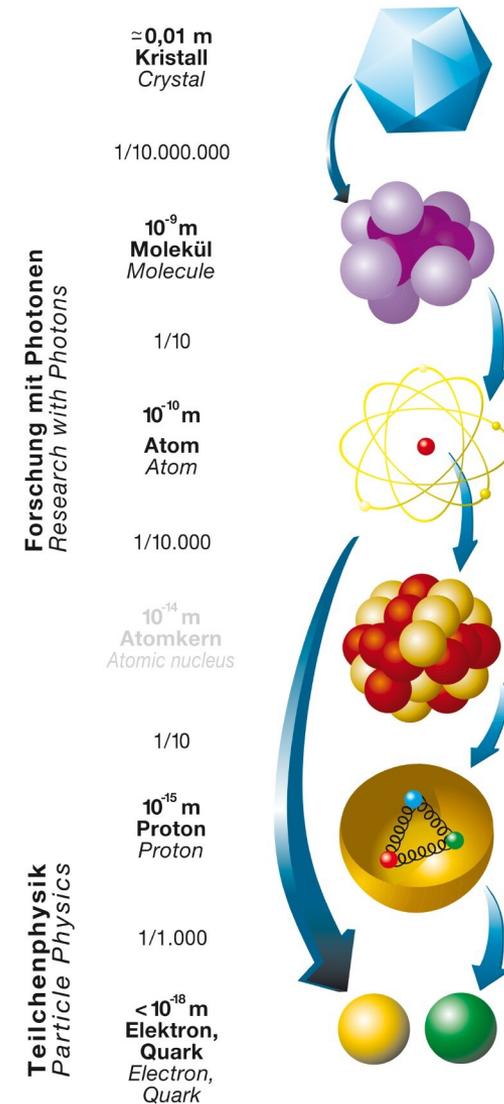


# Operational Experience - Viewpoint from Tier-2s

Andreas Gellrich for the DESY Grid Team  
Viewpoint from Tier-2s  
WLCG Collaboration Workshop  
11-13 July 2011 DESY, Hamburg

# Outline

- > Introduction
- > DESY Grid Center
  - Grid
  - NAF
- > Statistics
- > Experiences (→ [thnx to NGI\\_DE Tier-2s](#))
  - Paradigms
  - Operations
  - Services & Middleware
  - Support
  - Issues & Concerns
- > Conclusions



# Introduction: DESY

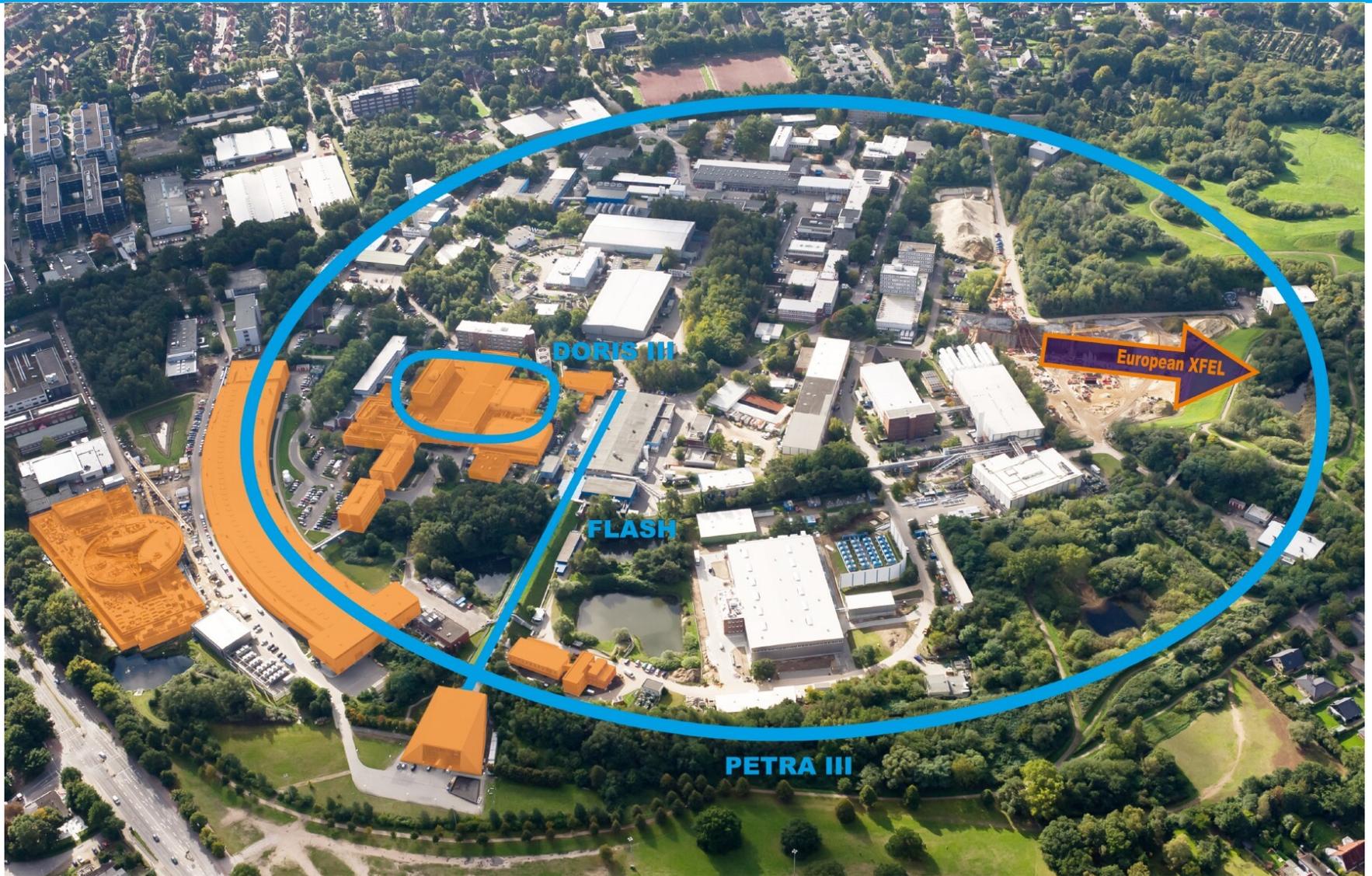
- > DESY is one of the world's leading **accelerator centres** and a member of the **Helmholtz Association**. DESY develops, builds and operates large particle accelerators used to investigate the structure of matter. DESY offers a broad research spectrum of international standing focusing on three main areas: **accelerator** development, construction and operation; **photon** science; **particle** and **astroparticle** physics.
- > Thanks to its expertise and worldwide unique diversity of excellent light sources, DESY is a very attractive venue for more than 3000 scientists from over 40 countries a year, and a sought-after partner in national and international cooperations and projects. The DESY research programme is not restricted to the facilities at its **two locations** in **Hamburg** and **Zeuthen**. DESY is closely involved in a number of major **international projects**, including the X-ray laser European **XFEL** in Hamburg and Schleswig-Holstein, the Large Hadron Collider **LHC** in Geneva, the neutrino telescope **IceCube** at the South Pole and the International Linear Collider **ILC**.
- > DESY was founded **1959** and celebrated its 50th birthday in 2009



