Accounting Infrastructure in EGEE

Pablo Rey, Carlos Fernandez, Javier Lopez/ CESGA
Dave Kant, John Gordon / RAL

IberGrid

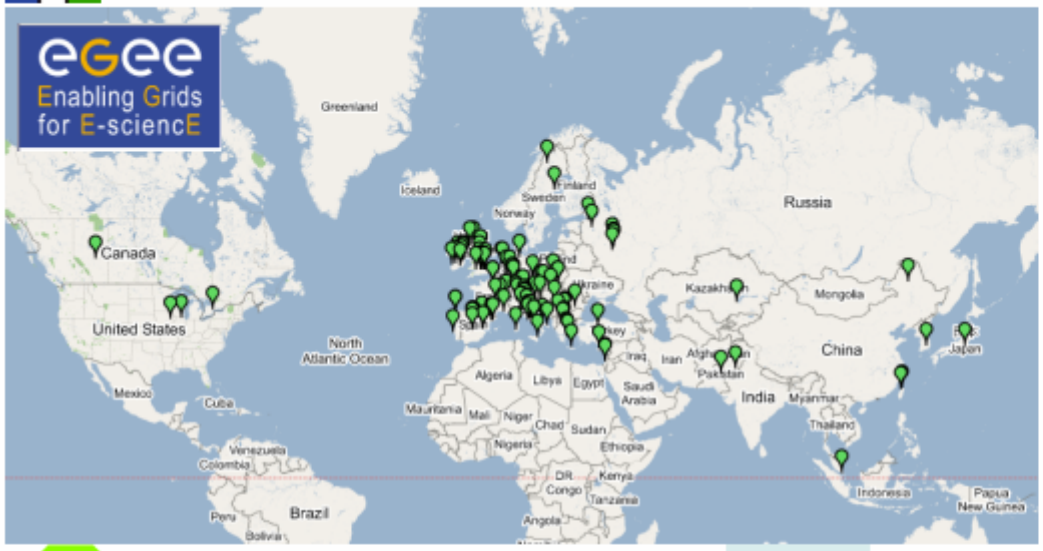
Santiago de Compostela, May 15, 2007

Outline

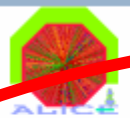
- **Introduction**
- **Accounting Architecture**
- **Accounting Portal**
- **Grid usage in depth**
  - Who?
  - How?
  - When?
- **Conclusions**

