

Resource provisioning for biomed

**Patrick Guterl, Franck Michel, Tristan Glatard
for the biomed VO**

OMB, May 29th 2012

biomed virtual organization

- **Users**

- 280 users from ~20 different countries
- Two SMEs (non commercial activities)
- Application fields: Bioinformatics, Drug discovery, Medical Imaging

- **Large infrastructure, loosely controlled**

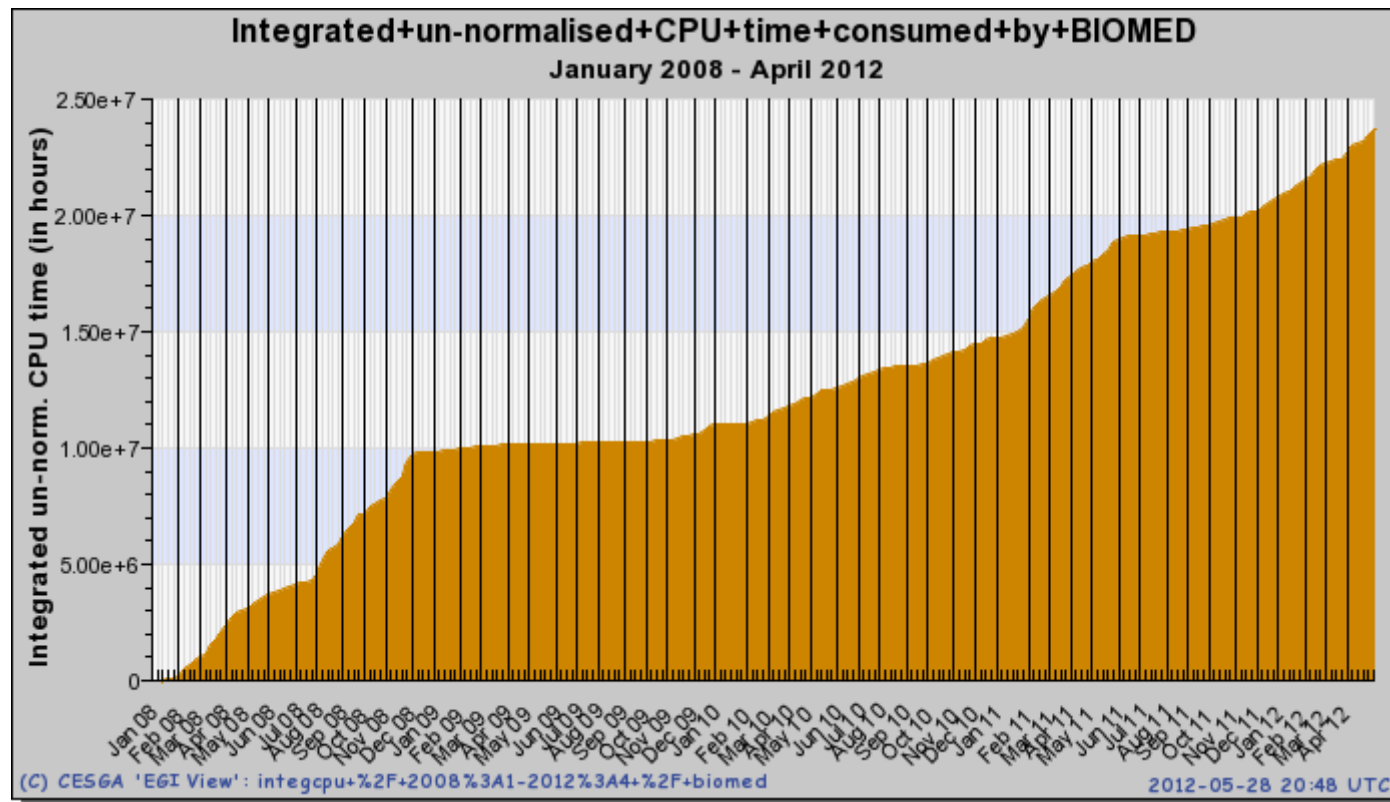
- 238 CEs (batch queues) from 129 sites ; 74 Storage Elements
- No formal agreement with sites
- VO support and management from user groups on a voluntary basis