# Introduction: DESY 1965



# Introduction: DESY 2011



# DESY Grid Center: History

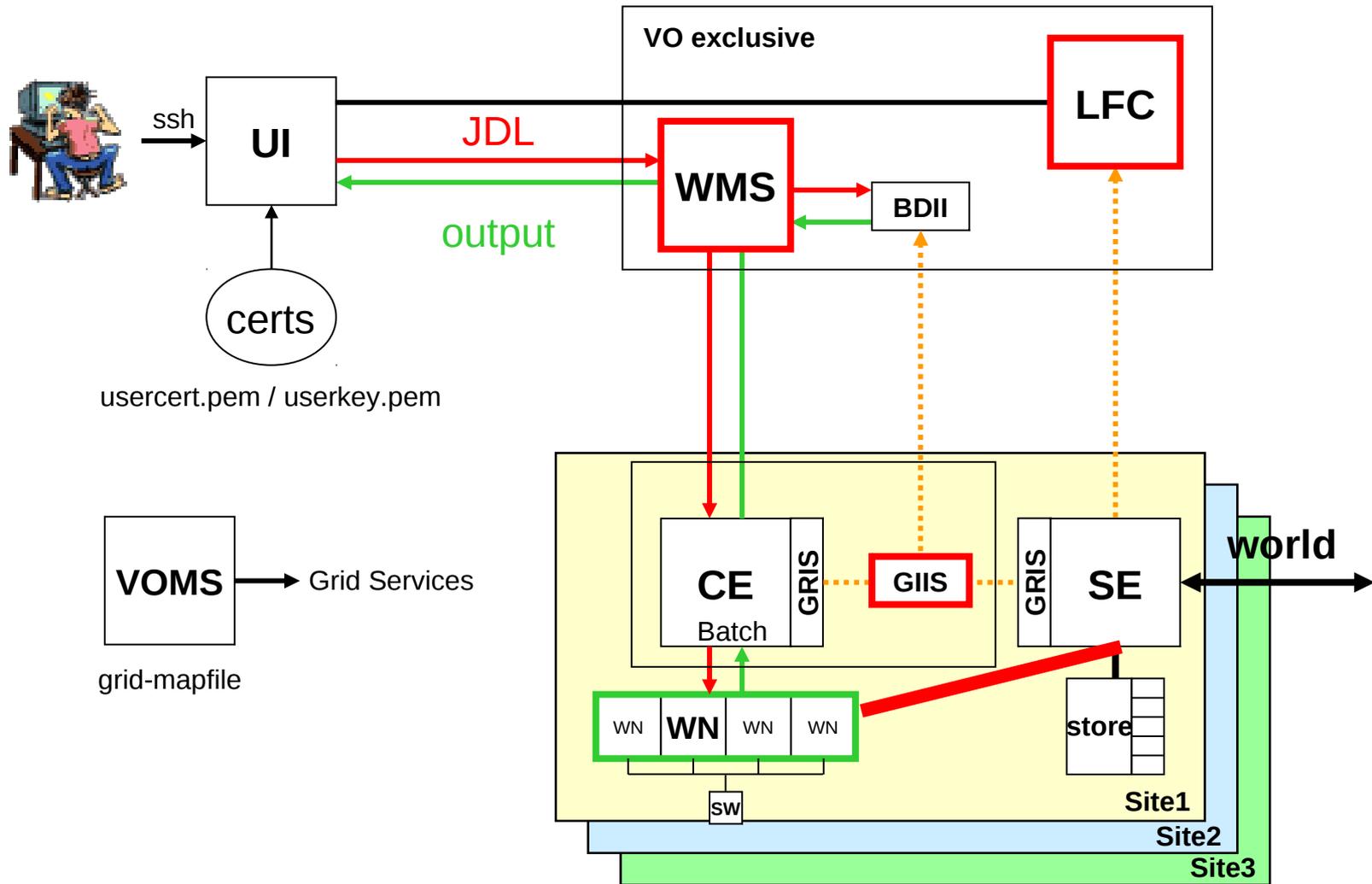
- > DESY has a long tradition in HEP computing (serving experiments on site)
- > 2003: First look into Grid computing
- > 2003: Future key technology to access resources
- > 2004: LCG\_2-1 Grid infrastructure H1 and ZEUS, IceCube, ILC, ILDG
- > 2004: EGEE(2/3) / DECH
- > 2004: Tier-2 for ATLAS, CMS, LHCb
- > 2005: Foundation of national D-Grid (DGI(2) and HEPCG)
- > 2007: National Analysis Facility (NAF) (ATLAS, CMS, LHCb, ILC)
- > 2010: Start of EGI / NGI\_DE and EMI (dCache)



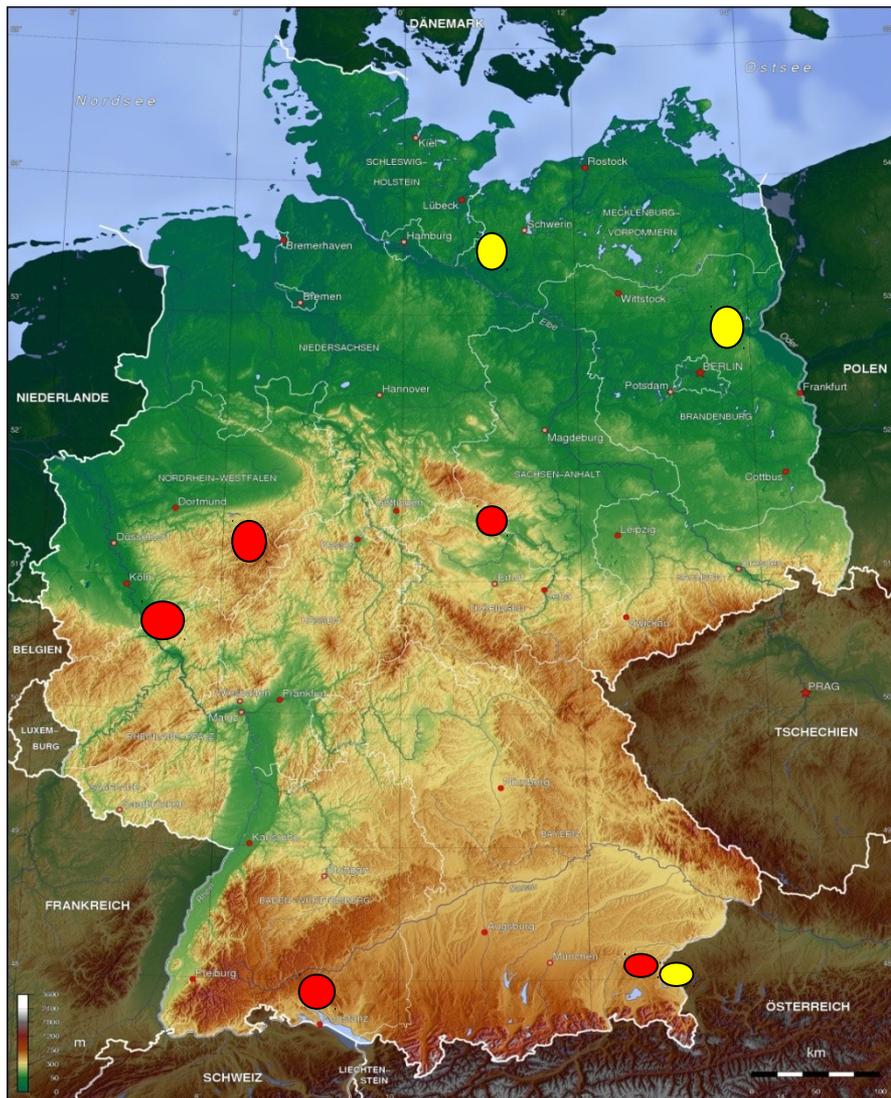
- > The Grid infrastructure reflects DESY's manifold scientific programme
- > DESY is the home of 10 VOs (6 global), incl. non-HEP
- > **Tier-2** for ATLAS, CMS, and LHCb in Germany (Tier-1: GridKa)
- > **Tier-0/1** for HERA; ILC VOs incl. testbeams (DESY, FNAL, CERN); CTA
- > Tape back-end for non-Tier-2
  
