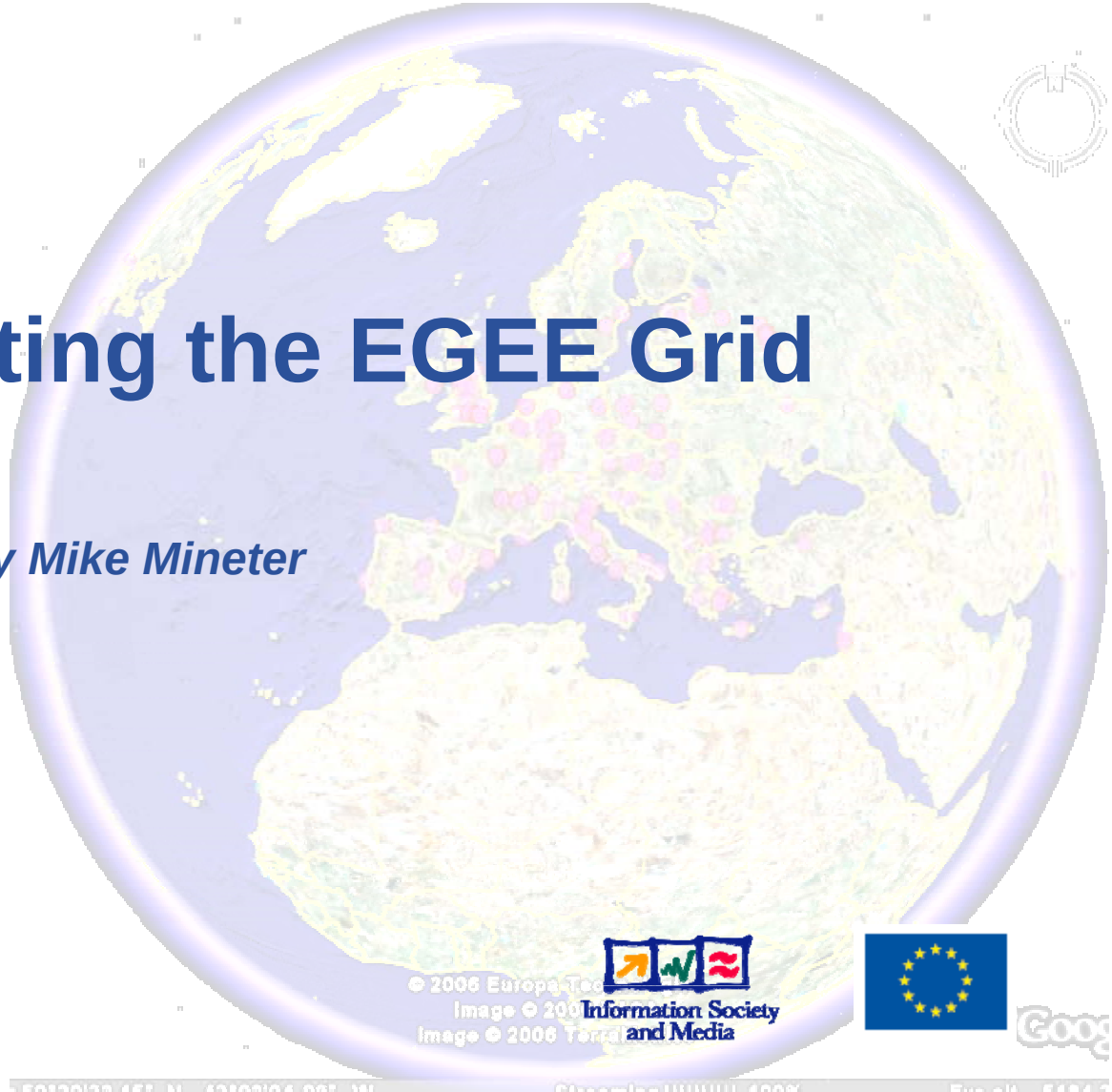
Operating the EGEE Grid

Presented by Mike Mineter



www.eu-egee.org

Outline

A selection of slides from a talk by the EGEE Operations Manager, Ian Bird to the EGEE Conference in Geneva.

http://indico.cern.ch/materialDisplay.py?contribId=262&sessionId=12&materialId=slides&confId=1504

Included in this event as an overview of the Operations activity.

The EGEE Production Grid

A Bird's-Eye View

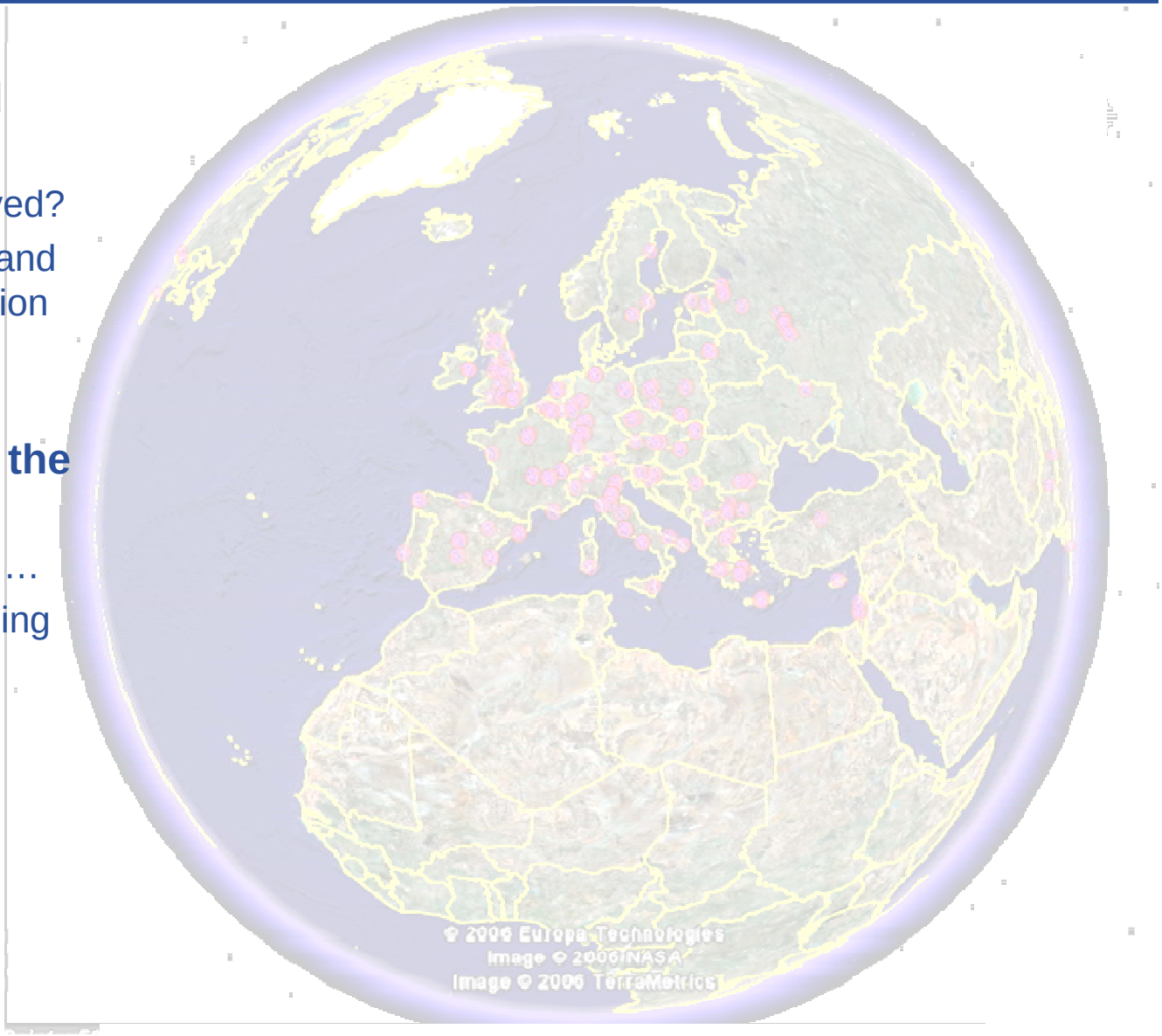
Ian Bird
EGEE Operations Manager

EGEE'06

Geneva, 27th September 2006

Outline

- **What is the EGEE grid infrastructure today?**
  - What has been achieved?
  - How does it compare and relate to other production grids?
- **What is the outlook in the short term?**
  - Timescale of EGEE-II …
  - What are the outstanding issues?
- **What should happen next?**



Some history … LHC → EGEE Grid

- **1999 – Monarc Project**
  - Early discussions on how to organise distributed computing for LHC
- **2000 – growing interest in grid technology**
  - HEP community was the driver in launching the DataGrid project
- **2001-2004 - EU DataGrid project**
  - middleware & testbed for an operational grid
- **2002-2005 – LHC Computing Grid – LCG**
  - deploying the results of DataGrid to provide a production facility for LHC experiments
- **2004-2006 – EU EGEE project phase 1**
  - starts from the LCG grid
  - shared production infrastructure
  - expanding to other communities and sciences
- **2006-2008 – EU EGEE-II**
  - Building on phase 1
  - Expanding applications and communities …
- **… and in the future – Worldwide grid infrastructure??**
  - Interoperating and co-operating infrastructures?

Some history …

- **EGEE grew out of the EDG and LCG projects**
  - LCG built the first production middleware distributions and set up the initial grid infrastructure → became EGEE in 2004
  - HEP (and LHC in particular) are very strong drivers for EGEE
  - Branding is changing slowly (gLite-3.0)

- **Difficult to get the right balance:**
  - LCG is pushing the boundaries
    - Data sizes and rates
    - Workloads 50k jobs/day now; ~500k jobs/day in 1 year
    - MoU for service reliability/availability is first real SLA
  - Crucial that other applications push as hard
    - Biomedical – application security aspects
    - And others with their own requirements …
    - Data challenges are a good way to move forward: LCG, Wisdom

- **Is the balance right?**
  - Everyone complains equally …

The EGEE Infrastructure

Test-beds & Services

- Certification testbeds (SA3)
- Pre-production service
- Production service

Support Structures

- Operations Coordination Centre
- Regional Operations Centres
- Global Grid User Support
- EGEE Network Operations Centre (SA2)
- Operational Security Coordination Team

Infrastructure:

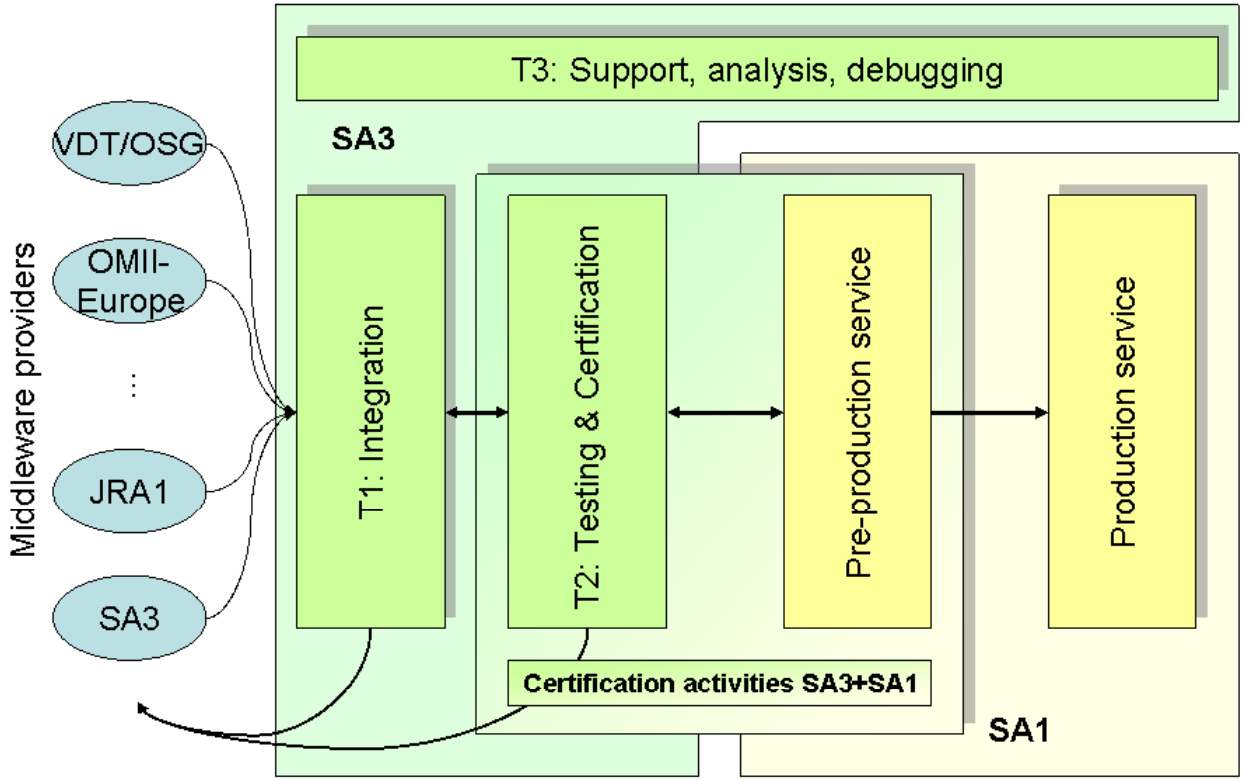
- Physical test-beds & services
- Support organisations & procedures
- Policy groups