- **Heterogeneous application environments**

- Heterogeneous tooling (portals, workflow engines, pilot-job systems): DIANE, DIRAC, WISDOM, OpenMole, Moteur, etc
- No central control point

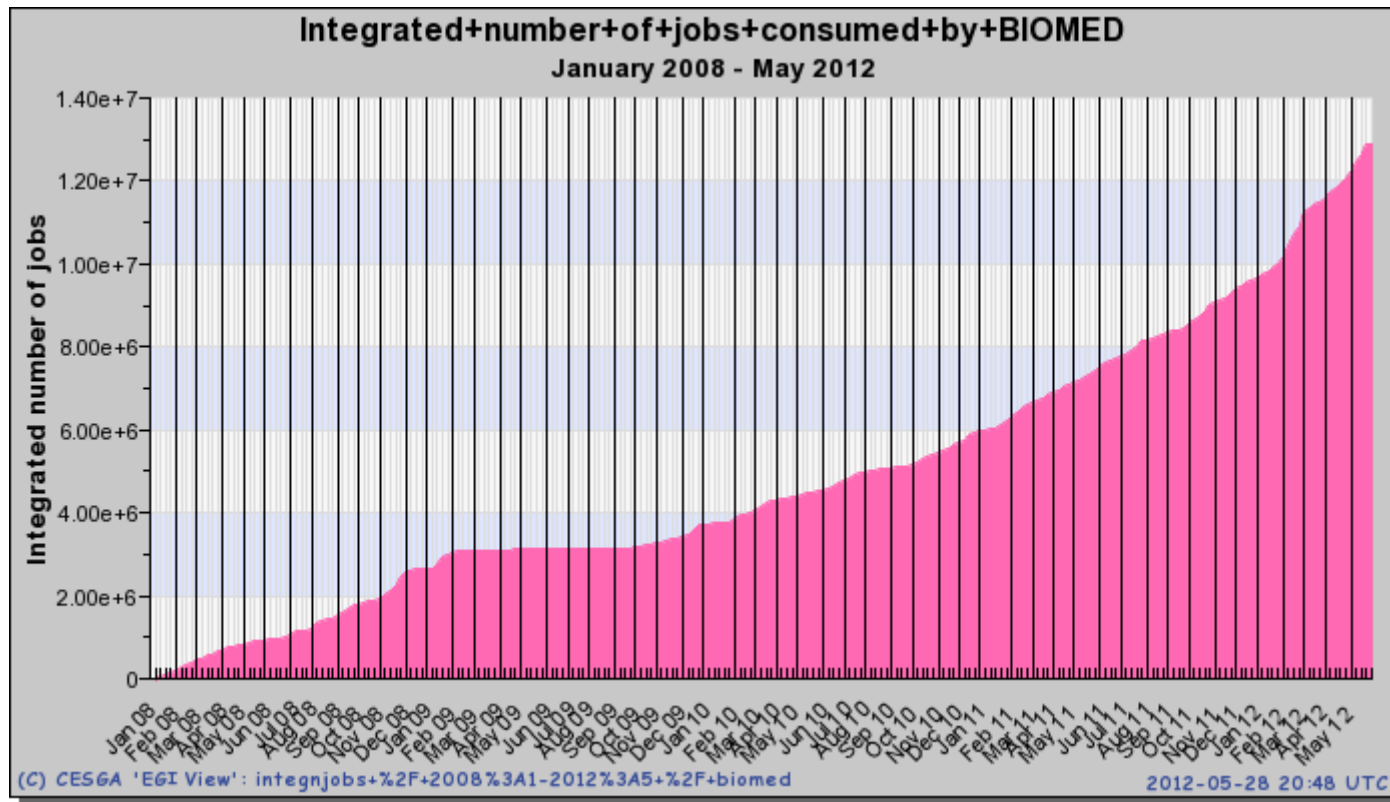
Activity

- **Consumed CPU**