- > One *complete generic* Grid infrastructure for *all* VOs
- > *Federated* resources w/ *opportunistic* usage
  
- > Currently, roughly 2/3 of the resources are assigned to the Tier-2 VOs
- > Grid is *complemented* by the National Analysis Facility (**NAF**) [size: ~1 Tier-2]



# DESY Grid Center: Grid Infrastructure



# DESY Grid Center: NGI\_DE Tier-2s



## Pledges for Germany

[http://lcg.web.cern.ch/LCG/Resources/WLCGResources-2010-2012\\_15DEC2010.pdf](http://lcg.web.cern.ch/LCG/Resources/WLCGResources-2010-2012_15DEC2010.pdf)

	CPU 2011 [HS]	CPU 2012 [HS]	Disk 2011 [TB]	Disk 2012 [TB]
<b>A DESY</b>	6200	6600	1050	1350
<b>C DESY</b>	11800	12900	640	900
<b>A Goettingen</b>	3800	4000	400	590
<b>C Aachen</b>	6600	8700	330	435
<b>A Munich</b>	9220	11560	1040	1340
<b>A FR/Wupp</b>	4633	4917	633	733
<b>A FR/W Freiburg</b>	4610	5780	518	668
<b>Summe CMS</b>	18400	21600	970	1335
<b>Summe Atlas</b>	28463	32857	3641	4701
<b>SumDESY</b>	18000	19500	1690	2250
<b>SumNonDESY</b>	28863	34957	2921	3786
<b>GrandTotAllT2 WW</b>	725324	776203	60454	71998



# DESY Grid Center: Grid Infrastructure

## > Two DESY Grid sites:

- **DESY-HH** ldap://grid-giis.desy.de:2170/mds-vo-name=DESY-HH,o=grid
- **DESY-ZN** ldap://lcg-giis.ifh.de:2170/mds-vo-name=DESY-ZN,o=grid

## > VOs at DESY:

- **DESY:** 'calice' & 'ilc' 'hermes' & 'hone' & 'zeus' 'cta' & 'icecube' 'ildg' 'xfel.eu'
- **Tier-2:** 'atlas' & 'cms' & 'lhcb'
- **non-HEP:** 'biomed' & 'enmr.eu'
- **Operations:** 'dech' 'desy' 'dgops' 'dteam' 'ghep' 'ops'

## > Grid Core Services:

- lcg-CE, CREAM-CE, VOMRS/VOMS, top-level BDII, 6 WMS, LB, LFC, PX
- Load balancing via DNS aliases for BDII, WMS, CREAM-CE
- VODIRs on NFS

## > Network:

- **LHCone:** dedicated connection between (big) sites



# DESY Grid Center: Grid Infrastructure (cont'd)

## > Computing Resources: (CE) [SL5 / 64-bit]

- `grid-ce4/5/-cr5.desy.de` 4232 slots @ 372 hosts
- `lcg-ce0/-cream.ifh.de` 712 slots @ 89 hosts
- **Total:** 33.9 + 8.6 kHS06

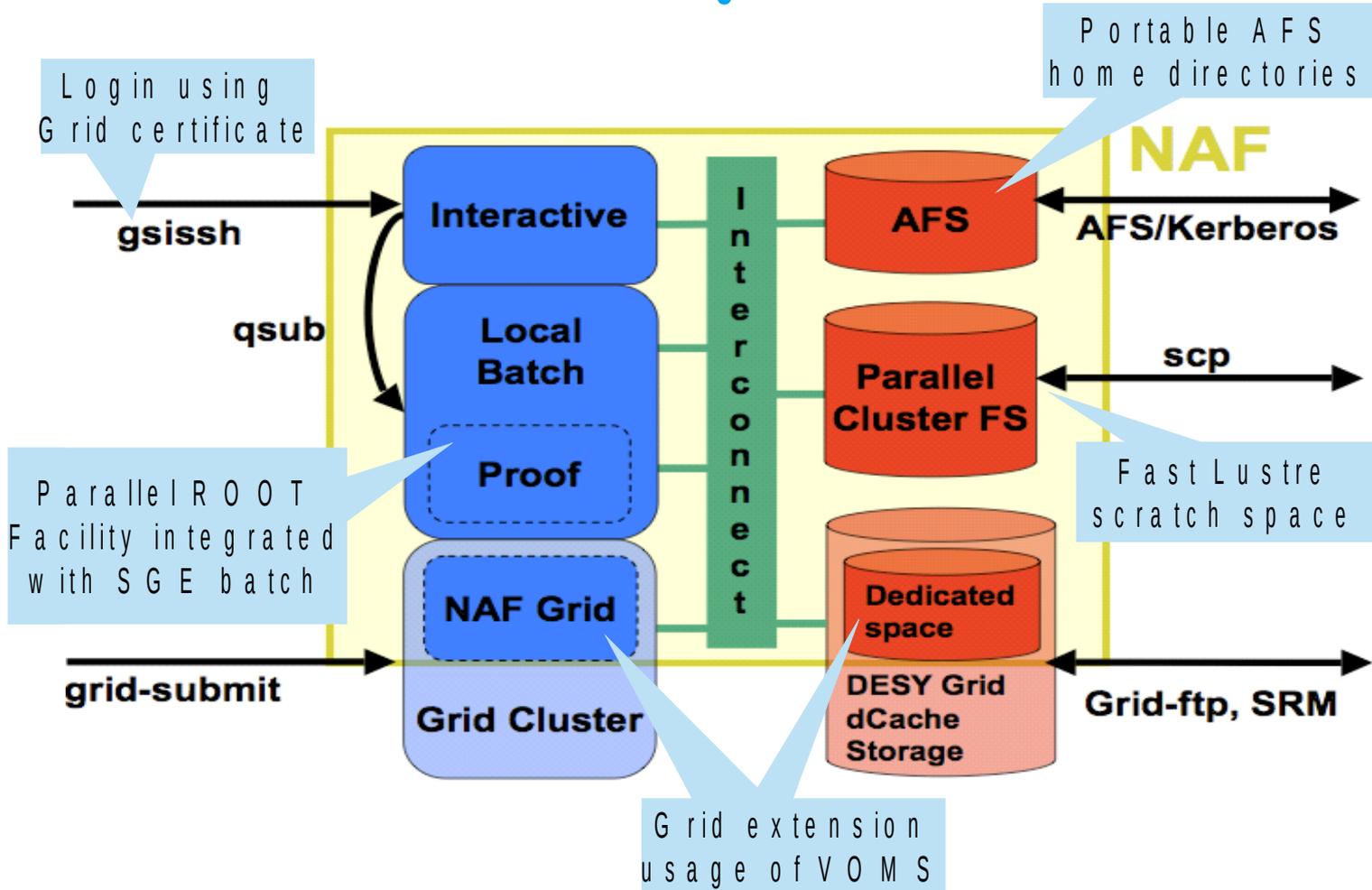
## > Storage Resources: (SE) (dCache)

- `dcache-se-atlas.desy.de` ATLAS: 800 TB
- `dcache-se-cms.desy.de` CMS: 1000 TB
- `dcache-se-desy.desy.de` others: 350 TB
- `lcg-se0.ifh.de` ATLAS: 550 TB
- `lcg-se1.ifh.de` LHCb: 180 TB
- `globe-door.ifh.de` others: 500 TB
- **Total:** 3380 TB