Security & Policy Groups

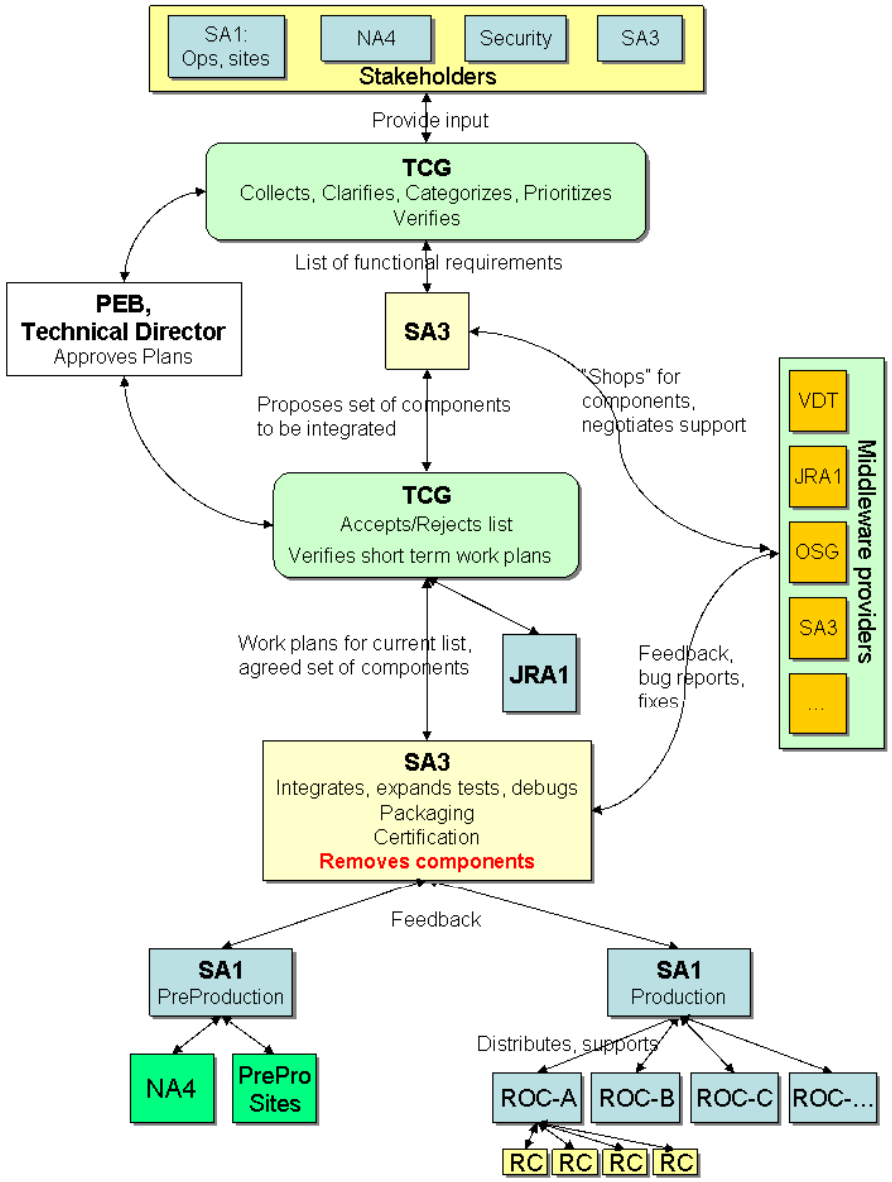
- Joint Security Policy Group
- EuGridPMA (& IGTF)
- Grid Security Vulnerability Group
- Operations Advisory Group (+NA4)

Certification & release preparation

- **The goal is to produce a *middleware distribution* that can be deployed widely**
  - Not the same as middleware releases from development projects
  - More like a Linux distribution – bringing together many pieces from several sources



Middleware release



- **Technical Coordination Group**
  - Agrees the contents and priorities for what goes into the integration and testing process
- **Not all desired new components or updates may make the next distribution**
  - Depends on priorities and urgency for other pieces
- **Moving away from big-bang releases to component upgrades**
  - Concept of a baseline release and then updates and patches
  - New baseline when significant changes (dependencies, …)

Certification

- **Extensive certification test-bed:**
  - Close to 100 machines involved
  - Main test-bed at CERN, test-beds for specific tasks at SA3 partner sites
- **Emulate the deployment environments**
  - Or at least the main ones …
- **Certification testing:**
  - Installation and configuration
  - Component (service) functionality
  - System testing (trying to emulate real workloads and stress testing)
  - Beginning to use virtualization to simplify the testing environment
- **Deployment into the pre-production system**
  - Final step of certification – validation by real sites
  - Validation by applications – also allows to prepare apps for new versions

Snapshot of (part of) certification tb

Enabling Grids for E-sciencE

	a	b	c	d	e	f	g	h	i	j	k	l
0		2057 UIcomb Configured	1765 UIcomb Configured	1762 RB Configured				1738 VOBOX Configured			config	1778 UI Configured
1	2016 RB Configured		lxb2020 PX Configured	2032 WMS Configured		0744 WMS Configured		1928 VOMS Configured		stable		1919 gLiteCE Configured
2	lxb2017 BDII Configured			2033 BDII Configured		2				1917 UI Configured		1912 WMS Configured
3	1A			1B		0730 RB Configured				1794 RB Configured		1779 CE Configured
4	2018 CE Configured			1905 gLiteCE Configured		0743 gLiteCE Configured		lxb0714 SFT Configured		1936 BDII Configured		1776 DPM_mysql Configured
5	1921 DPM_mysql Configured			2034 CE Configured		2035 CE Configured				1938 CE Configured		0727 DPM_pool Configured
6				1720 WNcomb Configured		lxb2036 DPM_mysql Configured				1915 DPM_mysql Configured		1777 dCache_mysql Configured
7	0731 WNcomb Configured			1716 WNcomb Configured						1916 DPM_pool Configured		1775 SEclassic Configured
8	lxb1909 FTS Configured			0735 WNcomb Configured		0724 dCache_mysql Configured		opalpro17 CE Configured		1751 dCache_mysql Configured		0718 WNcomb Configured
9	2019 HON Configured			lxb1917 DPM_mysql Configured		0741 WNcomb Configured		opalpro16 SEclassic Configured		1758 WNcomb Configured		
10	1941 LFC_mysql Configured			lxshare0297 dCache_mysql Configured				opalpro23 WNcomb Configured				1774 VOBOX Configured

Middleware releases & deployment

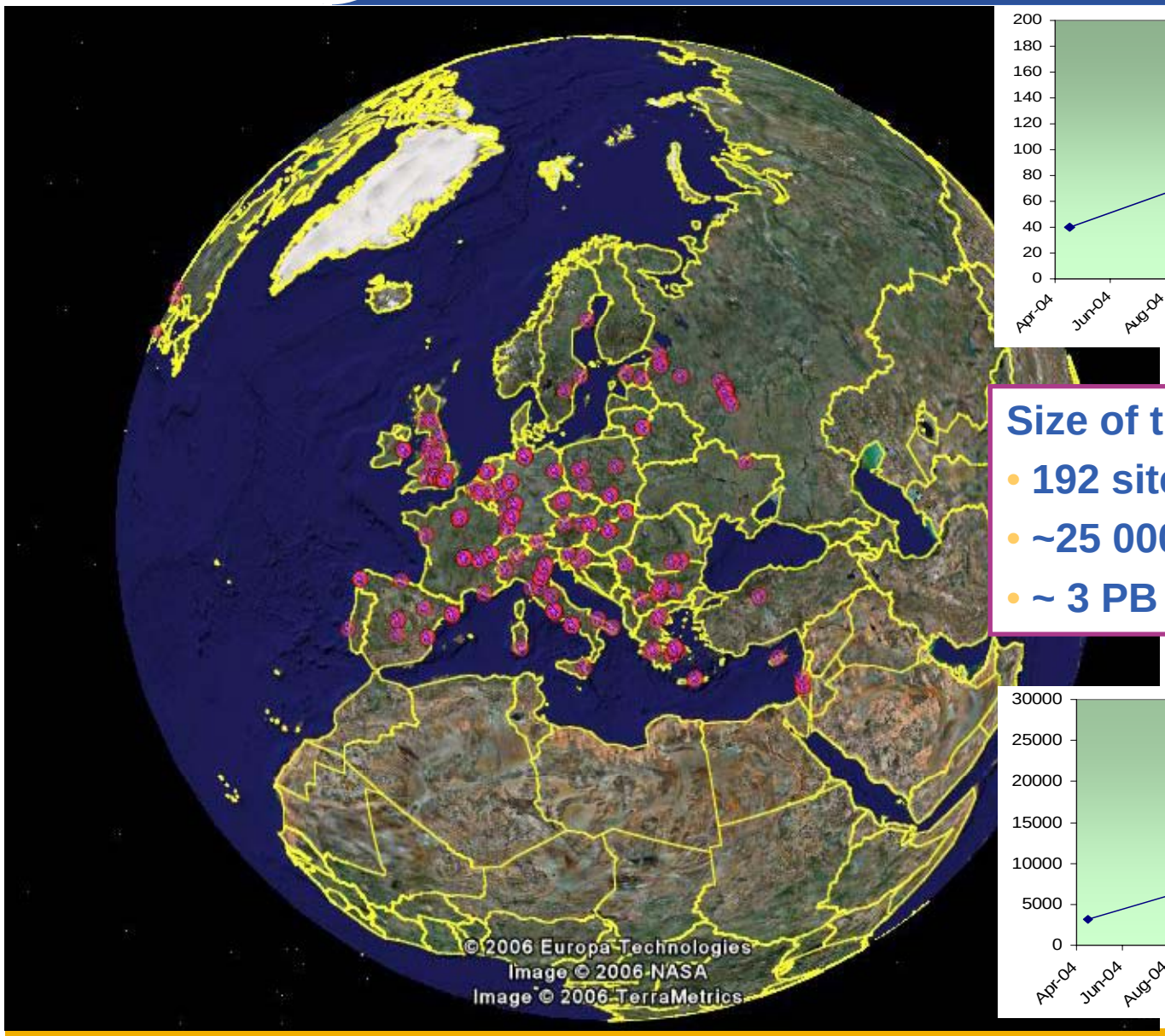
- **Once a distribution is ready for deployment, it takes several months to get this to the majority of sites**
  - Seems to be a constant
  - Advantage of decoupling the components – VOs can encourage sites to update the pieces they require
  - Client tools can be simply installed (remotely) even without site upgrade

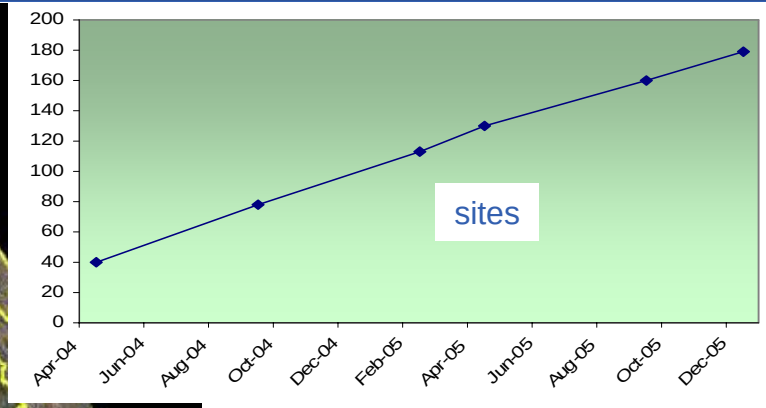
- **Deployment onto the EGEE infrastructure is managed and supported by the Regional Operations Centres**

Pre-production service

- **Pre-production service is now ~ 20 sites**
- **Provides access to some 500 CPU**
  - Some sites allow access to their full production batch systems for scale tests
- **Sites install and test different configurations and sets of services**
  - Try to get good feeling for the quality of the release or updates before general release to production
  - Feedback to: certification, integration, developers, etc.