Introduction

Grid Projects Collaborating in LHC Computing Grid

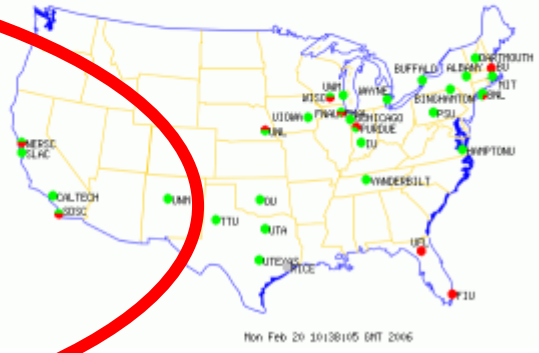


Open Science Grid



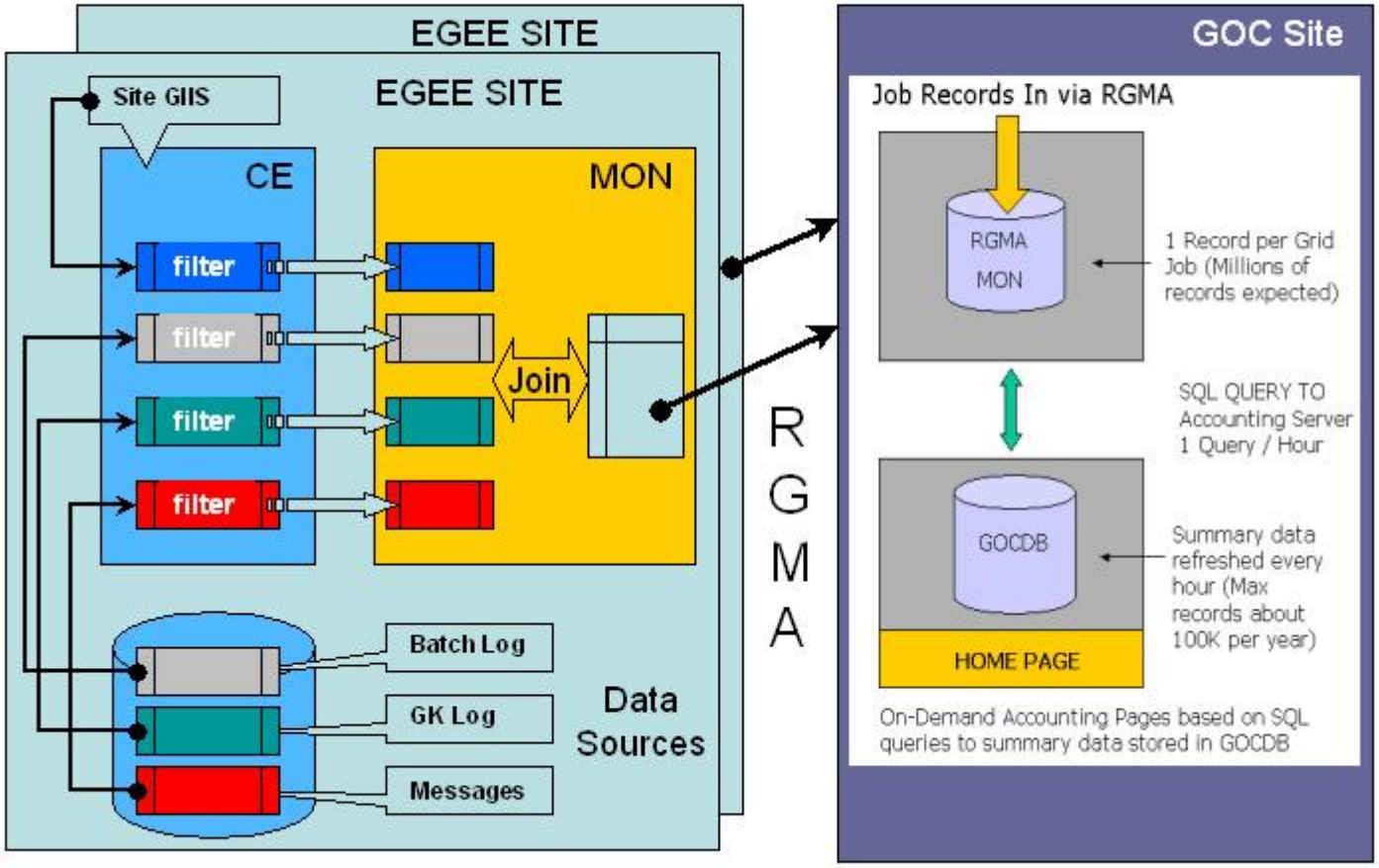
EGEE Operations Information	
Active Sites	177
Available CPU	29097
Available Storage (TB)	13831

LastBuild:Fri May 4 09:16:01 BST 2007 GstatQuery:2006-12-15



Accounting Architecture

Accounting Flow Diagram



OSG, Nordugrid, INFN-Grid and GridPP sites collaborating with WLCG also publish their accounting data into GOC accounting database (by other means) → Single view of WLCG VOs

Accounting Portal

EGEE View | VO MANAGER View | VO MEMBER View | SITE ADMIN View | USER View

Hierarchical Tree

- Tier1
- Countries
- EGEE
  - Production
  - PPS
- OSG
- UNREGISTERED
- VO_Discipline
- VO_Metrics

Data to graph:	Norm. Sum CPU — Normalised CPU time to a reference value of 1000 SpecInt2000
Period:	Start year: 2006 — Start month: 6 — End year: 2007 — End month: 5
Groupings:	Show data for: REGION — as a function of: VO
VO Groups:	LHC — Official EGEE — ALL — Custom
Chart:	Type: ACCUM BAR — Scale: LINEAR
dteam VO:	Exclude **dteam** and **ops** VOs jobs information