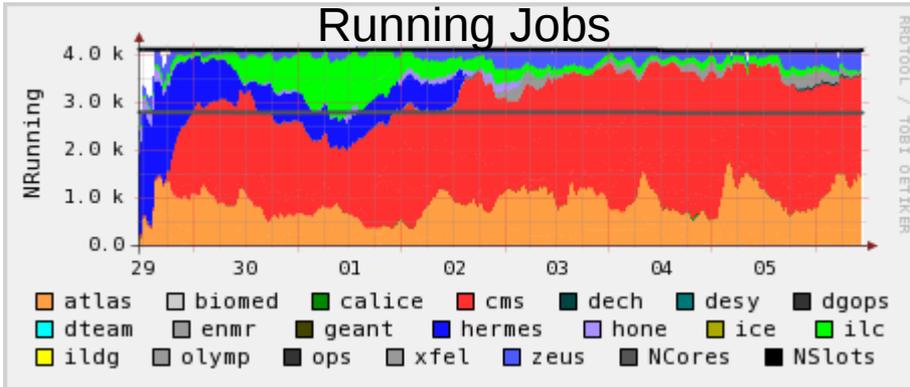
# DESY Grid Center: National Analysis Facility (NAF)

## The NAF Building Blocks

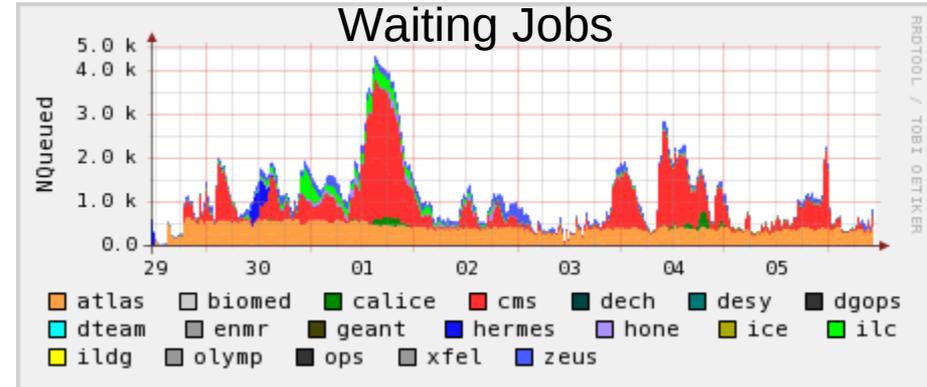


# Statistics: Jobs at DESY-HH (weekly)

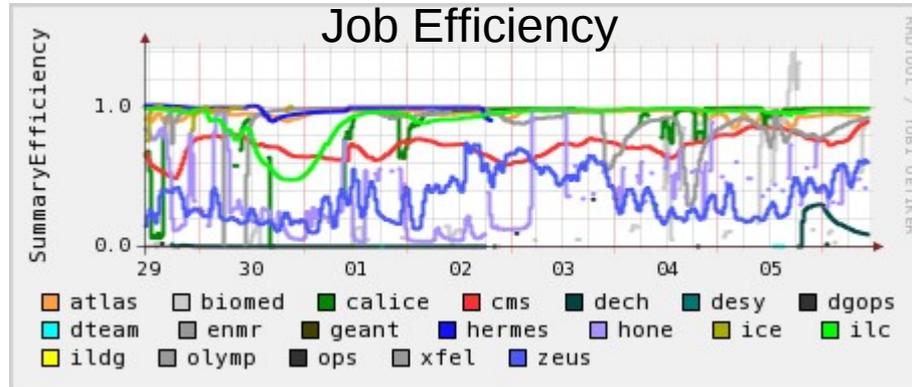
## Running Jobs



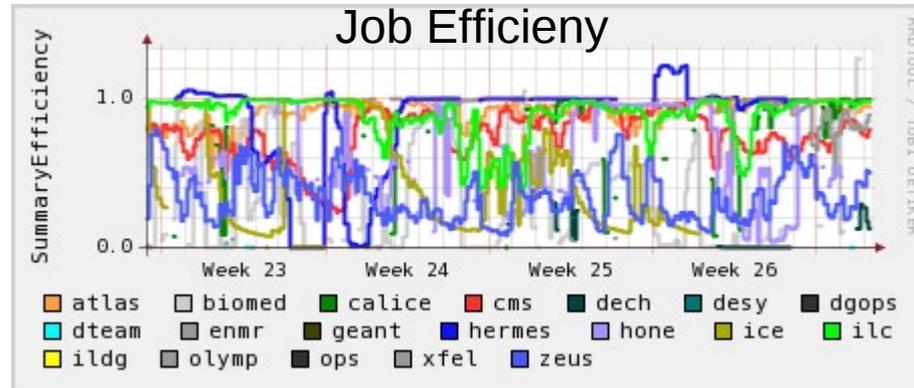
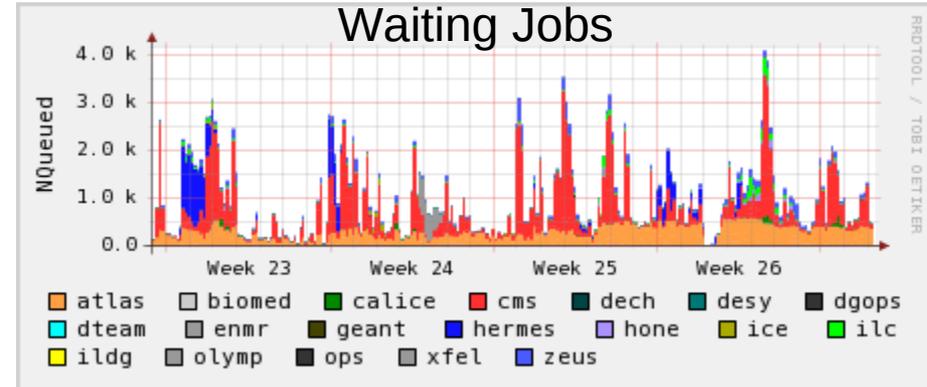
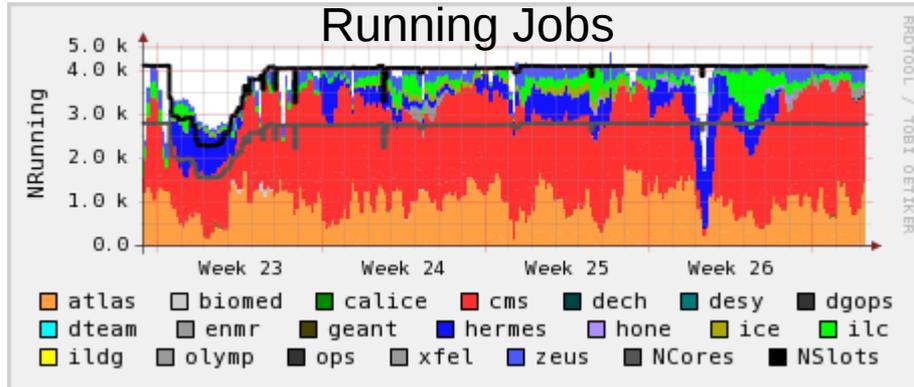
## Waiting Jobs