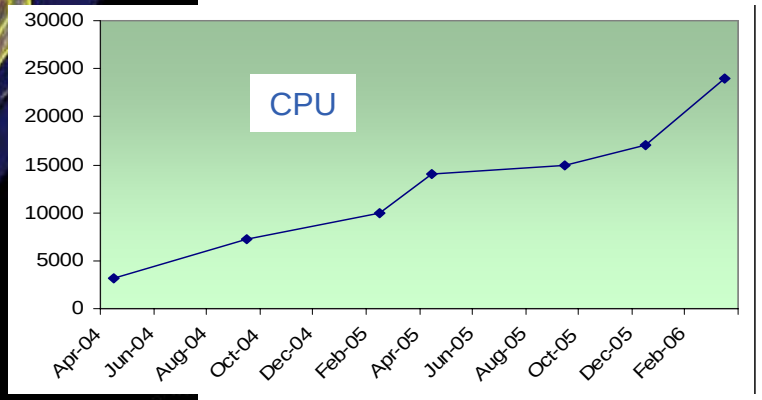
Production service





Size of the infrastructure today:

- **192 sites in 40 countries**
- **~25 000 CPU**
- **~ 3 PB disk, + tape MSS**

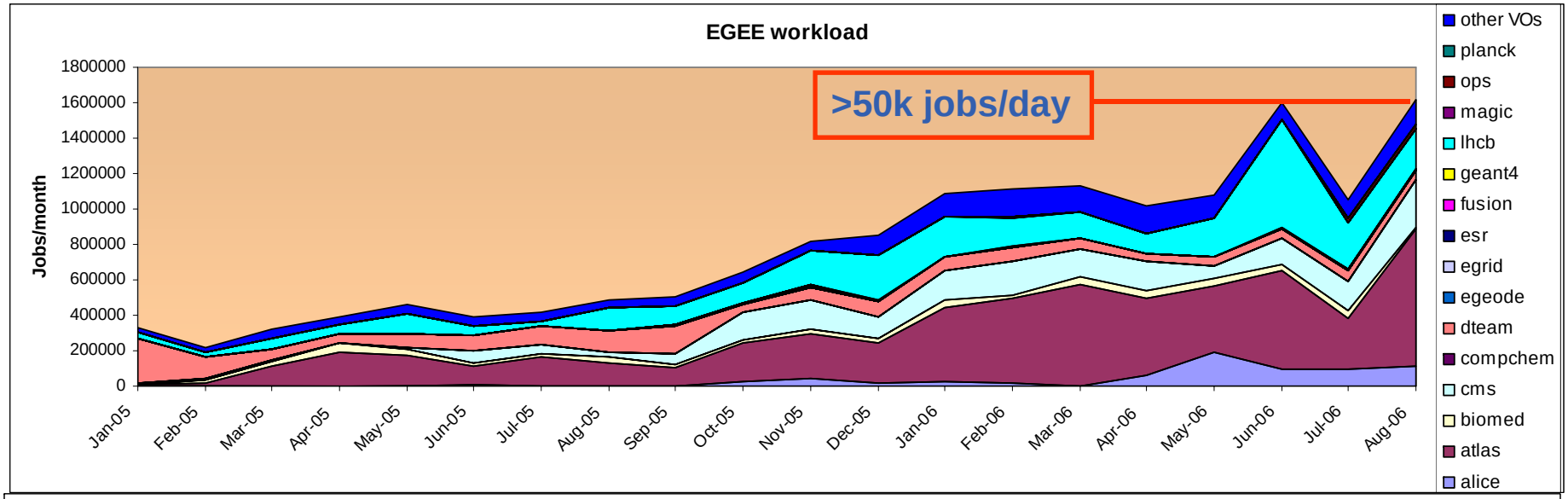


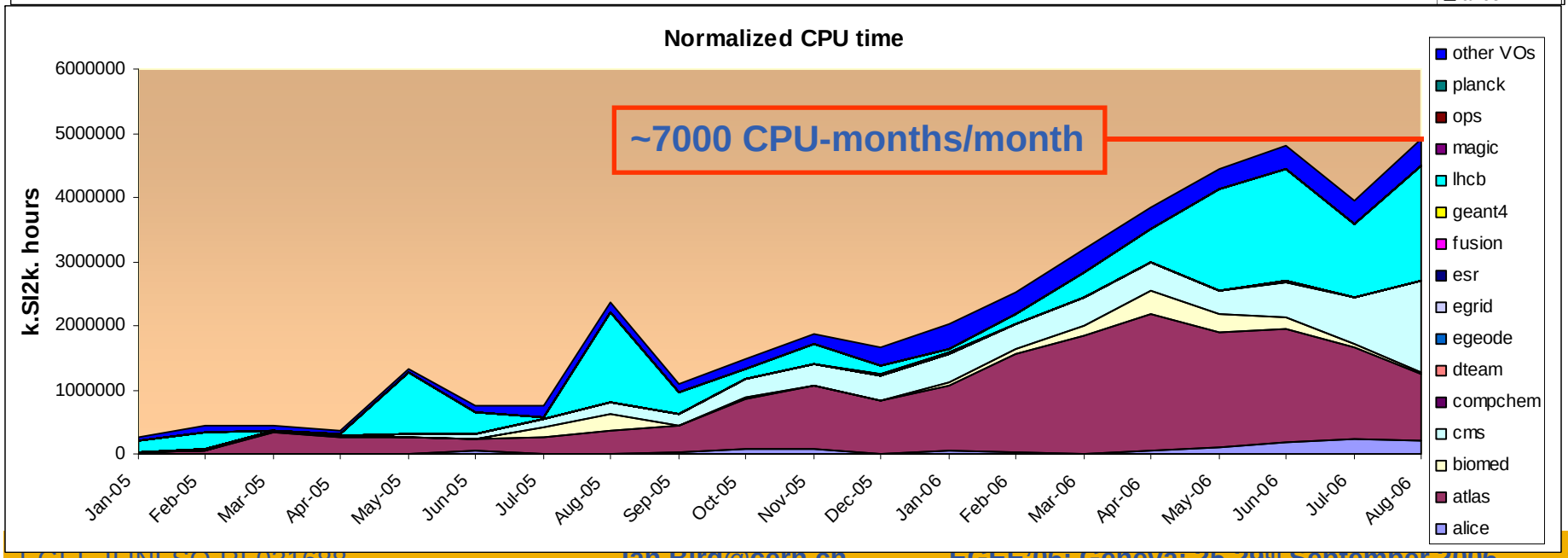
EGEE Resources

Region	#countries	#sites	#cpu	#cpu DoW	disk (TB)
CERN	0	1	4400	1800	770*
UK/I	2	23	4306	2010	310
Italy	1	27	2800	2280	373
France	1	10	2316	1252	300*
De/CH	2	13	2895	1852	280*
Northern Europe	6	16	2379	1860	64
SW Europe	2	13	956	898	16*
SE Europe	8	26	1101	1189	30
Central Europe	7	21	1584	1163	70
Russia	1	15	515	445	38
Asia-Pacific	8	19	840	751	72
North America	2	8	4069	-	229
Totals	40	192	28161	20265	2552

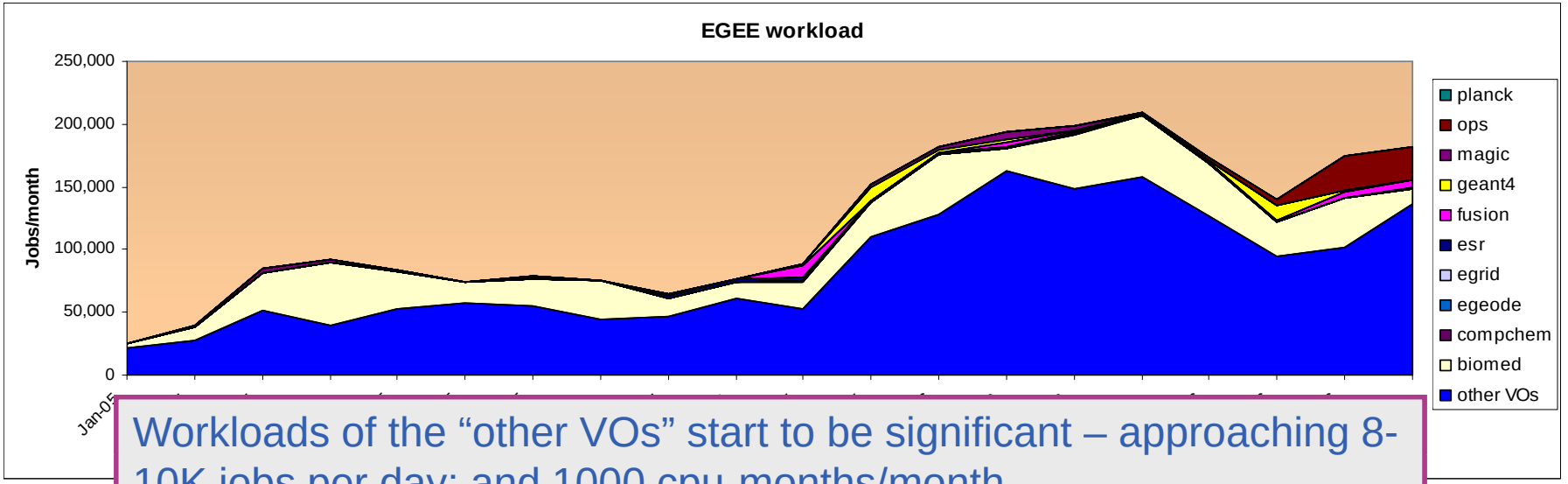
* Estimates taken from reporting as IS publishes total MSS space

Usage of the infrastructure



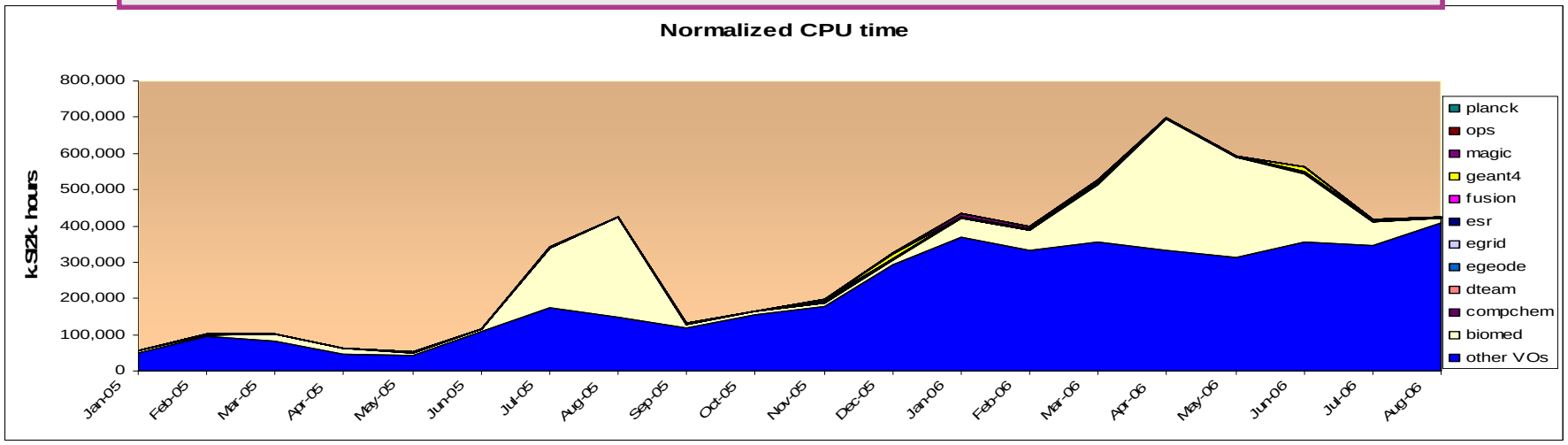


Non-LHC VOs

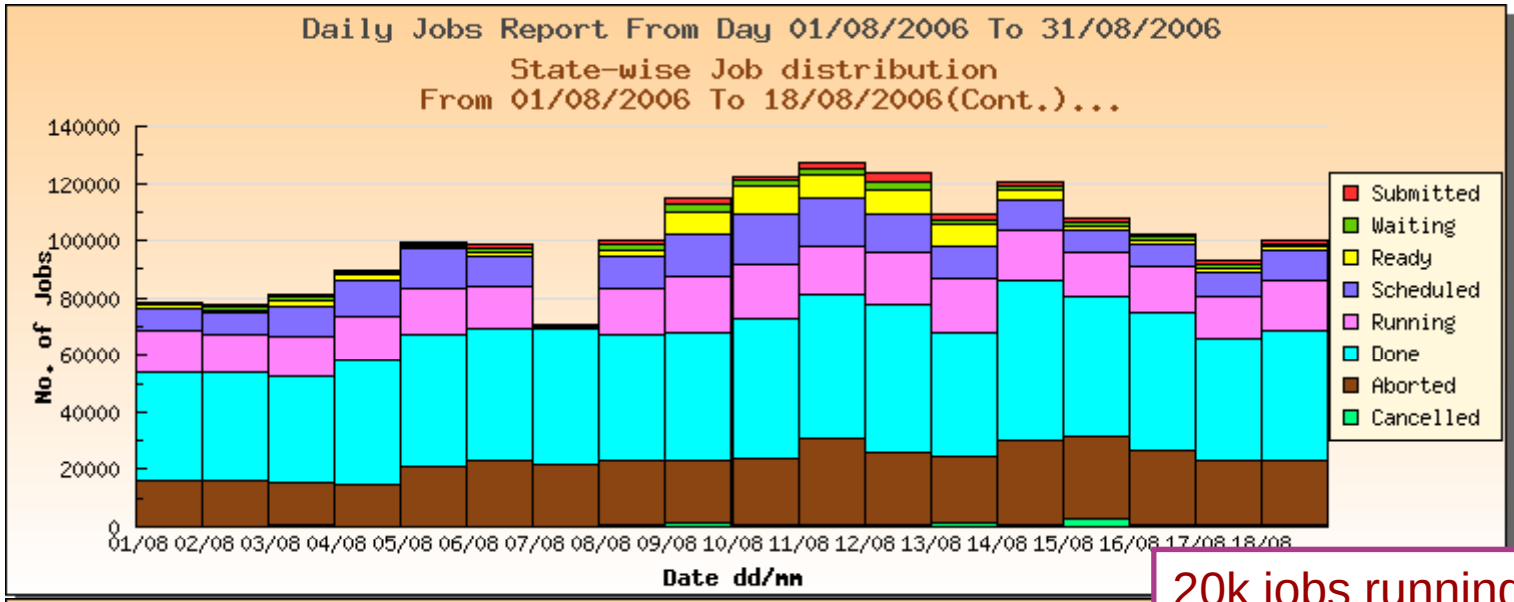


Workloads of the "other VOs" start to be significant – approaching 8-10K jobs per day; and 1000 cpu-months/month