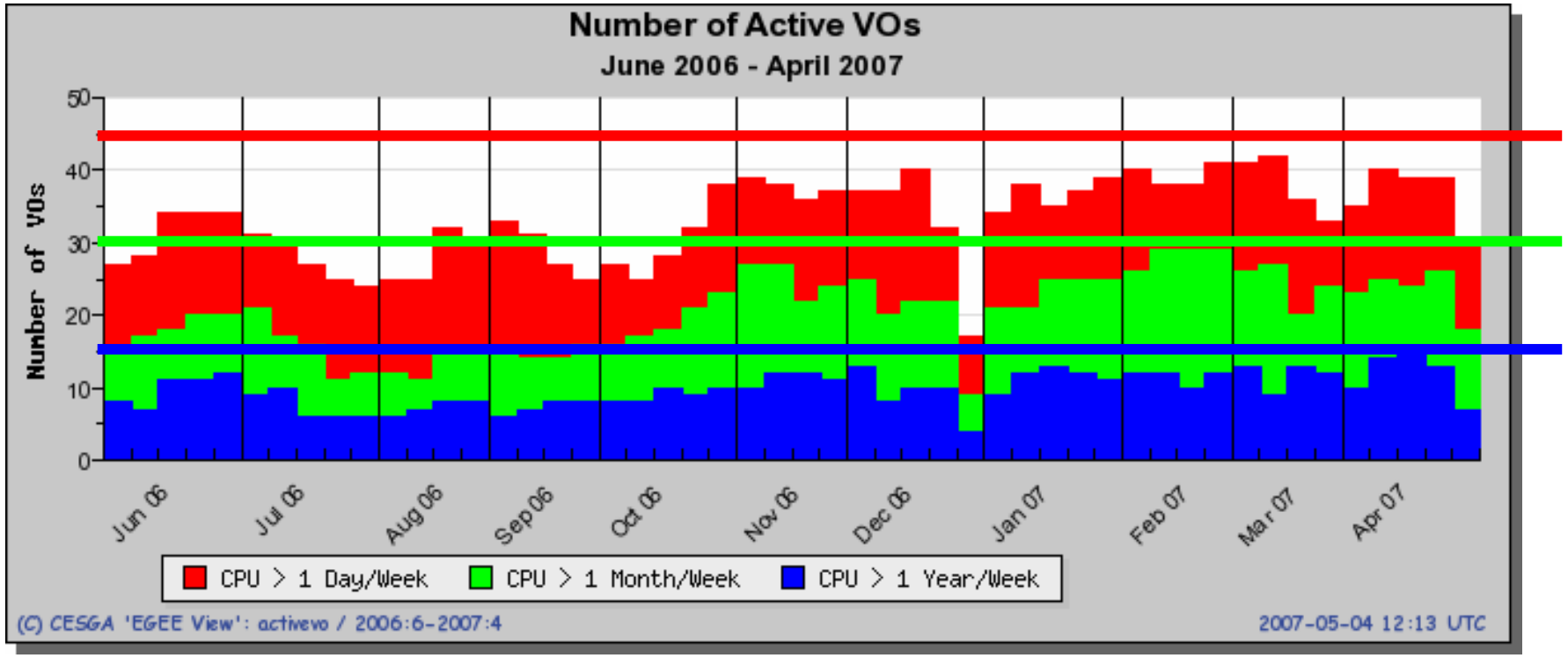
Grid usage in depth

Who

- **Who is using the Grid? …**
- **… or:**
  - How many VO are using the Grid?
  - What are these VO?

