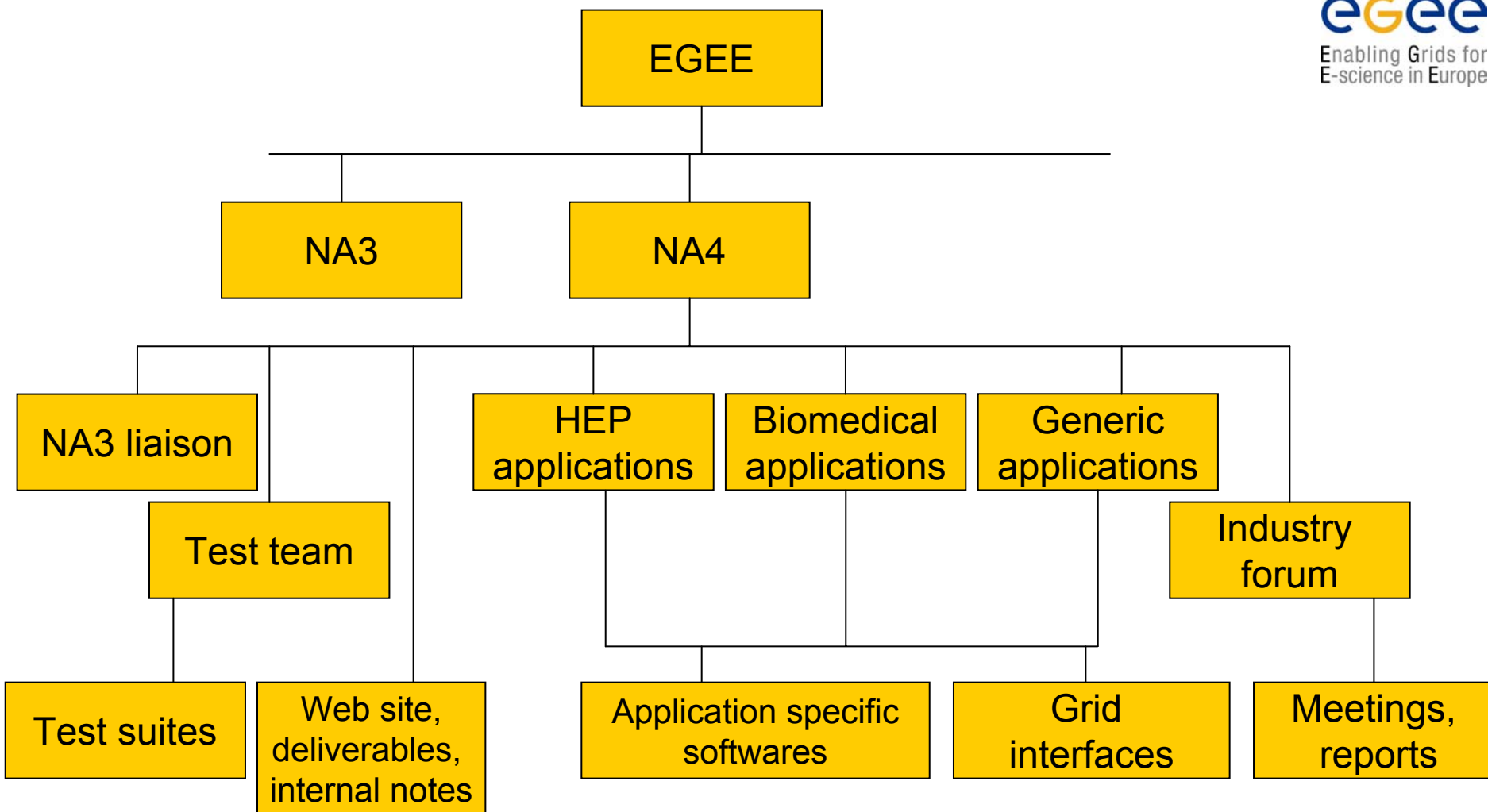
# Milestones

Milestones and expected result		
MNA4.1	M6	First applications migrated to the EGEE infrastructure
MNA4.2	M12	First external review of Applications Identification and Support with feedback
MNA4.3	M24	<ul style="list-style-type: none"> <li>• Second external review of Applications Identification and Support with feedback</li> </ul>