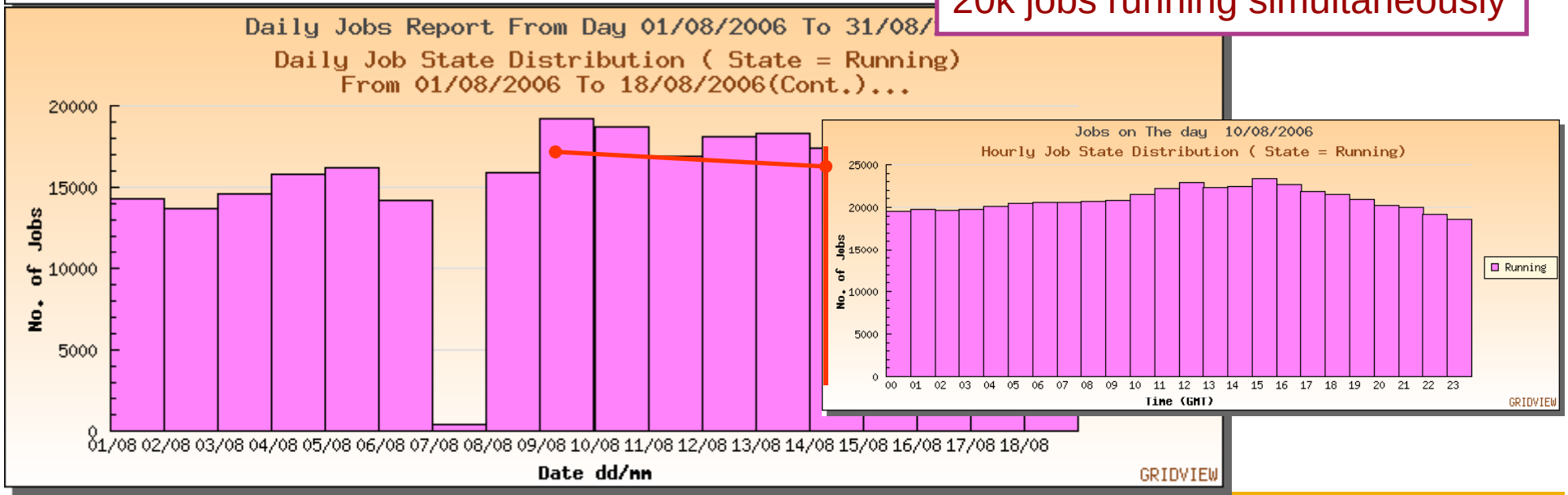
- one year ago this was the overall scale of work for *all* VOs