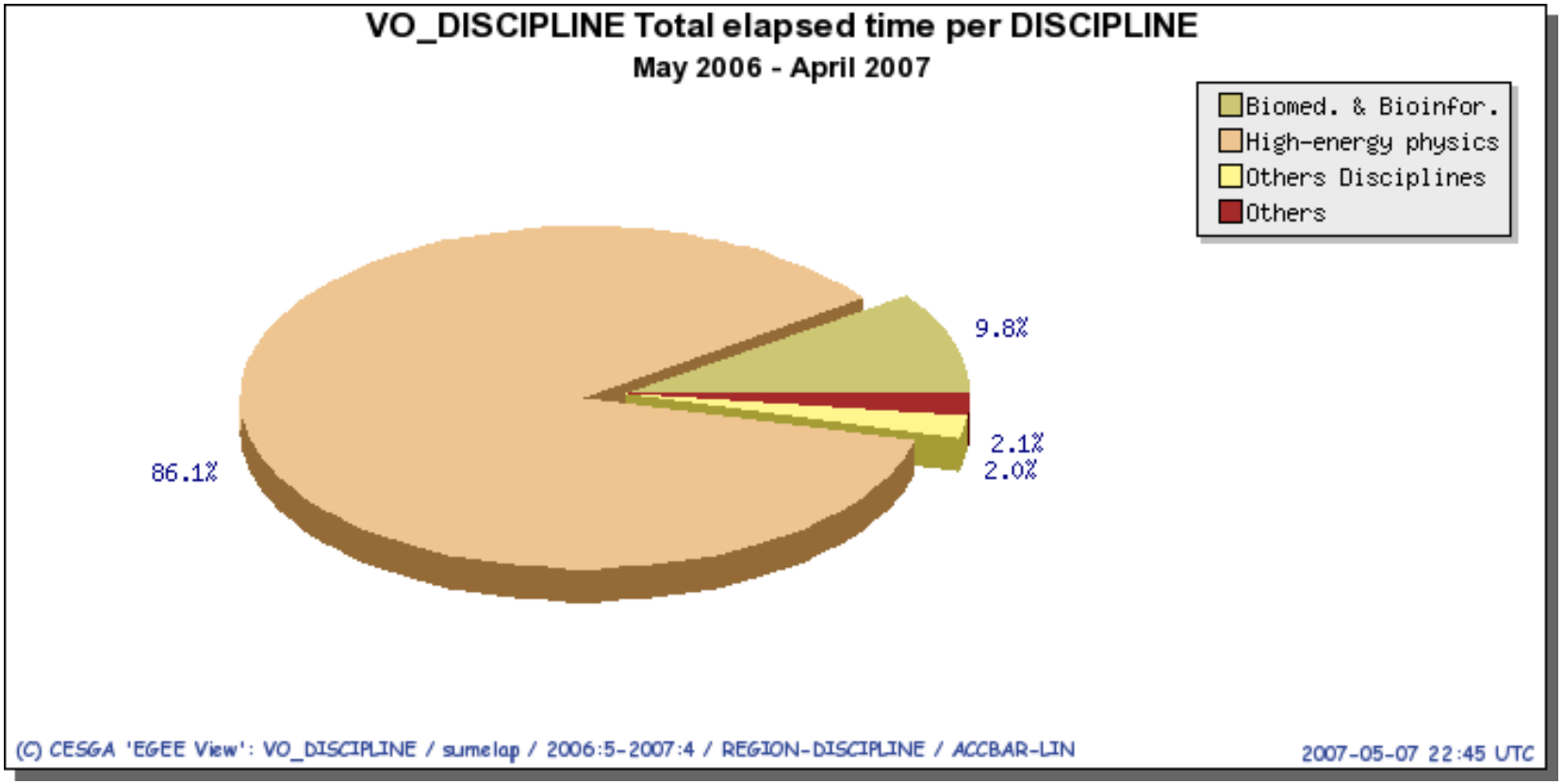
- **First things first: What is the current number of VO in EGEE?**
  - How many VO have published data into the accounting database?
    - 188
  - How many VO are registered in the CIC?
    - 118