 - No significant increase lately



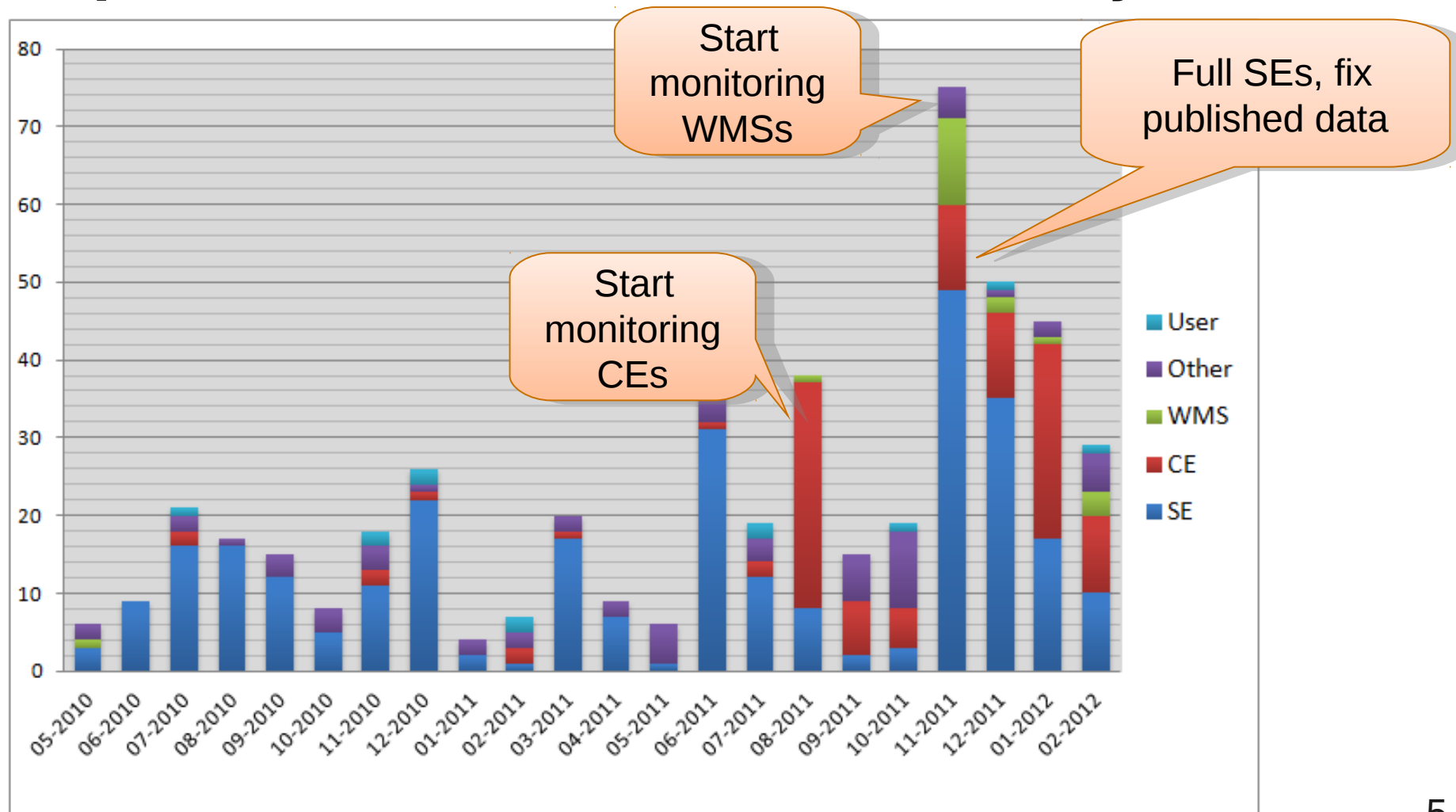
Activity

- **Number of jobs**
 - Slight increased since end 2011



Reported incidents

- **Monitored resources**
 - 108 SEs, 186 CEs, 36 WMSes, 1 LFC
- **450+ operational tickets handled over a year**