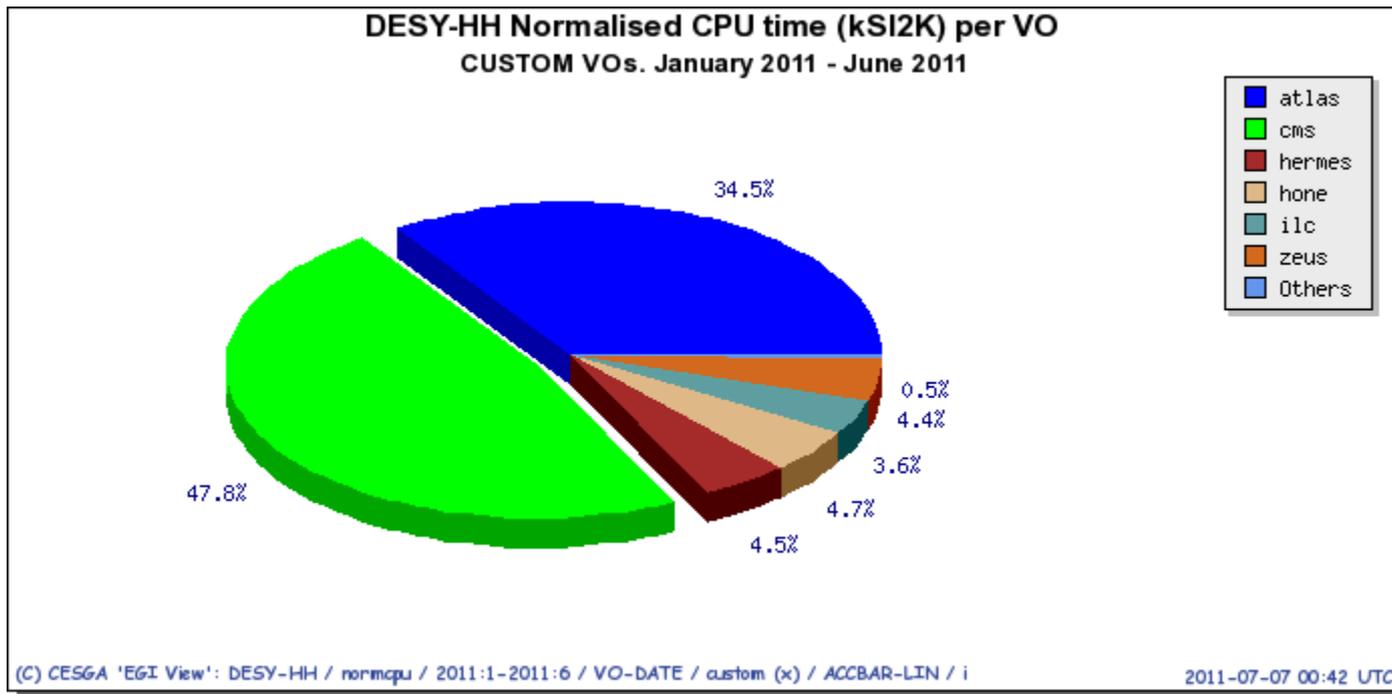
## Job Efficiency



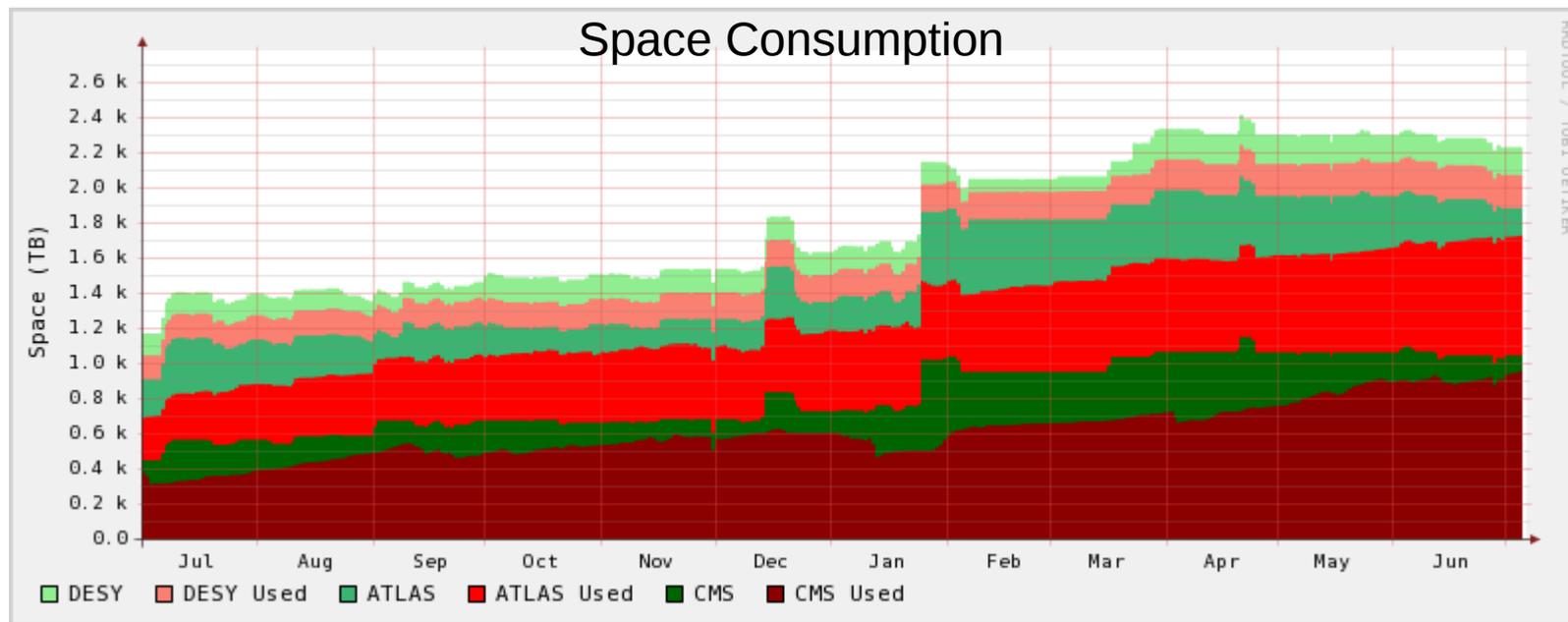
# Statistics: Jobs at DESY-HH (monthly)



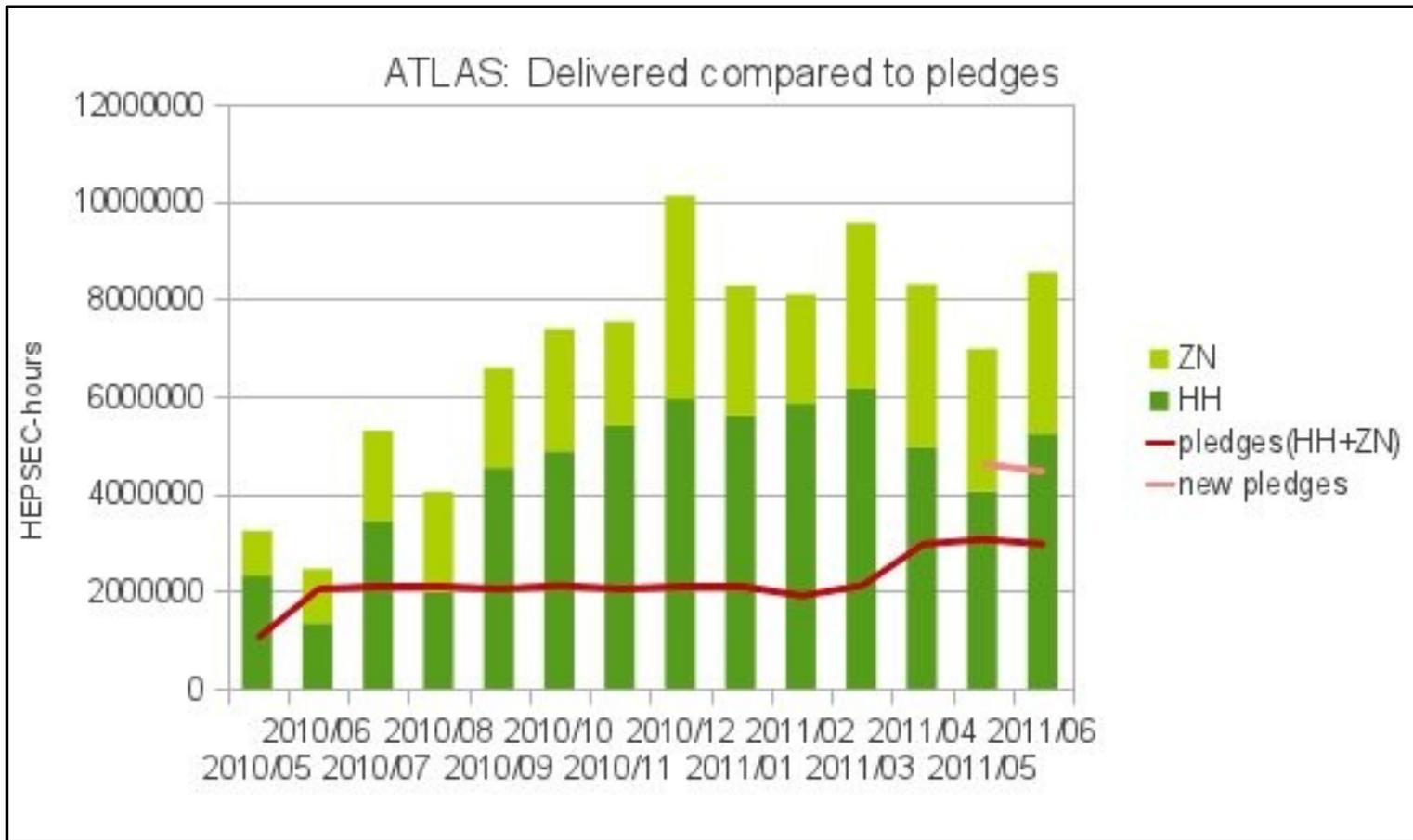
# Statistics: DESY Production (all VOs)