Active VOs

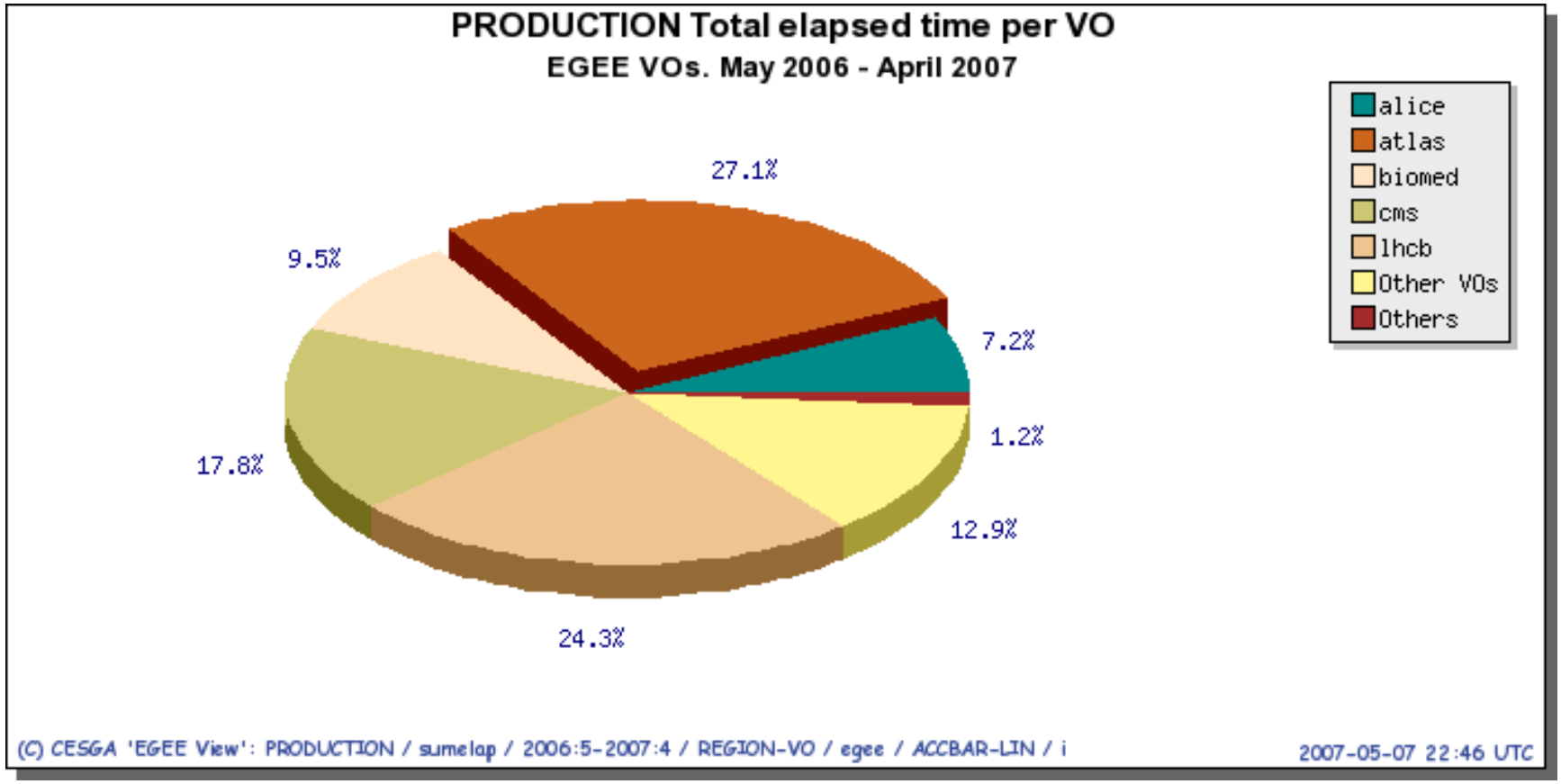


< 30/118

By Discipline



Most Active VOs



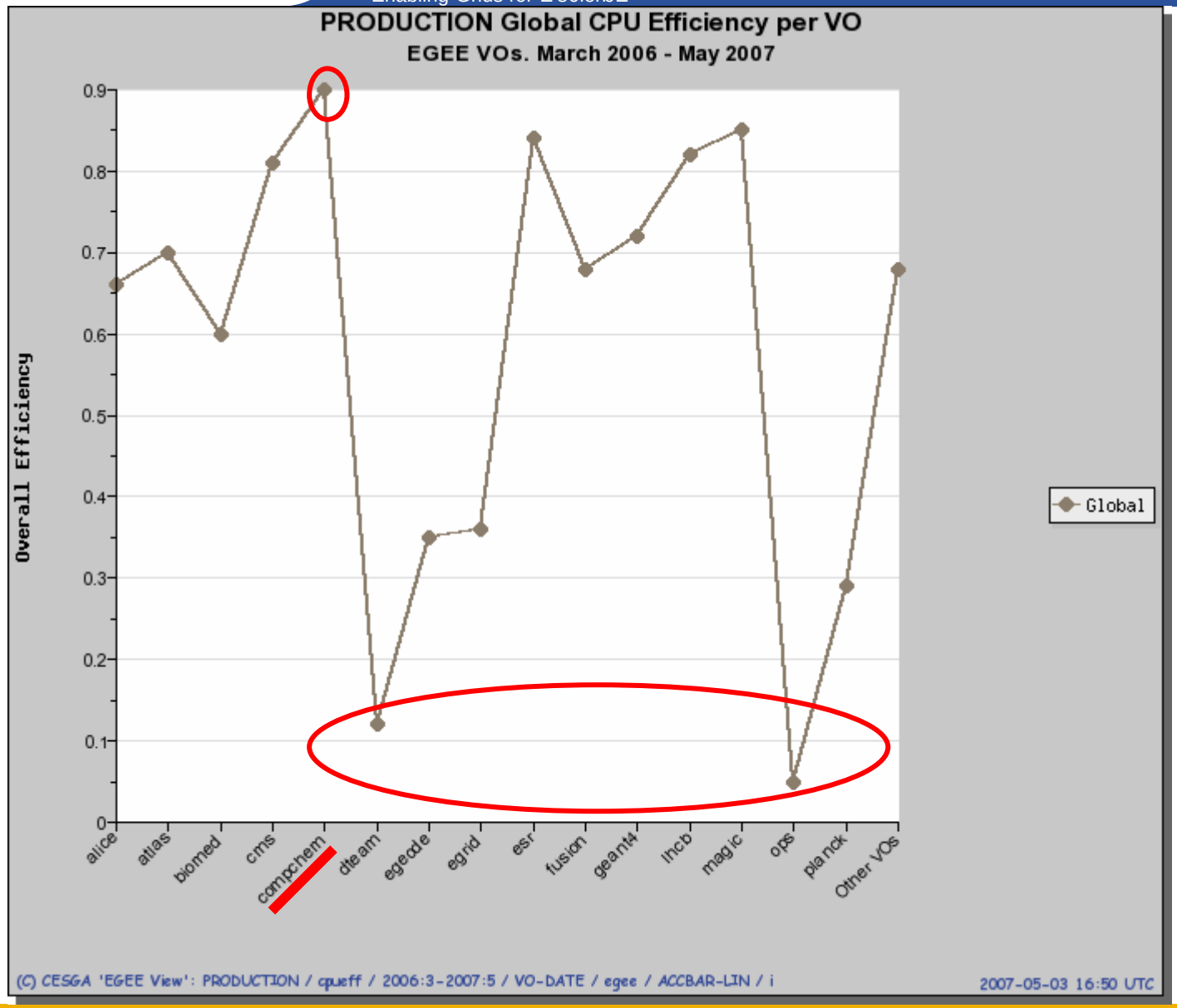
86% (5 VO)