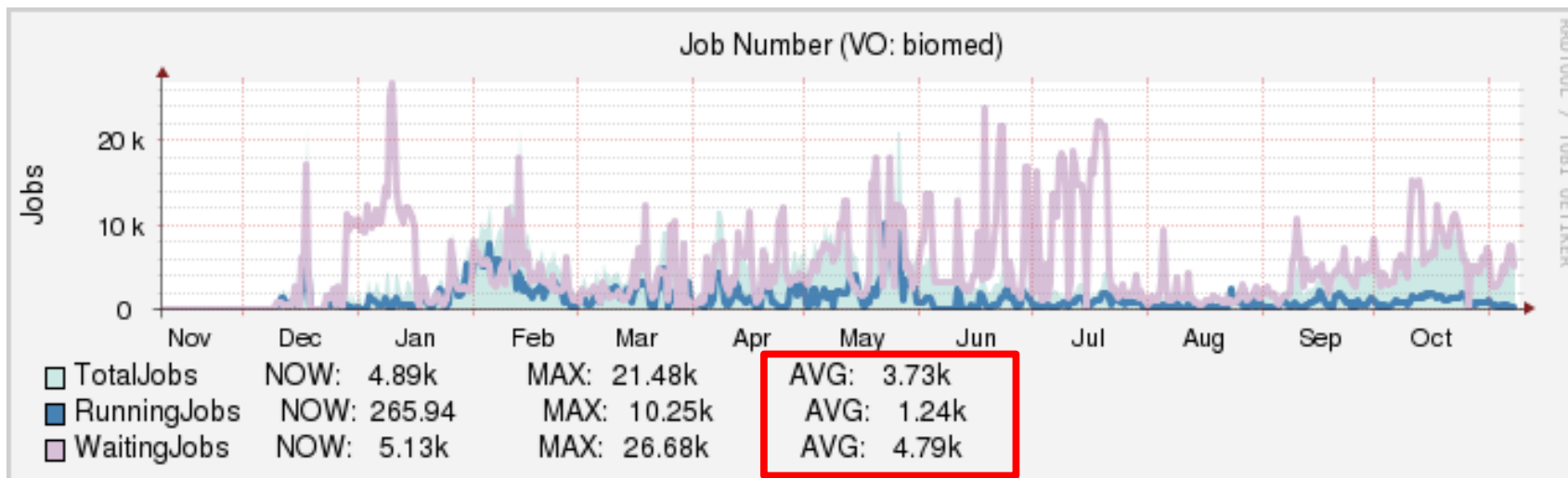


Active probing

- **Probe jobs submitted for 3 days**
 - Frequency: 1 probe every hour
 - To all CEs supporting the VO
 - Direct job submission (without WMS)
 - With 10-minute timeout on queuing
- **Results (see BilanRangeCE_3.xls)**
 - Total number of probes: 7278
 - Average waiting time: 220s
 - Timeout: 30%
 - Failure: 19%