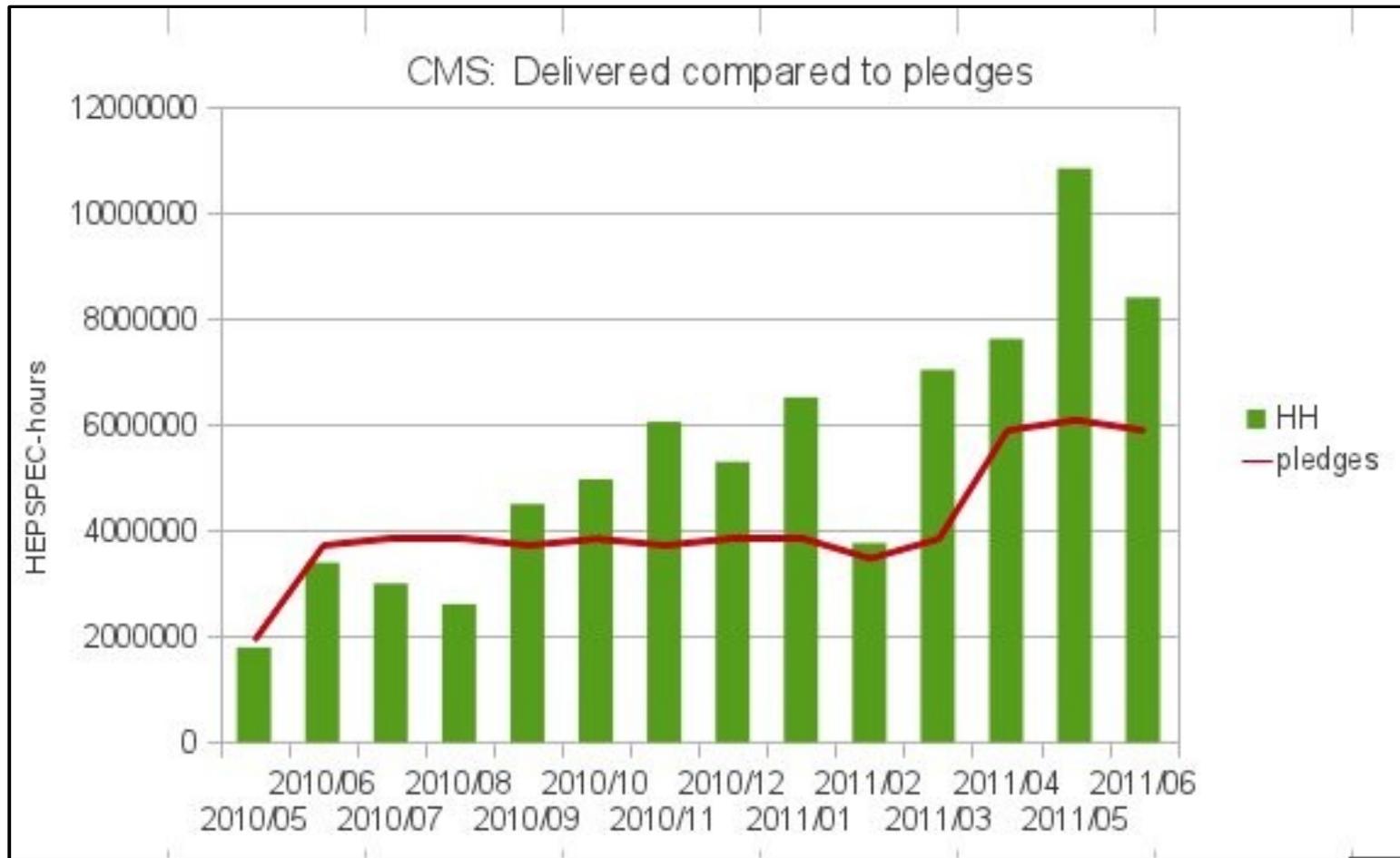
# Statistics: Space Consumption (DESY-HH)