Active VOs

- **An analysis of accounting data published in 2005/2006 by 283 sites and comprised of 25 million job records, shows that the usage is dominated by the LHC and biomedical VOs which together consumed a total of 87% (equivalent to 8269 CPU Years) of the reported usage across EGEE.**

How efficient are the jobs running in the Grid?

CPU Efficiency



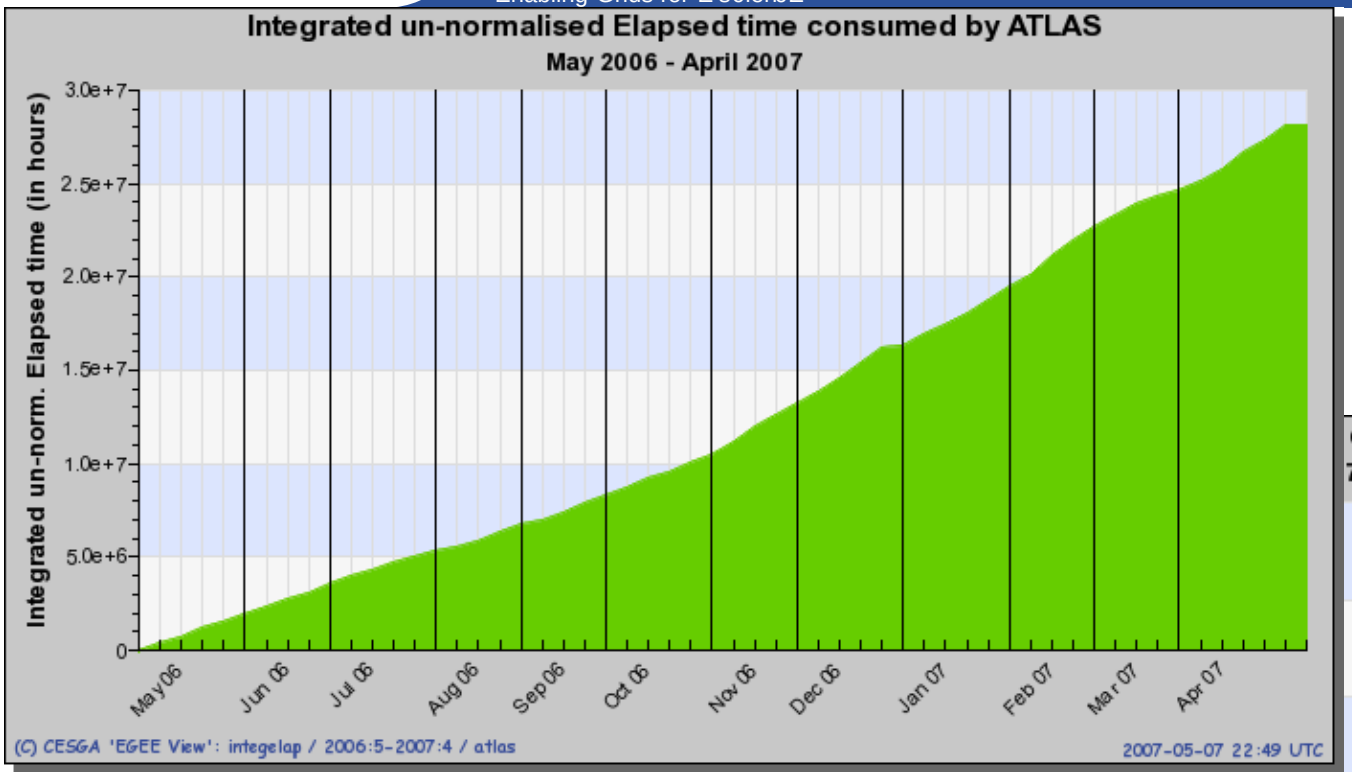
Alice	0.7
Atlas	0.7
Biomed	0.6
CMS	0.8
LHCb	0.8