# Deliverables

Deliverables		
DNA4.1	M3	Definition of Common Application Interface and Planning Document
DNA4.2	M6	Target Application Sector Strategy document
DNA4.3	M9	<ul style="list-style-type: none"> <li>•EGEE Application Migration Progress report (revision M15 and M21)</li> </ul>
DNA4.4	M24	Final Report of Application Identification and Support Activity



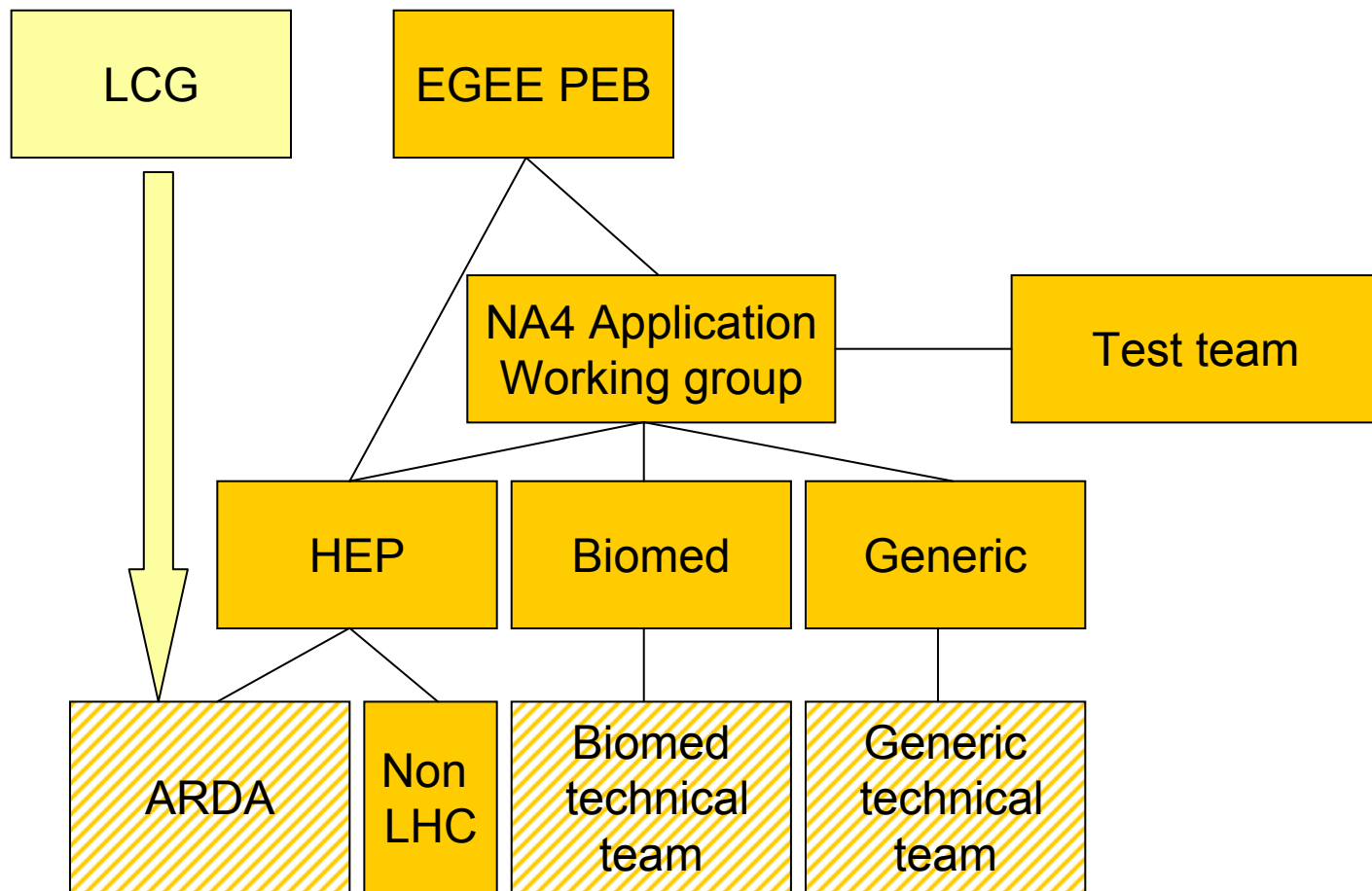


# List of tasks

Tasks	Month start	Month end	Resources allocated
<b>4.1 Consolidation of the existing requirements</b>	1	1	4
<b>4.2 Basic foundations of common application interface and associated roadmap (DNA4.1)</b>	1	3	21
<b>4.3 Definition of the strategy for application integration and deployment (DNA4.2)</b>	1	6	28
<b>4.4 Migration of first applications to EGEE infrastructure (DNA4.3.1)</b>	1	12	231
<b>4.5 Deployment of first generic applications (DNA4.3.1)</b>	1	12	116
<b>4.6 NA4 management</b>	1	12	21
<b>4.7 Test suite</b>	1	12	39
<b>4.8 Industry forum</b>	1	12	14

Task	HEP	Biomed	Generic	Test/NA3 liaison /CSSA	Coordination	Total
Tx NA4.1	CERN (1PM)	CNRS(1PM)	INFN(1PM)		CNRS(1PM)	4PM
Tx NA4.2	CERN (5PM)	CNRS(4PM) UPV/CSIC (1PM)	INFN(3PM)		CNRS(8PM)	21PM
Tx NA4.3	CERN (5PM)	CNRS(4PM) UPV/CSIC (1PM)	INFN(8PM)	UNEDI(3PM)	CNRS(7PM)	28PM
Tx NA4.4	CERN (4PM) +HEP (141PM)	CNRS(6PM) +UPV/CSIC (1PM) + bio (78PM)		UNEDI(2PM)	CNRS(6PM)	238PM
Tx NA4.5			INFN(18PM)+ generic (80PM)	UNEDI(7PM)	CNRS(11PM)	116PM
Tx NA4.6	CERN (3PM)	CNRS(3PM)	INFN(3PM)		CNRS(12PM)	21PM
Tx NA4.7	CERN (1PM)	CNRS (1PM)	INFN (1PM)	CNRS(12PM) CSSI (24PM)		39PM
Tx NA4.8				CSSA(12PM)	CNRS (2PM)	14PM
<b>TOTAL</b>	<b>CERN (19PM) + 141PM for application migration =160PM</b>	<b>CNRS (19PM) + UPV/CSIC (3PM) + 78PM for application migration =100PM</b>	<b>INFN(34PM) + 80PM for applicati on migratio n = 114PM</b>	<b>UNEDI(12PM) CNRS(12PM) CSSI(24PM) CSSA(12PM)= 60PM</b>	<b>CNRS(47PM)</b>	<b>481PM</b>

# NA4 technical organization



# NA4 Application Working Group

- Technical overview of application deployment