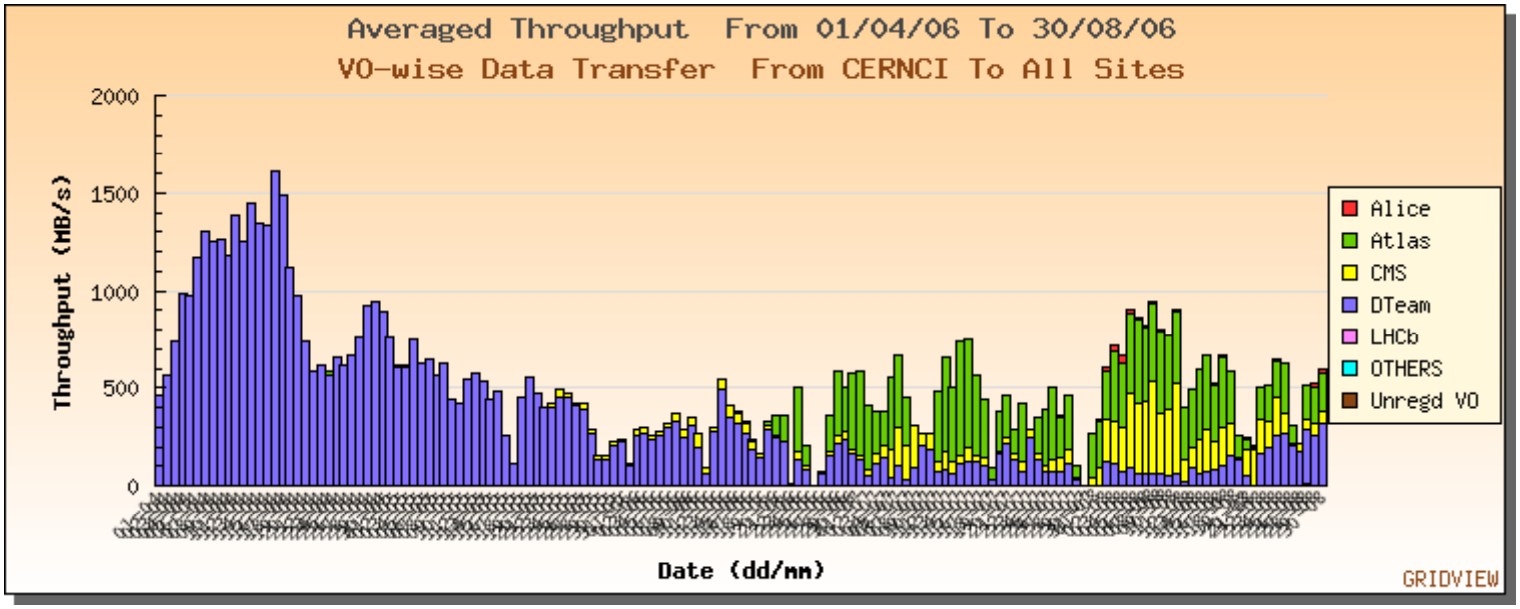
Use of the infrastructure



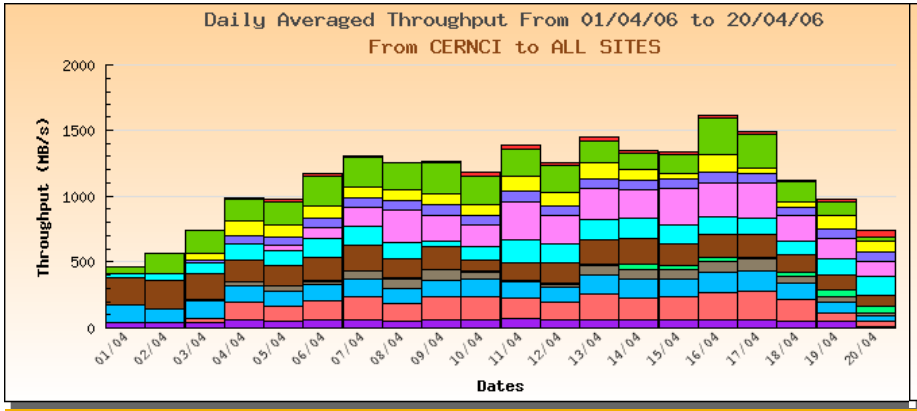
20k jobs running simultaneously

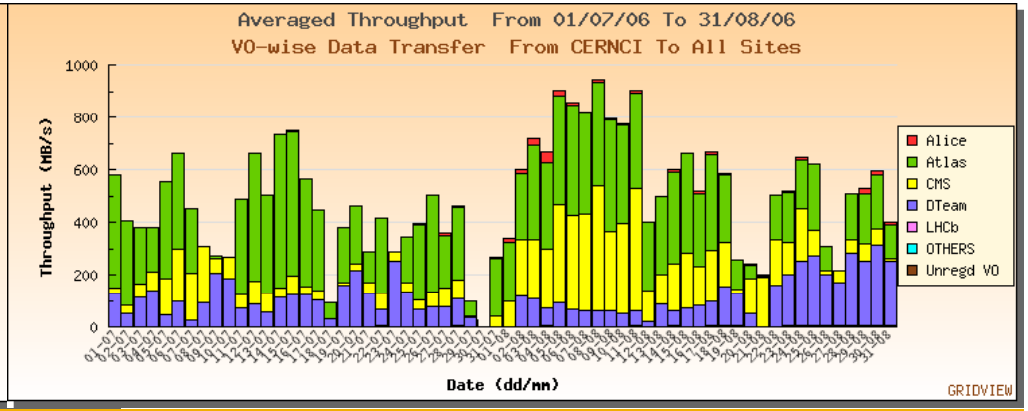


Use for massive data transfer

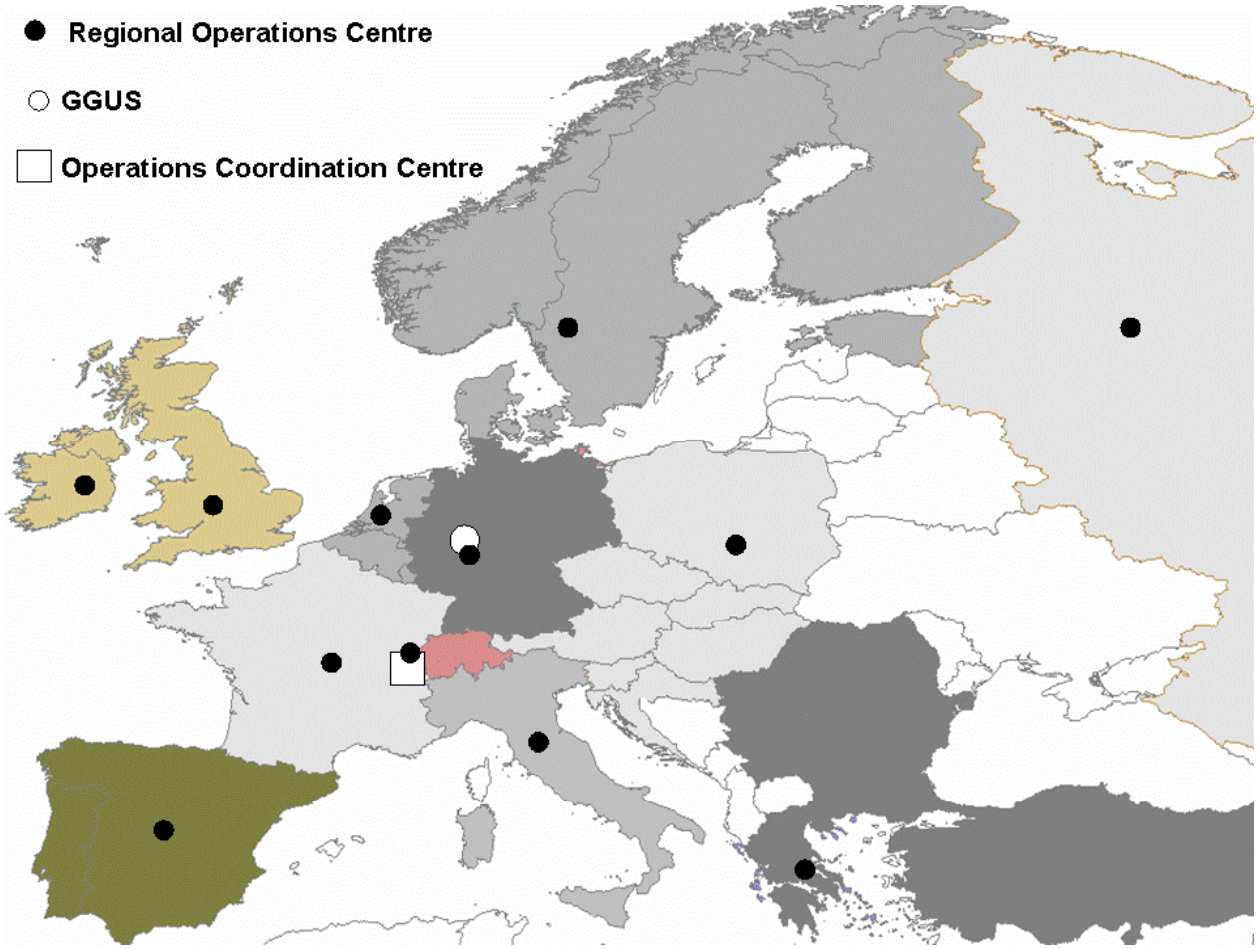


Large LHC experiments now transferring ~ 1PB/month each





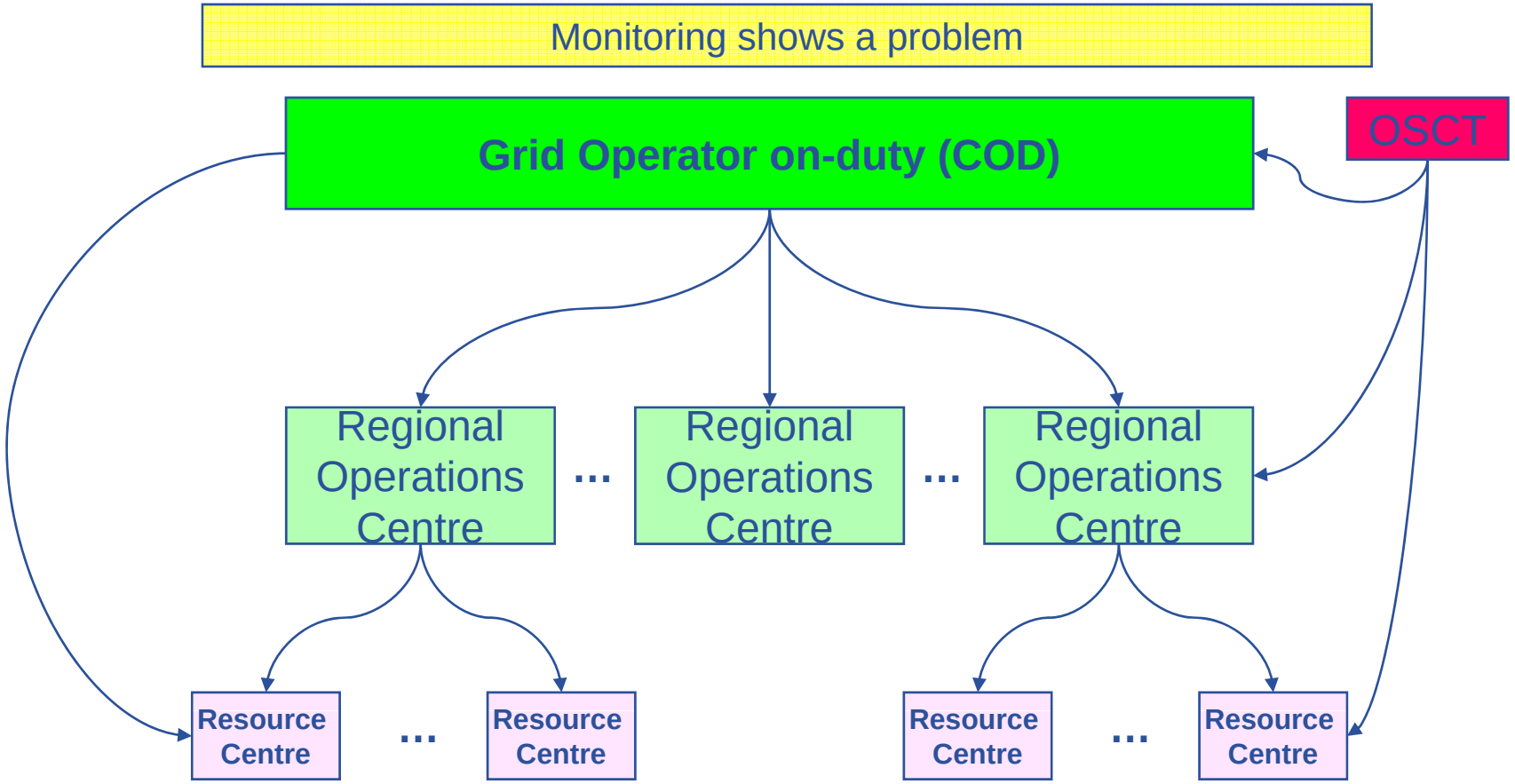
Grid management: structure



- **Operations Coordination Centre (OCC)**
  - management, oversight of all operational and support activities
- **Regional Operations Centres (ROC)**
  - providing the core of the support infrastructure, each supporting a number of resource centres within its region
  - **Grid Operator on Duty**
- **Resource centres**
  - providing resources (computing, storage, network, etc.);
- **Grid User Support (GGUS)**
  - At FZK, coordination and management of user support, single point of contact for users

Grid monitoring

The goal is to proactively monitor the operational state of the Grid and its performance, initiating corrective action to remedy problems arising with either core infrastructure or Grid resources



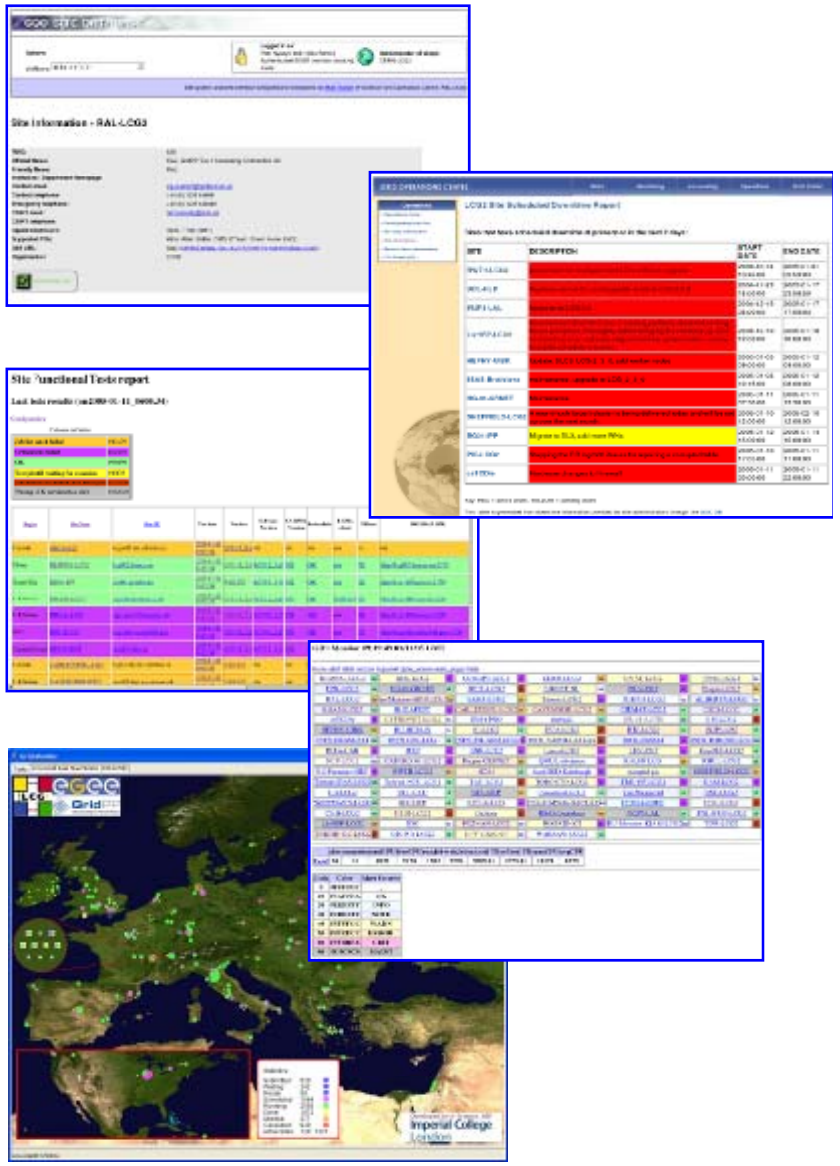
Grid Operator on Duty

- **Role:**
  - Watch the problems detected by the grid monitoring tools
  - Problem diagnosis
  - Report these problems (GGUS tickets)
  - Follow and escalate them if needed (well defined procedure)
  - Provide help, propose solutions
  - Build and maintain a central knowledge database (WIKI)

- **Who?**
  - 9 ROC teams working in pairs (one lead and one backup) on a weekly rotation
  - CERN, France, Italy, UK, Russia, Asia-Pacific, Southeastern-Europe, Central-Europe, Germany-Switzerland

Grid monitoring tools

- Tools used by the Grid Operator on Duty team to detect problems
- Distributed responsibility
- CIC portal
  - single entry point
  - Integrated view of monitoring tools
- Site Functional Tests (SFT) -> Service Availability Monitoring (SAM)
- Grid Operations Centre Core Database (GOCDB)
- GIIS monitor (Gstat)
- GOC certificate lifetime
- GOC job monitor
- Others



Site Functional Tests

- **Site Functional Tests (SFT)**
  - Framework to test (sample) services at all sites
  - Shows results matrix
  - Detailed test log available for troubleshooting and debugging
  - History of individual tests is kept
  - Can include VO-specific tests (e.g. sw environment)
  - Normally >80% of sites pass SFTs

- Very important in stabilising sites:
  - Apps use only good sites
  - Bad sites are automatically excluded
  - Sites work hard to fix problems

Test summary

	SD	JL	JS	CT	OK	total
dteam	15	12	4	6	139	176
lhcb	15	81	5	35	39	175

Colours definition

SD	Scheduled downtime	#a3a3a3
JL	Job list match failed	#aab3ff
JS	Job submission failed	#f4876b
CT	Critical tests failed	#f9d48e
NT	Non-critical tests failed	#f2f98e
OK	OK	#b2f98e

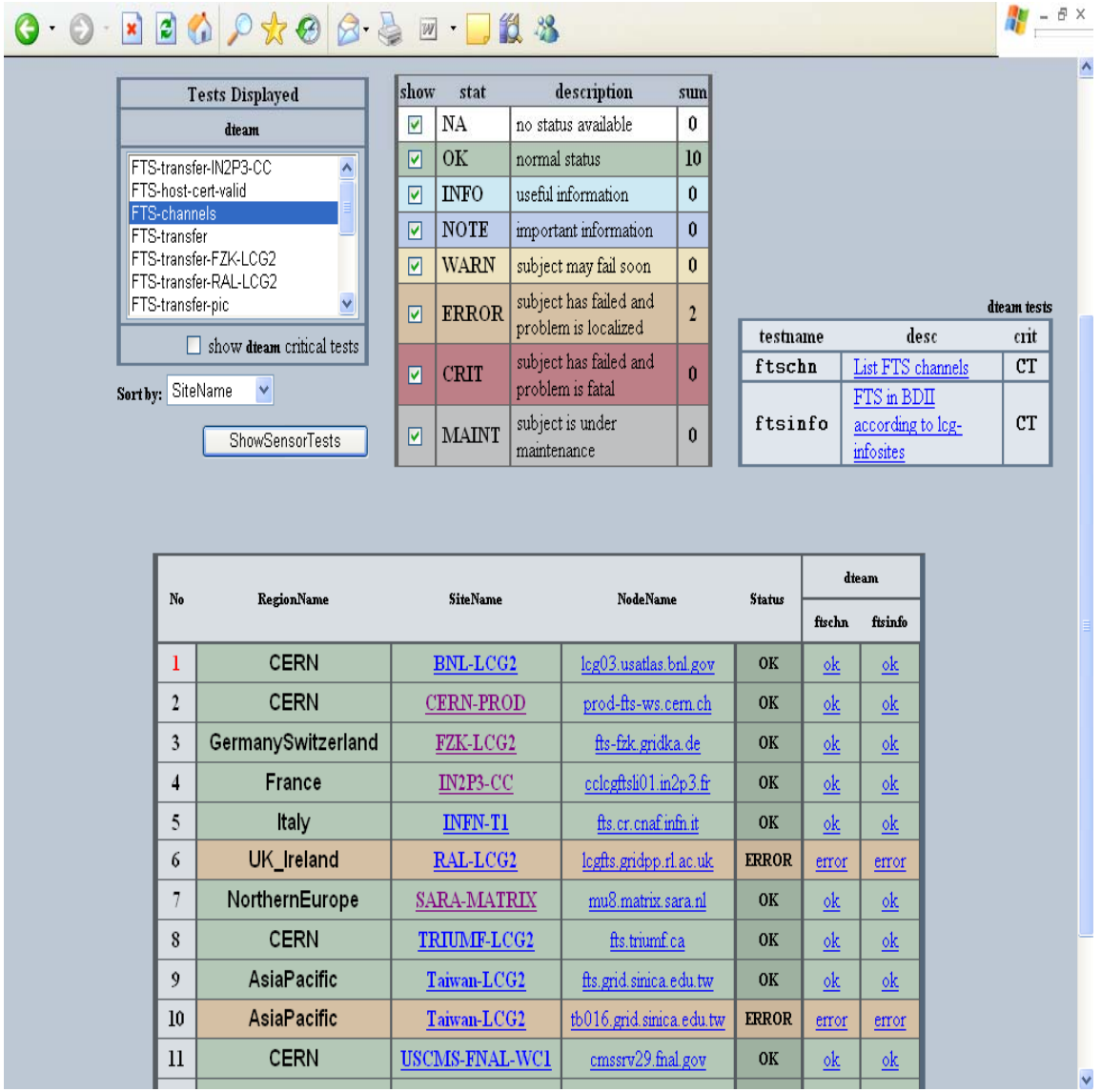
Test abbreviations

csh	CSH test
swdir	VO software directory
rgma	R-GMA
dirac-test	Dirac full test
ver	Software Version (WN)
wn	WN host name
ca	CA certs version
crl	CRL timestamp test
rm	Replica Management
votag	VO Tag management
js	Job submission
bi	BrokerInfo