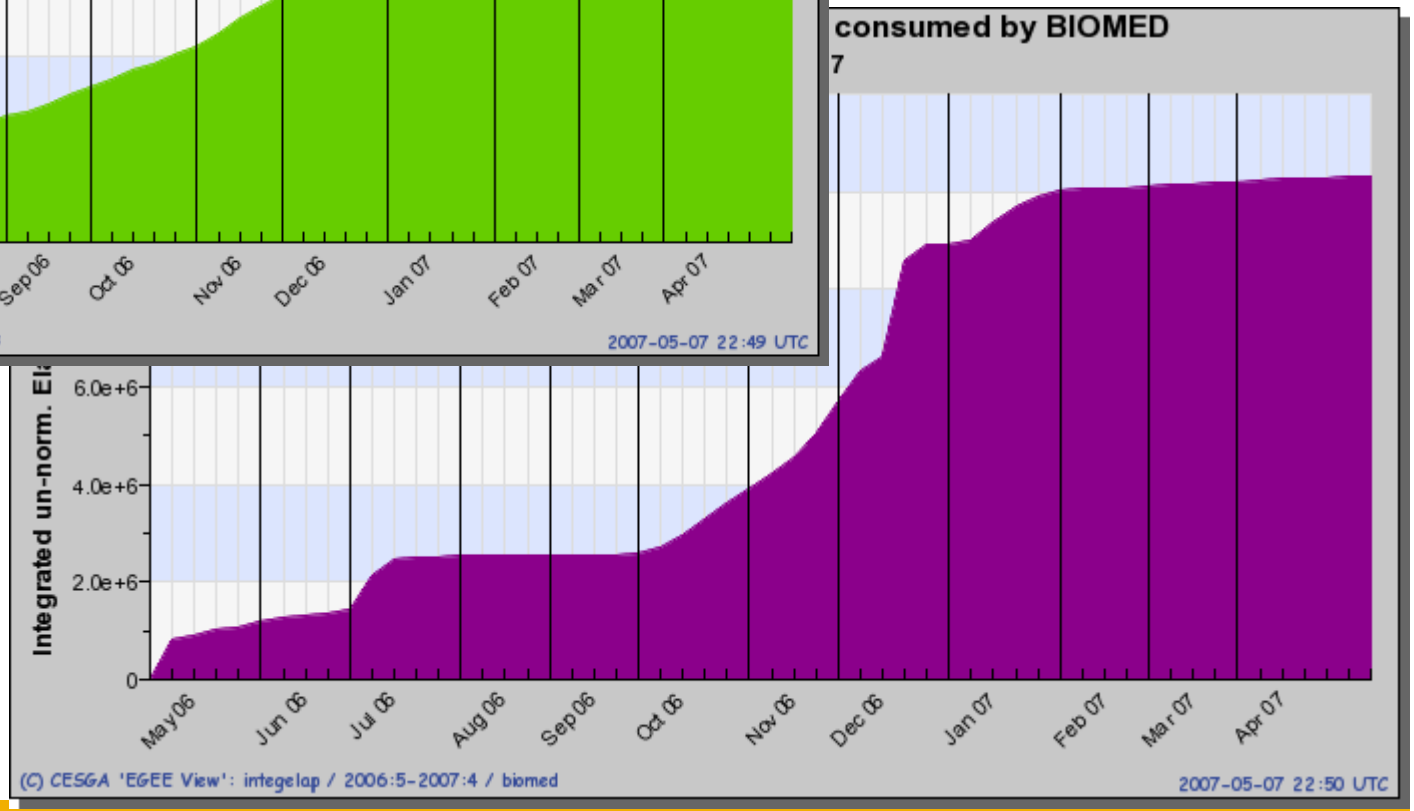
CPU Efficiency

- Only 5 VO have a mean efficiency higher than 0.75
- dteam and ops are the VOs with the lowest efficiency (operational VO)
- VOs that show average efficiencies below 0.50 should ***probably*** review their jobs (unless they already expect such low efficiencies)