  - Virtual organizations (Biomed, Generic)
  - Relationship with external/internal/pilot applications
- Preparation of NA4 deliverables
  - PM3: definition of a common application layer
- Technical relationship with other application-oriented projects
  - EU projects: Gridlab, Diligent, ...
  - National grid projects: Korea Grid initiative
  - ...
- Small group
  - 2 representatives (including application manager) of each application area (HEP, Biomed, Generic)
  - 2 representatives of NA4 coordination
  - 1 representative of test team

# Training requirements

- HEP
  - addressed partly/mostly within LCG for LHC experiments
  - Specific needs to address ? (non LHC, EGEE middleware)
- Biomedical community has important needs
  - External projects are mostly beginners (except GEMSS and Mammogrid)
  - Funded partners would benefit training to set up technical team
  - Internal applications would certainly benefit of training sessions to disseminate expertise
- Generic
  - Groups applying have not all the same previous experience of grids
  - Internal applications would certainly benefit of training sessions to disseminate expertise

# Selection process for external applications

- Need for an attractive and orderly integration process
- Proposal : EGEE Generic Applications Advisory Panel (EGAAP) selecting applications on the following criteria :
  - scientific interest of the proposed work, with particular emphasis on the grid added-value,
  - added value for EGEE to have such an application running on its infrastructure
  - coordination of the corresponding community,
  - grid-awareness of this community
  - minimum requirement that a small team followed the EGEE training, dedication of the community to this application,
  - agreement to the various EGEE policies and especially the security and resources allocation policies.
- EGAAP described in EDMS document  
<https://edms.cern.ch/document/451584/2>
- Need for a quick convergence (TODAY !!!)