	St.	Site Name	Site CE	VO dteam												VO lhcb		
				St.	js	ver	wn	ca	rgma	bi	csh	rm	votag	swdir	crl	St.	js	dirac-test
	AsiaPacific																	
1.	CT	INDIACMS-TIFR	ce.indiacms.res.in	CT	O	2 6 0	I	O	O	O	O	X	O	O	!!!	JL	X	??
2.	OK	TW-NCUHEP	grid01.phy.ncu.edu.tw	OK	O	2 6 0	I	O	O	O	O	O	O	O	!!!	JL	X	??
3.	OK	TOKYO-LCG2	dgce0.icepp.jp	OK	O	2 4 0	I	O	O	O	O	O	O	O	!!!	JL	X	??
4.	OK	Taiwan-LCG2	lcg00125.grid.sinica.edu.tw	OK	O	2 6 0	I	O	O	O	O	O	O	O	!!!	JL	X	??
5.	OK	Taiwan-IPAS-LCG2	testbed001.phys.sinica.edu.tw	OK	O	2 6 0	I	O	O	O	O	O	O	O	!!!	JL	X	??
6.	OK	GOG-Singapore	melon.ngpp.ngp.org.sg	OK	O	2 6 0	I	O	O	O	O	O	O	O	!!!	JL	X	??
7.	OK	Taiwan-NCUCC-LCG2	ce.cc.ncu.edu.tw	OK	O	2 6 0	I	O	O	O	O	O	O	O	!!!	OK	O	O
8.	OK	LCG_KNU	cluster50.knu.ac.kr	OK	O	2 5 0	I	O	O	O	O	O	O	O	!!!	CT	O	!!!
	BNL																	
9.	SD	BNL-LCG2	lcg-ce01.usatlas.bnl.gov	SD	X	??	??	? ?	??	? ?	??	??	??	??	??	SD	X	??
	Canada																	
10.	JL	TORONTO-LCG2	bigmac-lcg-ce.physics.utoronto.ca	JL	X	2 6 0	I	O	O	O	O	O	W	O	!!!	OK	O	O
11.	SD	CARLETONU-LCG2	lcg02.physics.carleton.ca	SD	X	??	??	? ?	??	? ?	??	??	??	??	??	SD	X	??
12.	OK	TRIUMF-LCG2	lcgce01.triumf.ca	OK	O	2 6 0	I	O	O	O	O	O	O	O	O	OK	O	O
13.	OK	Umontreal-LCG2	lcg-ce.lps.umontreal.ca	OK	O	2 6 0	I	O	O	O	O	O	W	O	!!!	OK	O	O

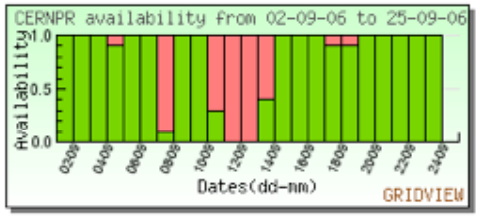
Service Availability Monitoring

- **Service Availability Monitoring (SAM)**
  - Monitoring of all grid services
  - web service based access to data
  - availability metric calculation
  - Will be used to generate alarms
    - to generate trouble tickets
    - to call out support staff



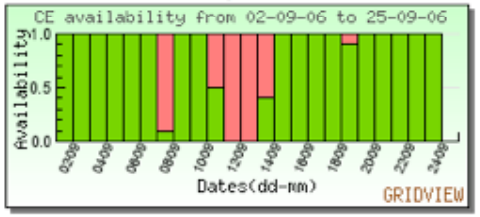
Availability metrics - GridView

Overall Service Availablity for site CERNPR : Daily Report

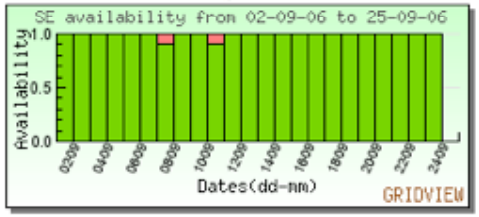


Individual Service Availablity for site CERNPR : Daily Report

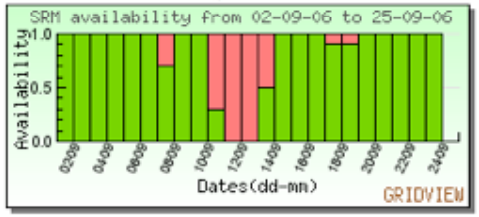
CE :



SE :



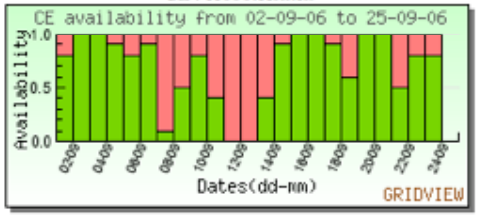
SRM :



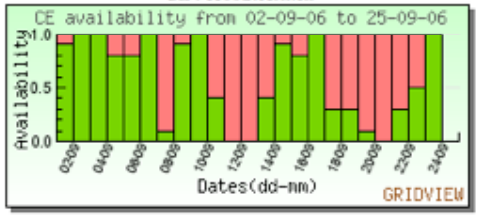
Service Instance Availablity for site CERNPR : Daily Report

