First Experience with LCG Operation and the future ...

CERN openlab

Board of Sponsors

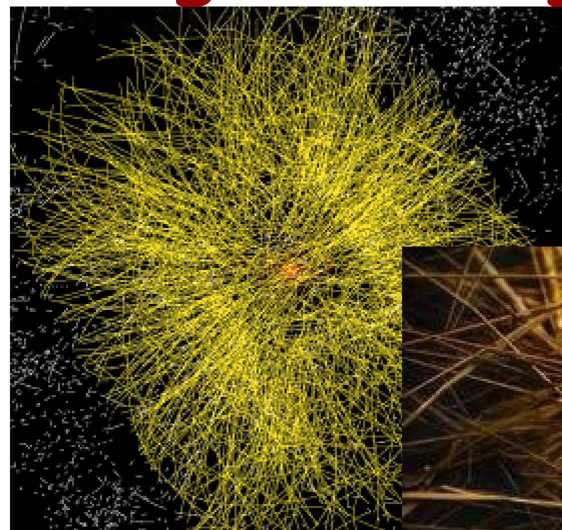
3rd April 2009

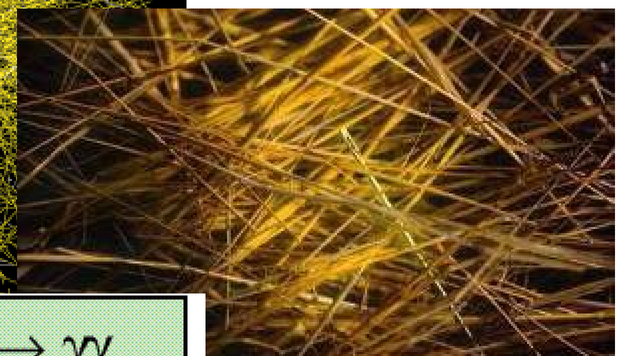
Ian Bird

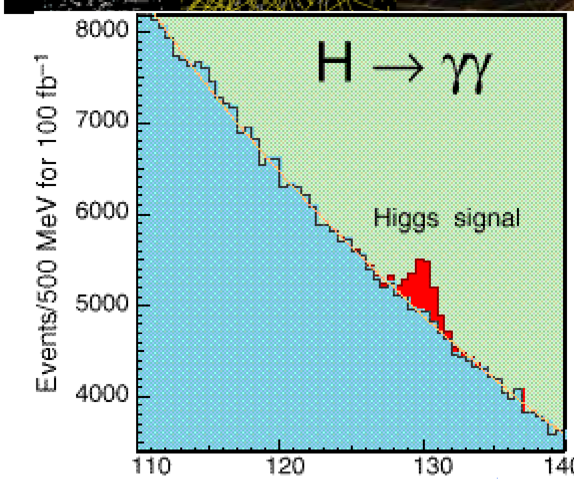
LCG Project Leader

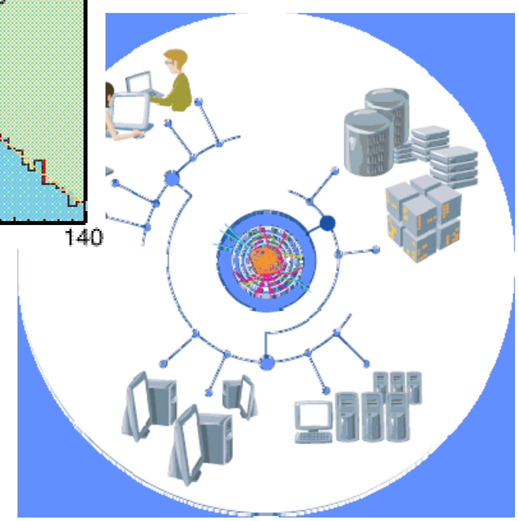
The LHC Computing Challenge

- Signal/Noise: 10^{-9}
- Data volume
  - High rate * large number of channels * 4 experiments
  - ➔ **15 PetaBytes of new data each year**
- Compute power
  - Event complexity * Nb. events * thousands users
  - ➔ **100 k of (today's) fastest CPUs**
  - ➔ **45 PB of disk storage**
- Worldwide analysis & funding
  - Computing funding locally in major regions & countries
  - Efficient analysis everywhere
  - ➔ **GRID technology**