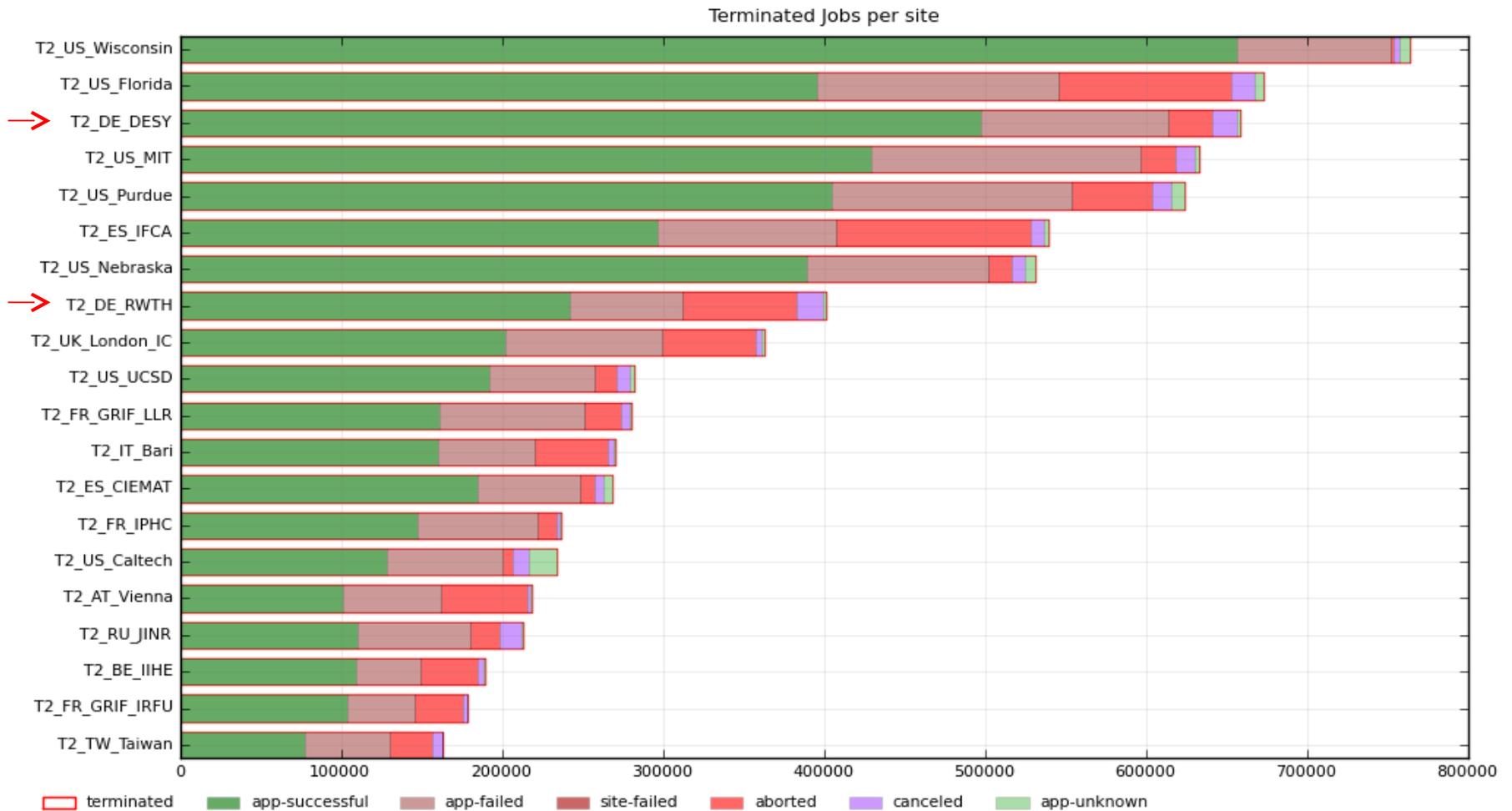
# Statistics: DESY Production (ATLAS)



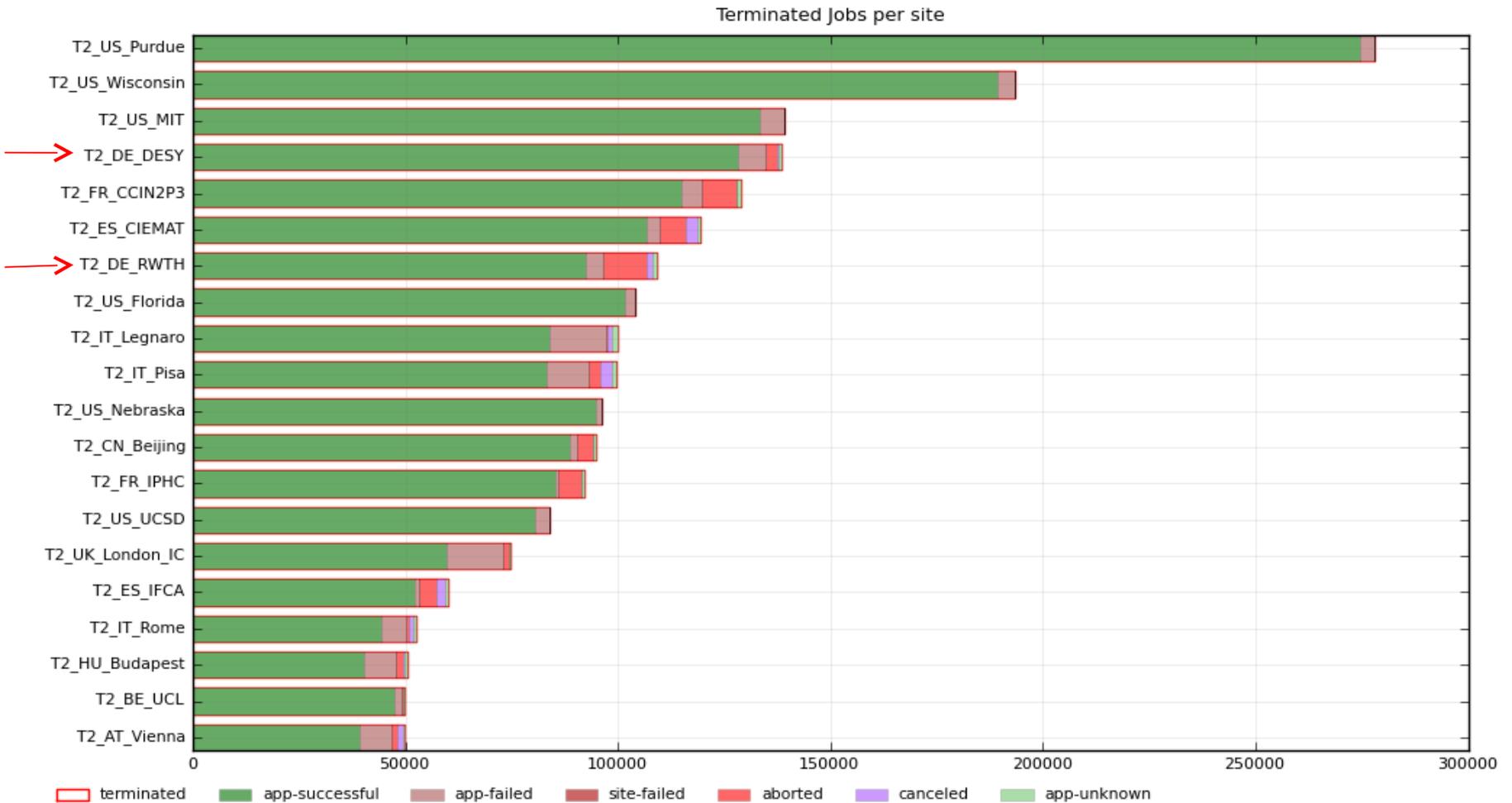
# Statistics: DESY Production (CMS)