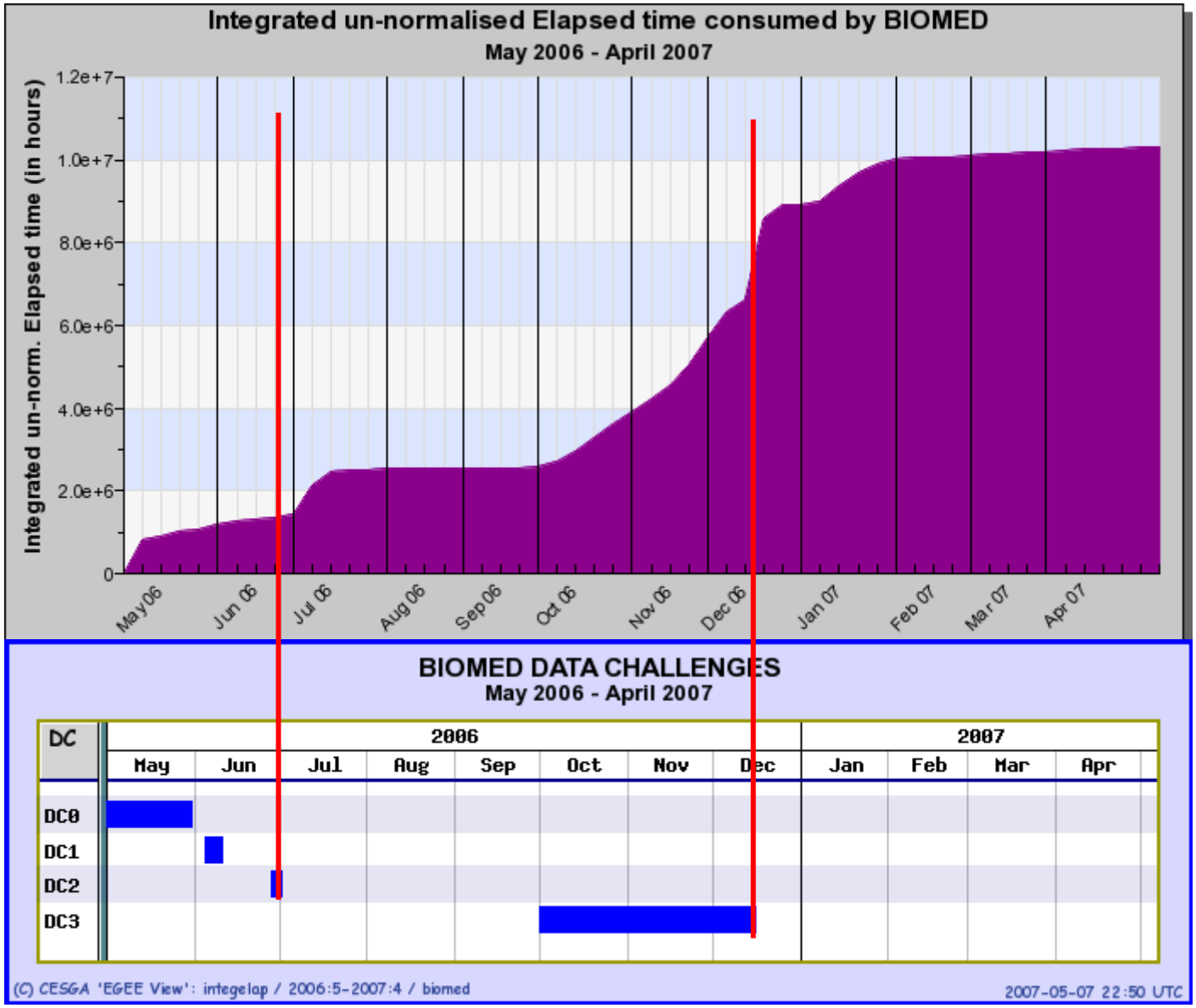
Trends of VO usage

Trends of VO usage





Trends of VO usage



Conclusions

- The accounting infrastructure is in place
- There is a single view of WLCG VOs (includes OSG and other non-EGEE sites)
- The accounting portal allows to analyze how the Grid is being used
- Although there are 118 registered VO less than 30 are actively using the Grid
- 5 VO consumed 86% of the resources
- Only 4 VO have a mean CPU efficiency higher than 0.75
- There are different trends using the Grid
  - Constant job submission
  - Data Challenges

References

- **Accounting Home**
  - http://goc.grid-support.ac.uk/gridsite/accounting/
- **EGEE View**
  - http://www3.egee.cesga.es/gridsite/accounting/CESGA/egee.php
- **EGEE View Wiki Page**
  - http://goc.grid.sinica.edu.tw/gocwiki/EGEE_View
- **Accounting Enforcement**
  - https://edms.cern.ch/document/753895
  - http://www3.egee.cesga.es/acctenfor/

Thanks

Thank you!