Site Services

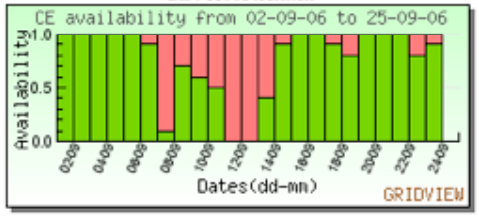
CE : ce101.cern.ch



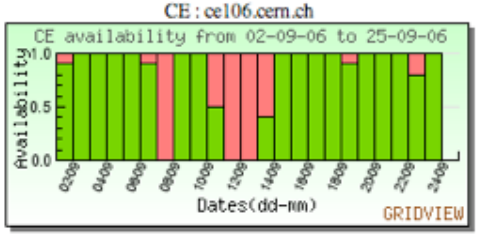
CE : ce102.cern.ch



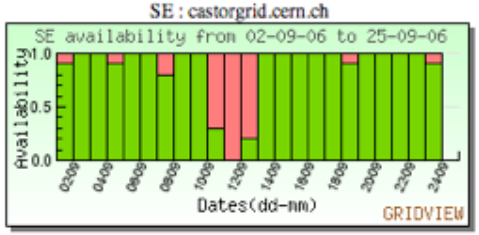
CE : ce105.cern.ch



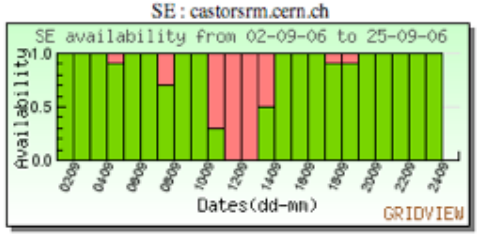
CE : ce106.cern.ch



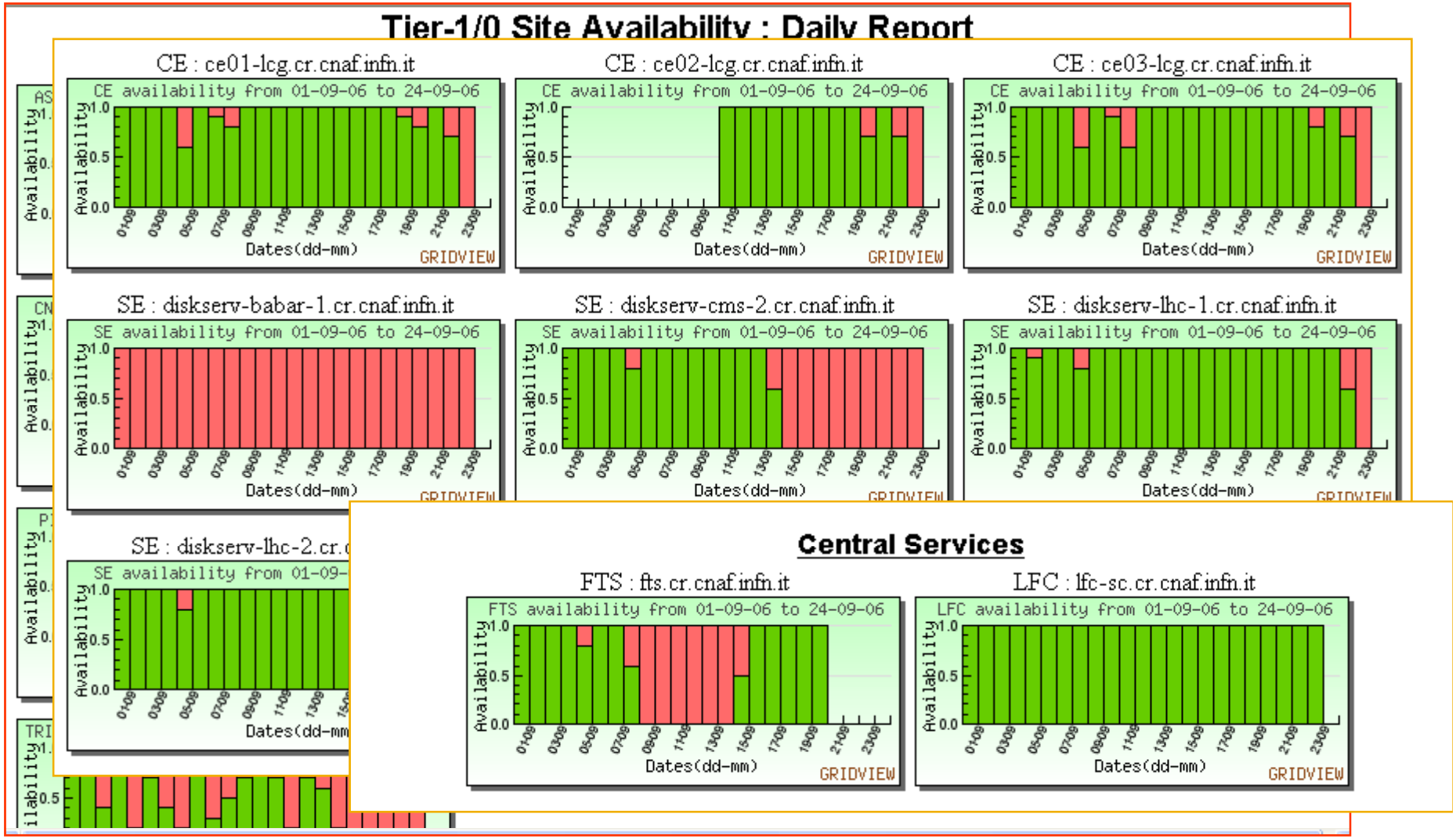
SE : castorgrid.cern.ch



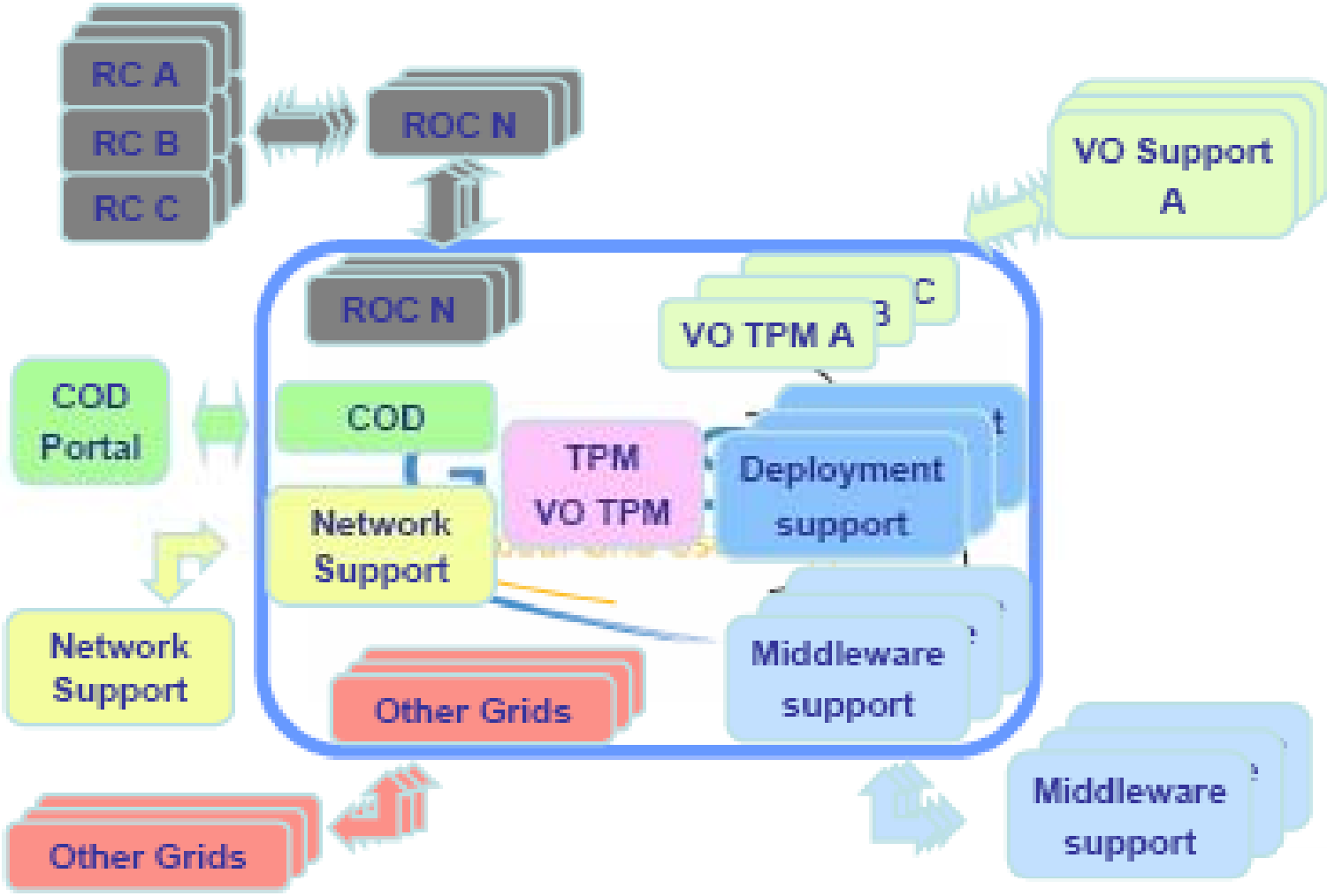
SE : castorsrm.cern.ch



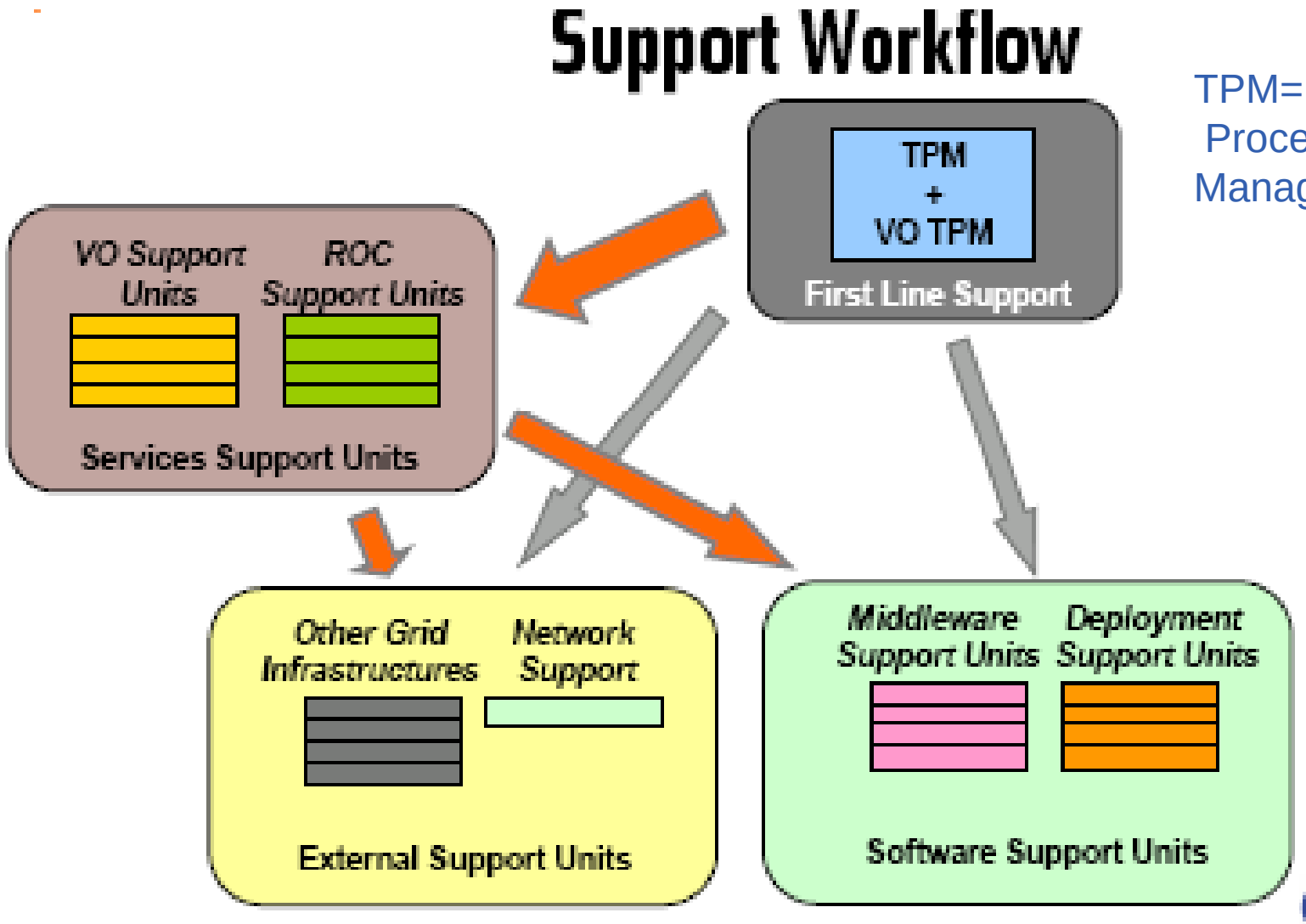
Site metrics - availability



Support - GGUS

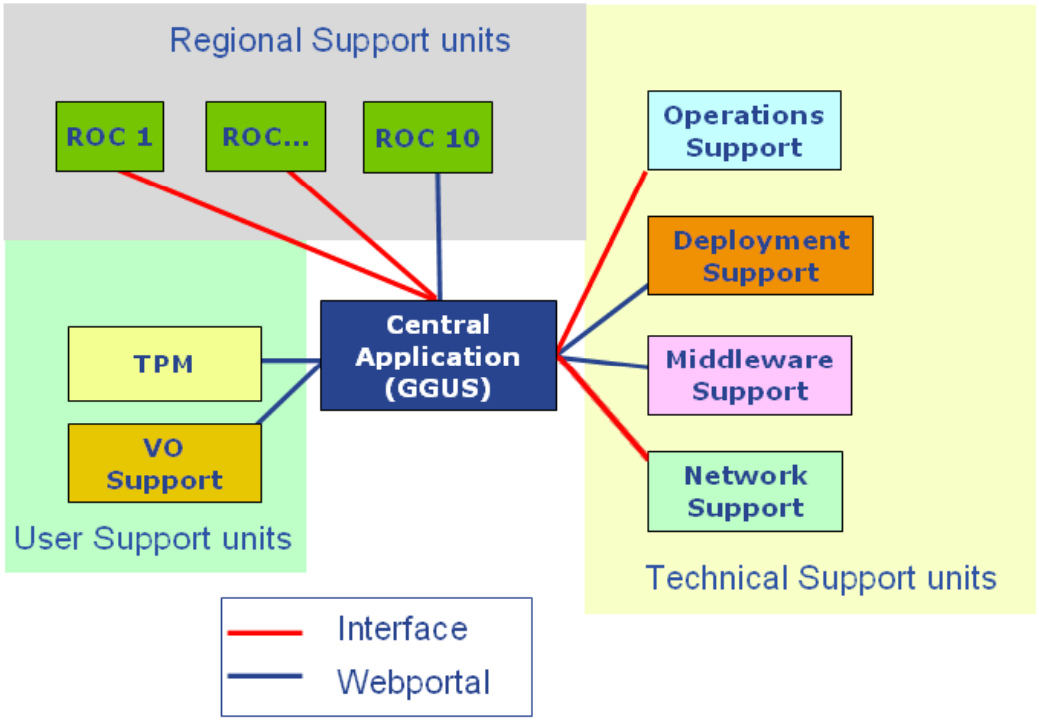


GGUS - Global Grid User Support



TPM= Ticket Processing Manager

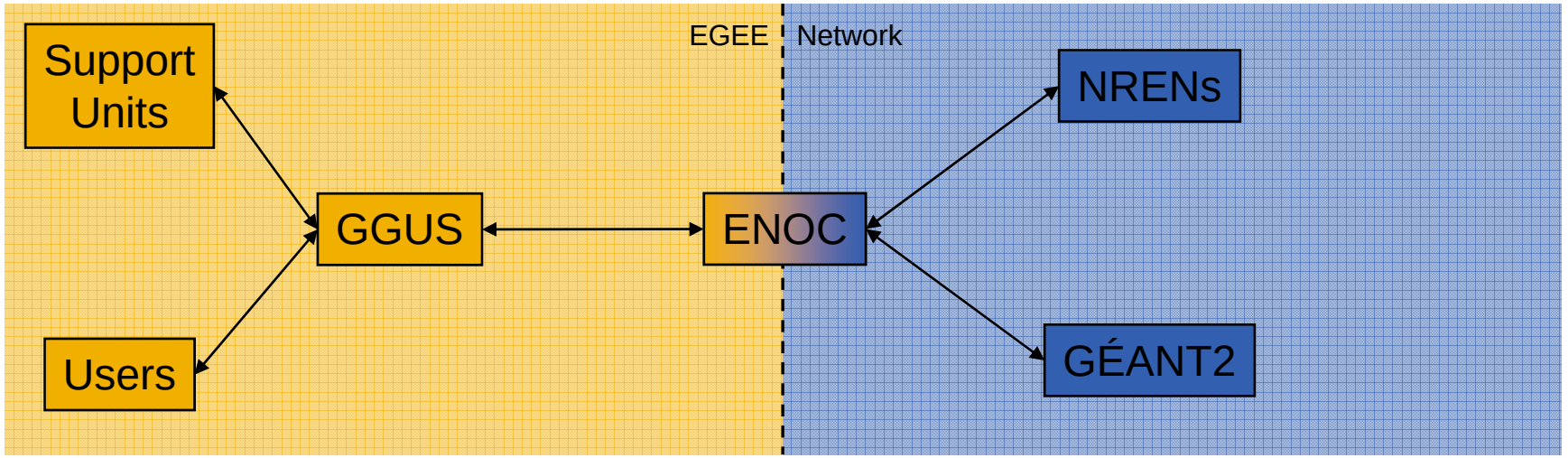
EGEE Network Operations Centre (ENOC)



- **Creating a "Network Support unit" in the EGEE operational model;**
- **Based on the work done during EGEE:**
- **First implementation in EGEE-II:**
  - First "iteration";
  - Planned developments in the next months.

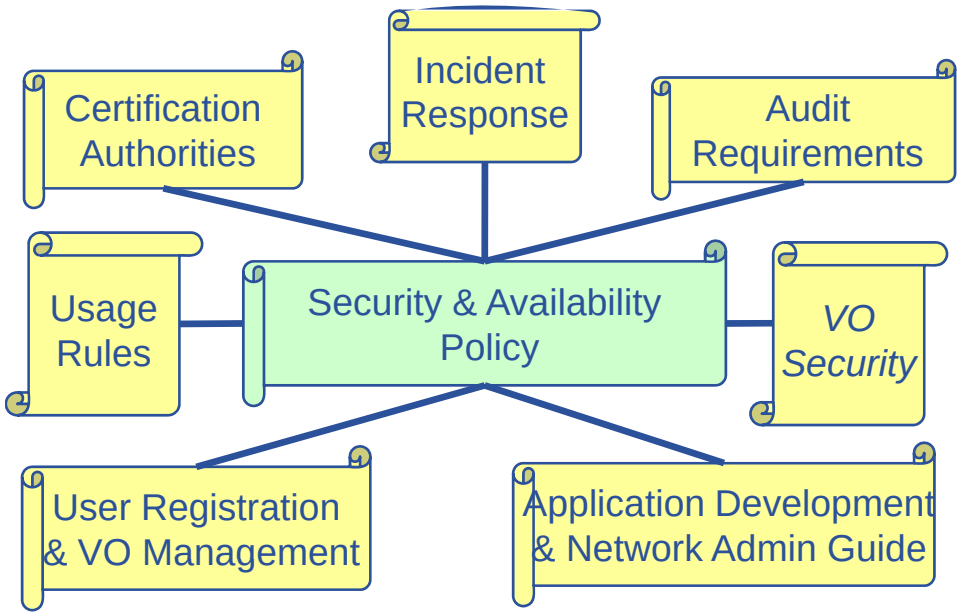
- **Tasks:**
  - Receive tickets from NRENs, and forward to GGUS if impact on grid
  - Receive tickets from GGUS if TPM determines a network issue
  - Troubleshoot them provided that the ENOC has access to suitable monitoring tools;
  - Contact identified faulty domains or reassign ticket to the associated site if there is no evidence of a backbone problem (e.g. LAN issue).

Current ENOC status



- **Interface with NRENs is running like in EGEE:**
  - ENOC receives Trouble Tickets (incident, maintenance) from GÉANT and the NRENs (currently France, Germany, Greece, Hungary, Ireland, Italy, Russia, Spain, Switzerland, and United Kingdom);
    - More to come: Poland, the Netherlands, Czech Republic;
  - Forward it to GGUS after analysis and if relevant to EGEE.
- **Identified as the Network Support unit in GGUS:**
  - 2nd level support for network related issues
- **Identified as the point of contact for EGEE by the NRENs and GEANT2**

Security & Policy



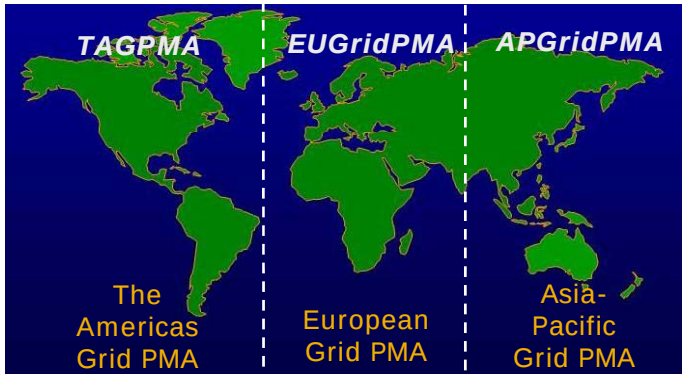
Collaborative policy development

- Many policy aspects are collaborative works; e.g.:

- **Joint Security Policy Group**
- **Certification Authorities**
  - EUGridPMA → IGTF, etc.
- **Grid Acceptable Use Policy (AUP)**
  - common, general and simple AUP
  - for all VO members using many Grid infrastructures
    - EGEE, OSG, SEE-GRID, DEISA, national Grids…
- **Incident Handling and Response**
  - defines basic communications paths
  - defines requirements (MUSTs) for Incident Response
  - not to replace or interfere with local response plans