Running vs Waiting over the last year

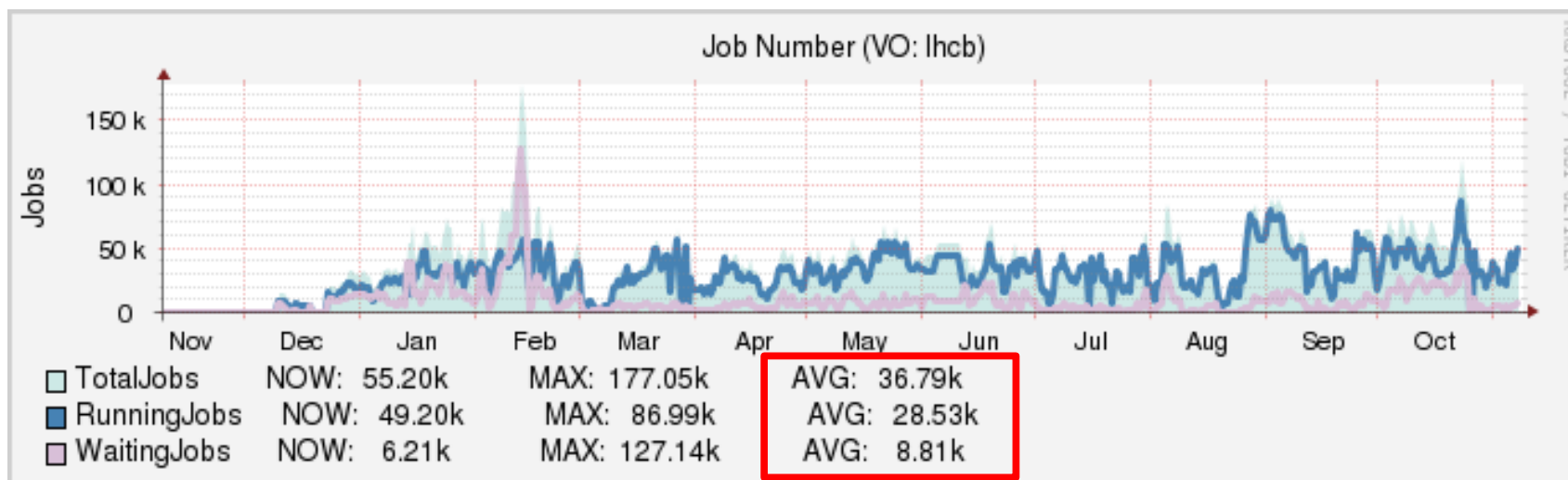
- **biomed**



Source:

<http://gstat-prod.cern.ch>

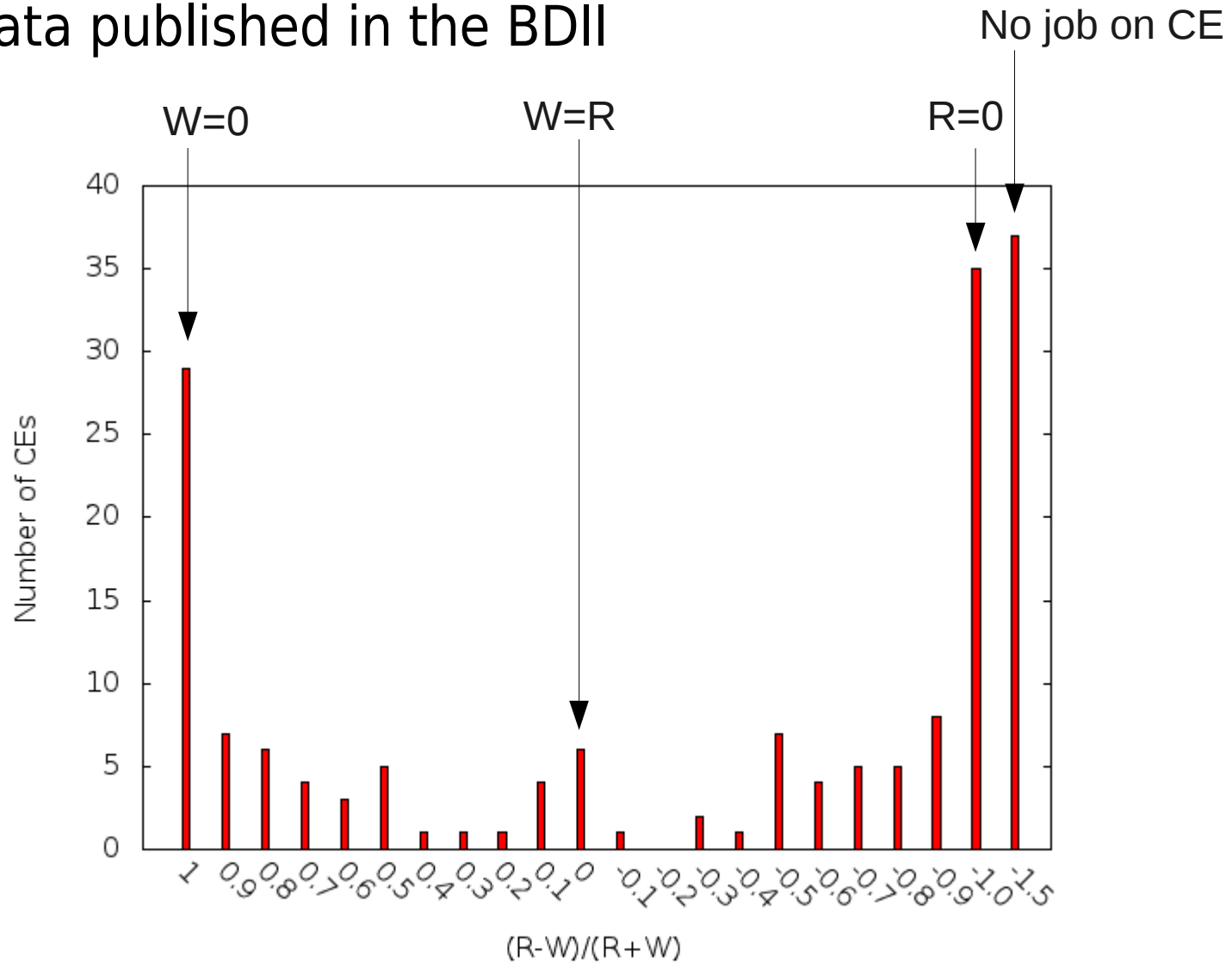
- **LHCb**



Per-CE analysis

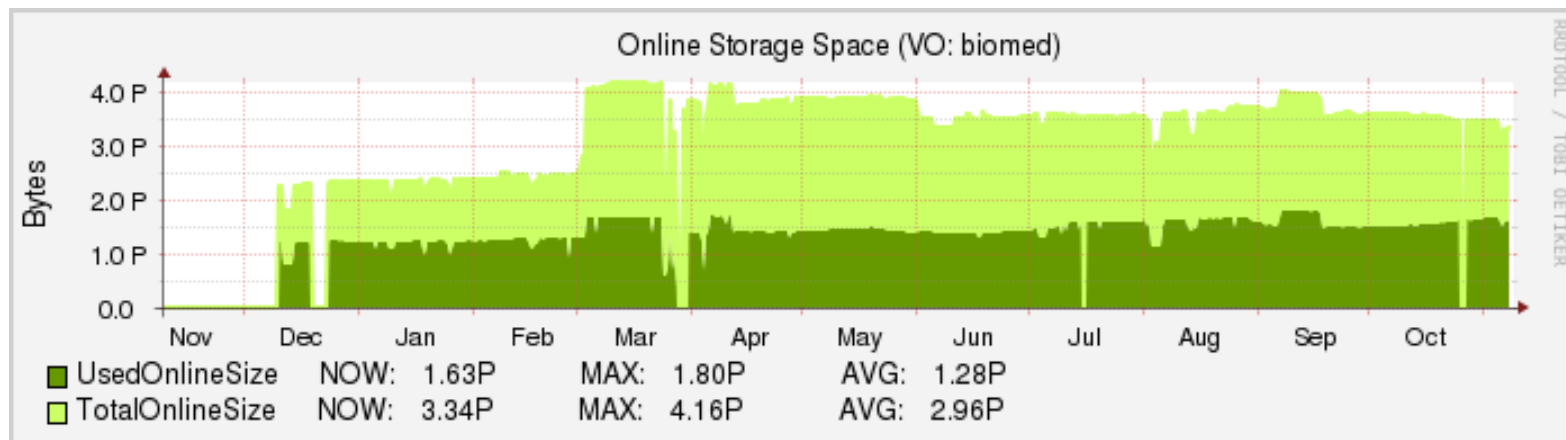
- **Computed $(R-W)/(R+W)$**
 - For all CEs supporting the VO
 - From data published in the BDII