# Industry forum (C. Saguez, G. Wormser)

- Steering committee well in place
  - Members : HP, IBM, Microsoft, Datamat, Nice, Gridexpert, U. Warsaw +CS+GW
- Agreed by-laws and financing rules
- Web site accessible from EGEE public page
- Preparation of Cork Industry Forum day essentially done
  - Most of the industrial participants to the two round tables are identified
  - Agreement with Datagrid and Crossgrid Industry Fora
  - Invitations are going out
- Expectations at Cork
  - Increase membership (currently 25 members)
  - Start the work in two working groups: industry-related technology issues and business models



# NA4 initial quality indicators (from G. Zaquine, EDG)

- QoS - Quality indicators specifications: <https://edms.cern.ch/document/386039>
  - Performance indicators (per virtual organization)

$$E_{crude} = \frac{\text{Number of jobs successfully completed}}{\text{Total Number of jobs submitted}}$$

$$E_{Exec} = \frac{\text{Time while job is running}}{\text{Total time between submission and completion}}$$

$$E_{System} = \frac{\text{Resources Delivered}}{\text{Min (Resources Requested, Total Resources Available)}}$$

- Utilisation (does the testbed fit the needs of the applications)
  - Integrated CPU used as a function of user, VO and site
- Bugzilla follow-up
  - Number of new anomalies / number of pending anomalies.
  - Percentage of anomalies resolution in each classes of time resolution range (low, medium, high).
- User support
  - Percentage of support requests resolved within (x) time - x should be defined (e.g.: 1 hour)

# Risk analysis

- Technical risks
  - Middleware fails to meet requirements of applications
    - Or fails to meet requirements beyond HEP
    - CPU cycles hidden from non LCG VOs
- Management risks
  - NA4 resources outside CERN, CNRS and INFN allocated to “internal” applications
    - Resources lost from project point of view because allocated to invisible tasks (national projects, non EGEE collaborations)
  - NA4 fails to become a team
- Scientific risks
  - Failure to integrate applications coming from external projects

## Issues for JRA1 and Bob : relationship between middleware and NA4 “loose cannons”

- Four NA4 technical teams need close interactions with middleware
  - ARDA members
  - Biomed and Generic technical teams
  - NA4 Test team
- A process is needed to
  - feed back requirements
  - discuss middleware issues (features, bugs)

## Issues for SA1 and Bob: deployment of applications from PM0

- What is the infrastructure to start application deployment at day 0 ?
- What are the policy and entry point to
  - Create new VOs (if LCG2, for HEP non LHC, biomed and generic)
  - Install needed RPMs
  - Have access to resources
- Response urgently needed...

## Issues for JRA3 and Bob: relationship with applications

- Security is one of the bottlenecks for deployment of many applications
- What is project security policy at day 0 ?
- How does it evolve during the project lifetime ?
- Is there a formal mechanism to feed back requirements ?

# Next steps before project conference

N°	Date in	Who	Subject	Deadline	Priority 1 upto 5
1	21/01/04	Frank	Mailing list of extended Steering Committee	Done	
2	21/01/04	Frank/Vincent	Set up dates for extended SC meetings	4/03	4
3	21/01/04	Frank/Roberto	Develop contacts with the GAT group in preparation for the July deliverable	1/04	3
4	21/01/04	Vincent/Frank/Roberto	Review the consolidation of the requirements (issue for the first SC meeting)	1/05	3
5	21/01/04	Guy/Frank/Roberto	Clarify roles and participation of groups remotely connected to NA4 activities (Russian groups, ...)	1/04	3
6	21/01/04	Vincent/Frank/Roberto	Planning of the Cork meeting (issue for the first SC meeting)	18/04	5
7	21/01/04	Vincent	Confirmation of the biomedical pilot applications	15/03	3
8	21/01/04	Roberto	NA4 web site	Done	
9	21/01/04	Vincent	Execution plan	?	5
10.	21/01/04	Roberto/Guy	Definition of process for selection of generic applications	15/3	5
11.	21/01/04	All partners	Hiring of the funded resources	1/04	5

## Preparation of NA4 deliverable DNA4.1 “Definition of Common Application Layer”

- Contacts with SAGA newly created research group at GGF
- Interest for
  - description of operations
  - use cases
  - usages of grids for short and mid/long term
  - extraction of functionalities out of the use cases
- Gridlab interested to provide template for definition of requirements and use cases
- Does this document include a list of requirements for a grid portal ?