Security groups

- **Joint Security Policy Group:**
  - Joint with WLCG, OSG, and others
  - Focus on policy issues
  - Strong input to e-IRG
- **EUGridPMA**
  - Pan-European trust federation of CAs
  - Included in IGTF (and was model for it)
  - Success: most grid projects now subscribe to the IGTF
- **Grid Security Vulnerability Group**
  - New group in EGEE-II
  - Looking at how to manage vulnerabilities
  - Risk analysis is fundamental
  - Hard to balance between openness and giving away insider info
- **Operational Security Coordination Team**
  - Main day-to-day operational security work
  - Incident response and follow up
  - Members in all ROCs and sites
  - Recent security incident (**not** grid-related) was good shakedown



Operational Advisory Group

- **Role**
  - Negotiate access to resources for applications and VOs
  - Manage procedures:
    - To recognize new VOs & define MoUs
  - Identify and manage major procedural problems between VOs and Operations
- **Membership**
  - Co-chaired by SA1 and NA4
  - Members: VO Managers, ROC managers
- **Status**
  - New simpler VO registration procedure in place
  - MoU with DILIGENT in progress
  - Tools to show high level resource allocation by region and VO are planned
- **Issues**
  - Resource negotiation procedures have to be developed
    - This has to be done by region
    - Resource allocation summary tools are a pre-requisite
    - Escalation procedures in case of unsatisfied requests have to be found
  - The operation of the OAG itself has to be changed
    - No EGAAP any longer
    - User Forum and EGEE Conference now more important for face-to-face meetings

Interoperation

- **Interoperability and interoperation (or co-operation)**

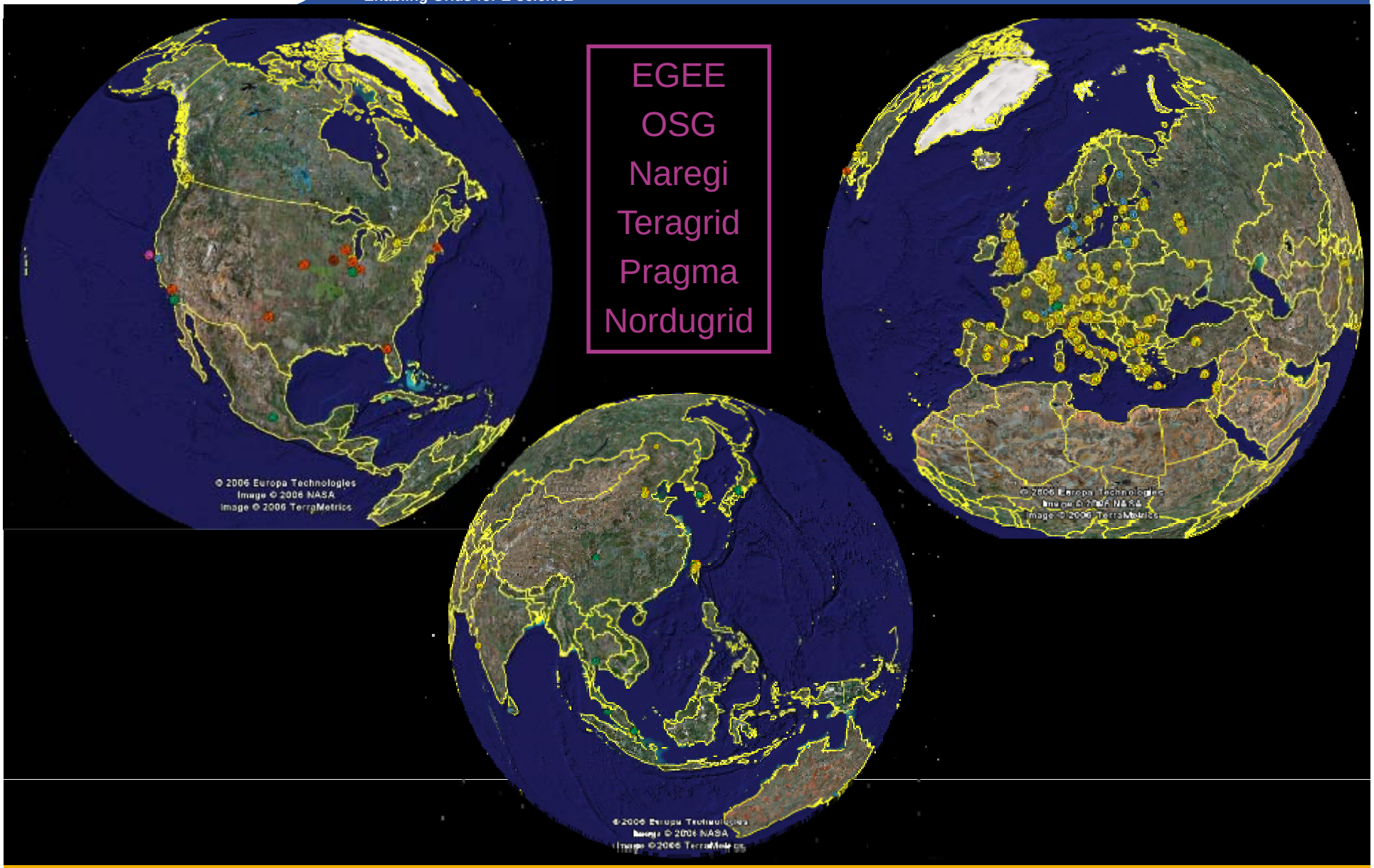
- **EGEE has interoperability activities with:**
  (enabling the middlewares to work together)

  - Open Science Grid (U.S.) – quite far advanced
  - Nordugrid (ARC) – task in EGEE-II, 4 workshops and ongoing activity
  - UNICORE – task in EGEE-II
  - NAREGI (Japan) – 1 workshop, continued activity
  - GIN (OGF) – active in several areas

- **EGEE has interoperation activities with:**
  (enabling the infrastructures to co-operate)

  - Open Science Grid – actually in use
  - Anticipated with NorduGrid (NDGF) for WLCG

Interoperating information systems

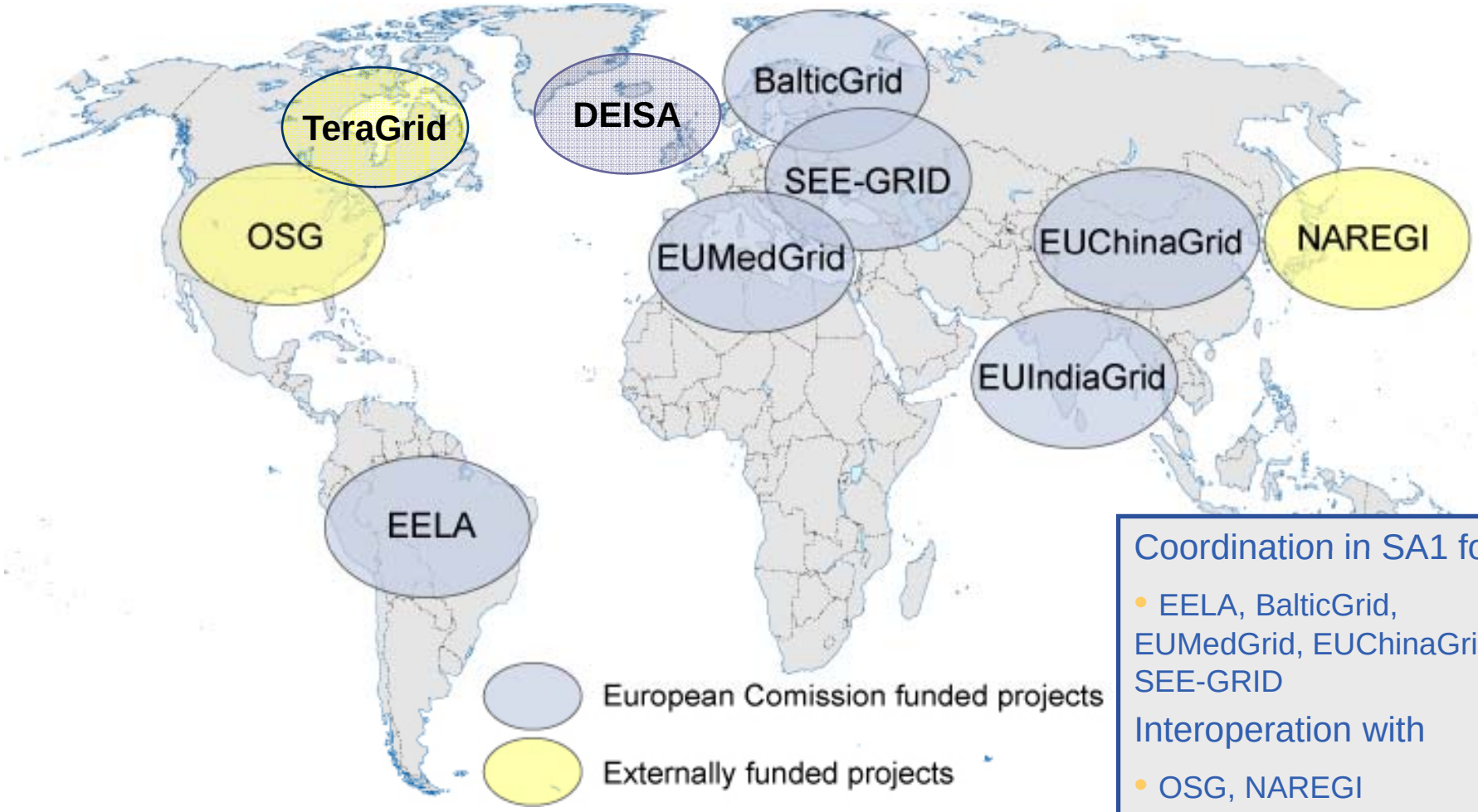


- **To obtain a Google map of the Grids in the Globus Interoperability Now (GIN) initiative go to:**

http://www.pparc.ac.uk/Nw/GIN.asp

- **(You will need to install GoogleEarth)**

Related infrastructure projects



Coordination in SA1 for:

- EELA, BalticGrid, EUMedGrid, EUChinaGrid, SEE-GRID

Interoperation with

- OSG, NAREGI

SA3:

- DEISA, ARC, NAREGI

Summary of status

- **Today we have an operating production infrastructure**
  - Probably the largest in the world, supporting many science domains
  - Relied upon by several as their primary source of computing
- **We have a managed operations process addressing most areas**
  - Constantly evolving
- **Inter/Co-operation is a fact and is becoming more important very quickly**
  - Several applications need to work across grids – and they need support for that
- **A large fraction of the value of the operations activity is in the intangibles – processes, structures, expertise, etc.**
- **We recognise that there are many outstanding problems with the current state of things**

Some (personal) observations

- **Production grids turned out to be a lot harder than anticipated**
  - Often not production quality software, not designed with services and service management in mind – really should regard as advanced prototypes
  - Rediscovered the wheel : takes a long time to go from prototypes to production quality
  - We have done a lot : but it is hard to use, hard to support, and hard to manage …
- **Complexity**
  - We have a lot of complexity – often a clue that something is not right
  - Perhaps it is necessary … ? But we should be careful.
- **Many of the reliability issues are not grid specific**
  - Site management problems are reflected in the overall service
- **… and any major changes we make have to be implemented in such a way that does not break the production service**

Short term issues

- **Urgently need to support more platforms**
  - Migration to GT4 (pre-WS)
  - Simplify effort involved in porting
- **Deployment of new services**
  - Especially secure data management
  - Better support for MPI
  - …
- **Site reliability, stability – Understanding of the system in detail**
  - Not more monitoring tools
  - We need more sensors (do we monitor enough or the right things?)
  - We need data interpretation (knowledge not information)
    - And use that to generate actions – alarms, self-repairing systems, …
  - We have to manage a dynamically changing distributed system with many levels of reliability, management, stability
    - And hide this from the applications !
- **How do we scale up to 10-50 times the workload we have now?**

Long term outlook

- **Reliability & management must be built in not added on**
  - We have been through 1 round of “re-engineering” – now we need the “real thing”
- **We need a real architecture (avoid “VO Boxes”)**
  - How do we deploy application level services in an acceptable way?
- **Information systems:**
  - Crucial to a grid infrastructure, more and more information needs to be published
  - See the limits of current system
  - Our experience and knowledge is encapsulated in GLUE, not the implementation
- **Security model**
  - Proxy renewals ? Is this what we want?
  - Complexity again
  - Where does Shibboleth fit?
  - Is there a better way?
- **Dynamic VOs (the original idea of what a VO could be)**
  - How to achieve this? It seems a long way off
  - cf ITU exercise where we hi-jacked an existing VO

Standardisation

- **Interoperation, interoperability is the best way to drive real standards**
  - But must take care not to constrain ourselves – things are still changing rapidly
- **See already practical work on information systems, data, job submission, etc**
  - Both in EGEE and related projects and in the GGF/OGF GIN work
- **EGEE Operations have a wealth of experience now**
  - Procedures, issues, what works, what does not work
  - The problem is finding the time to publish this knowledge
  - We need to start documenting this now – especially at the ROC level
- **The value of EGEE is in the infrastructure – not just the production service, but all the parts that fit around that**
  - This is more or less independent of any specific set of middleware (although that makes life more or less easy …)

Conclusions

- **We have come a long way in the last 5-6 years**
  - From specific solutions for HEP to a vision of a global infrastructure
  - But we need to be careful to walk before we run … and clarify expectations
- **Many complex issues need to be addressed**
- **We should expect to see major changes in implementations**
  - But the infrastructure should remain and evolve across these changes
- **Standardisation will come with co-operating grids**
- **Many opportunities for collaboration**
  - With other projects, with industry, with applications