- **Results**

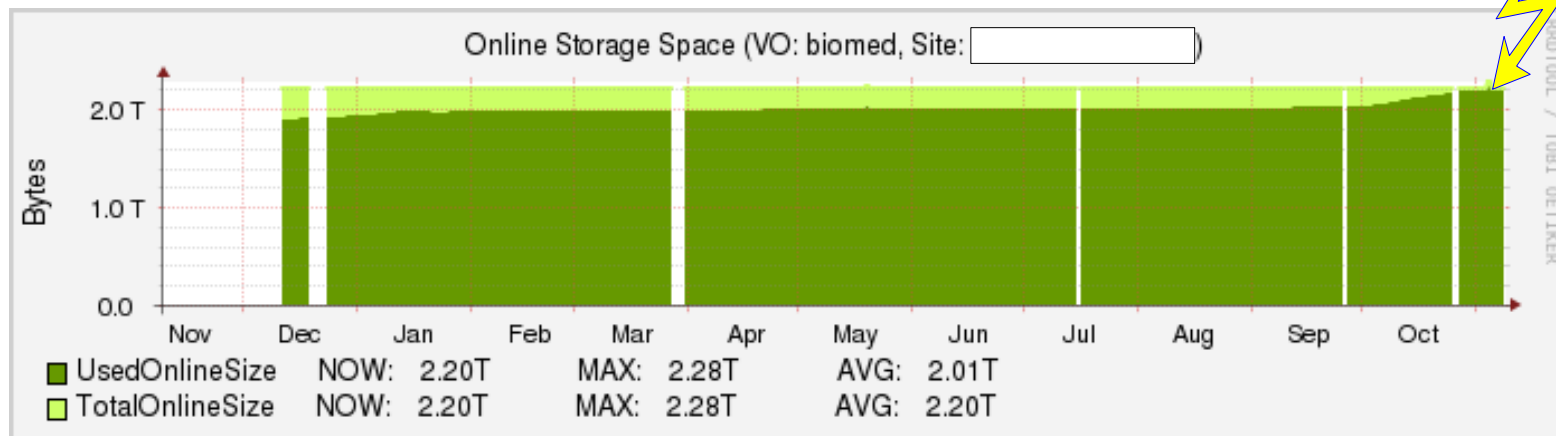


Storage management

- Overall: we don't need more...



- Local: we may need more... or better distribute



Jobs fail

- Related issues

- Data cleanup, especially of users who left the VO
- Small SEs (<1TB): reduce transfer cost, but very often full

Conclusion

- **Computing**

- Need better job scheduling
 - But distributed policies
 - And no consensus on the solution
- And/or more resources

- **Storage**

- Need better storage management
 - But application-specific constraints
 - And no consensus on the solution
- Need cleanup campaigns/tools
- And/or more storage at some sites

Actions identified from the OMB meeting

1. OMB reviews and proposes strategies used by WLCG to better use computing and storage resources. These could serve as advice or recommendation to biomed users.

- Includes rescheduling of jobs by WMS (has anyone tested this?)

2. biomed contacts those users whose jobs have low efficiency, to investigate potential improvements to data transfer issues

- Contacted 2 users about this problem.
- May improve their system in case they use the grid again

3. biomed discusses desired resource quotas per site, and updates VO ID Card accordingly

4. OMB sets up a wiki page advertising how additional resources could be officially requested to NGIs.