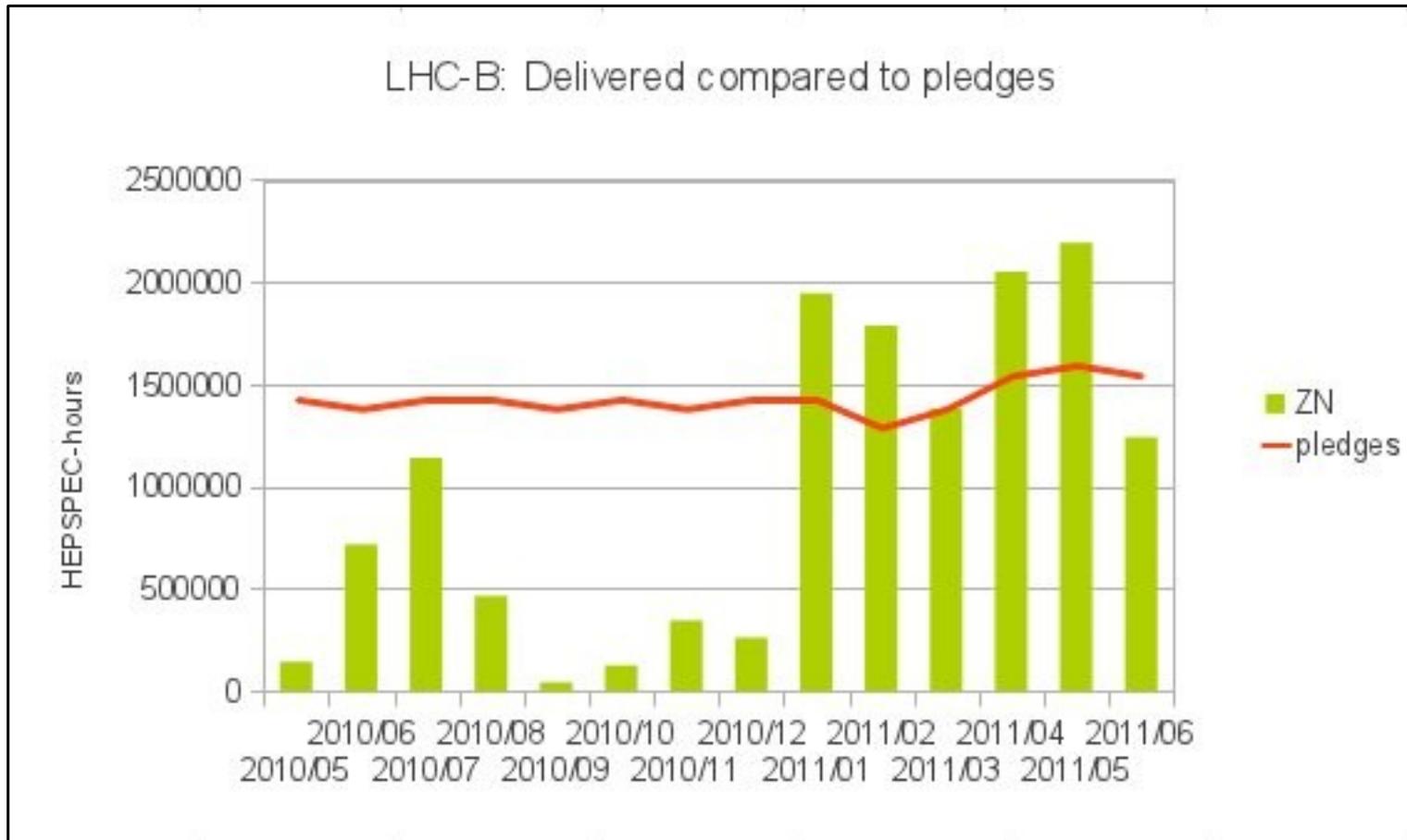
# Statistics: Tier-2 Production (CMS)



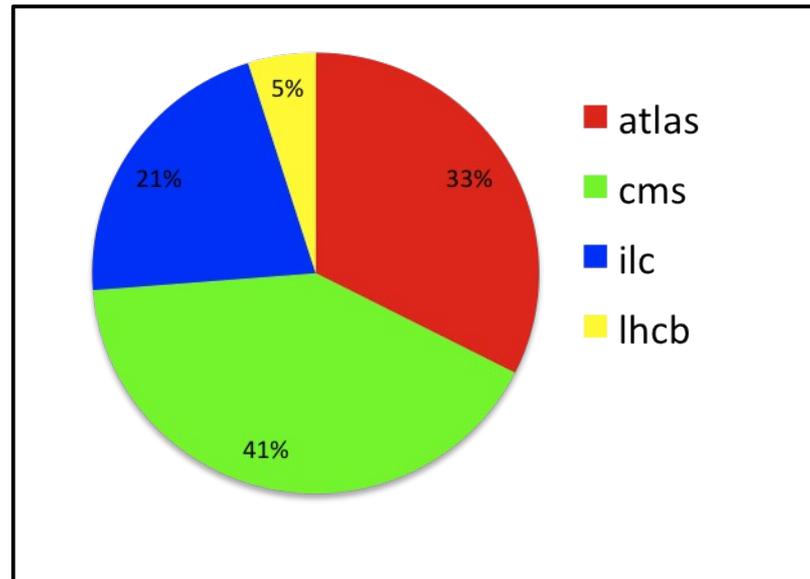
# Statistics: Tier-2 Production (CMS MC 2011)



# Statistics: DESY Production (LHCb)



# Statistics: NAF usage by VO's



# Experiences: Paradigms

- > Main goal: **Keep site in production!**
- > Multi-VO Grid infrastructure serving various science fields
  - Generic Grid infrastructure
  - DESY is the home of 10 VOs
  - Grid services
  - Grid resources
  - Tier-2 is a (big) part of the DESY Grid infrastructure
- > Federating resources
- > Opportunistic use of resources



# Experiences: Operations (Jobs)

- > Up to 24 (mostly 16 or 8) job slots per physical WN
- > Always 2 GB of physical memory per slot (+2GB swap)
- > Number of jobs per user per WN limited to 50% of slots
- > Local disk usage is limited to 15/20 GB
- > Local memory usage is limited to 6 GB
- > CPU/Wall-time limit on queues: 48h/72h
- > Jan – Jun 2011: (½ year)
  - ATLAS : CMS : OTH = 32.8% : 48.1% : 19.1% (CPU time)
  - 4M jobs (2.1% canceled)
  - 1748 CPUy (9.3% canceled)
  - 14kHS06y (1.3kHS06y canceled) [→ 2.6kHS06/y is a ~½ Tier-2]



# Experiences: Operations (Data)

- > Jobs are *transient* / data are *persistent* ...
- > dCache SEs
- > We must limit the number of user jobs to keep the job efficiency up
- > Depending on the usage patterns, increasing number of jobs does not increase utilized CPU time when job efficiency shrinks
  - The VO's data access patterns differ substantially
  - e.g. for CMS user analysis O(1)M files on hundreds of dCache pool nodes are accessed from the Grid and the NAF *concurrently*
- > Access patterns *could* be distinguished via VOMS groups/roles
- > This does *not* work with pilots!



# Experiences: **Services & Middleware**

- > Overall stability and performance
- > gLite packaging (BDII, openLDAP 2.4)
- > CREAM stability
- > GLEXEC / ARGUS
- > TORQUE upgrade (`/var/spool/pbs` → `/var/torque`)
- > TORQUE security advice
- > UI for SL5 / 32-bit used on desktops / laptops
- > EMI vs. gLite: yaim?



# Experiences: Support

- > Grid sites are part of global infrastructure and depend on it
  - Network
  - Data transfers
  - VOMS
  - catalogues
  
- > Big multi-VO:
  - Long term expertise
  - Different VOs with different requirements
  
- > Smaller university sites:
  - Clear assignment to a small number of VOs
  - Usually, less expertise and little long term personnel



# Experiences: Support (cont'd)

- > Downtime management
  
- > Accounting
  
- > WLCG: Alliance of sites *or* VOs *or* middleware devs?
  
- > Recent security drill:
  - Site admins feel primarily responsible for their sites and might not want to distinguish the pilot user from the payload user



# Experiences: Issues & Concerns

- > Conceptual / paradigmatic changes:
  - authorization via VOMS groups/roles vs. Pilots (GLESEC)
  - Keep Grid infrastructure generic vs. extras for VOs (CERN-VMFS)
  - Changing computing models (data distribution)
  
- > Future of TORQUE
  - Scalability
  - Alternatives? SGE? SLURM? Middleware?
  
- > „*whole node scheduling*“
  - Middleware capabilities
  - Accounting
  
- > Virtualization / Clouds (→ HEPiX)



# Experiences: **Issues & Concerns (cont'd)**

- > Does WLCG take EGI/EMI seriously?
- > Will WLCG be compatible with Grid infrastructure in the future?
- > Fraction of WLCG in the future?



# Conclusions

- > The DESY Grid Center reflects the scientific computing at DESY
- > We see (increasing) demands of non-LHC and non-HEP communities
- > We run *one* (generic) Grid infrastructure for *all* VOs, federating resources with opportunistic usage
- > The NAF complements the Grid for interactive data analysis
- > Operations and support is rather complex
- > We need middleware which allows us to run the Grid center
- > Generally, ad hoc solutions do not help ...