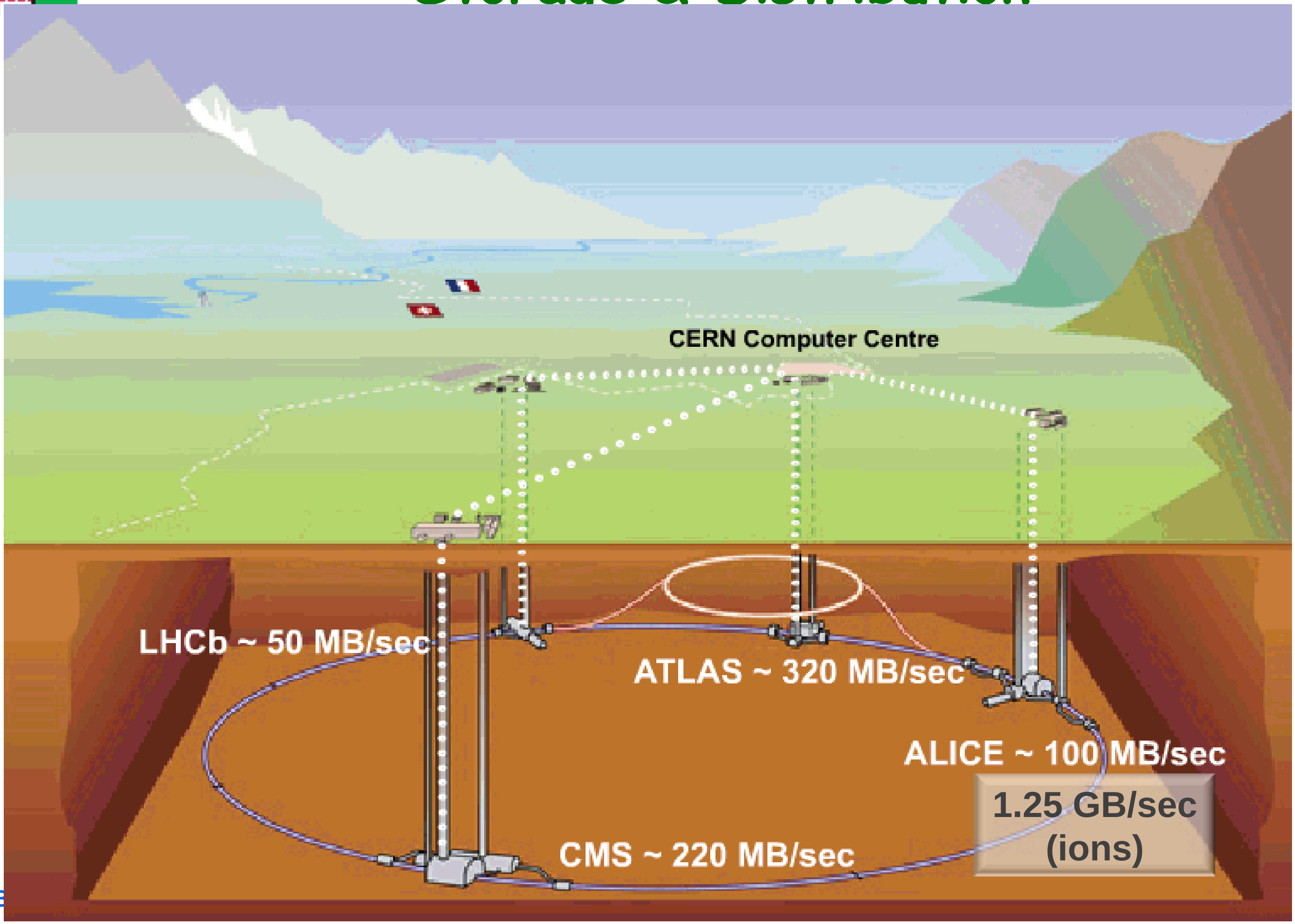
WLCG – what and why?

- A distributed computing infrastructure to provide the production and analysis environments for the LHC experiments
- Managed and operated by a worldwide collaboration between the experiments and the participating computer centres

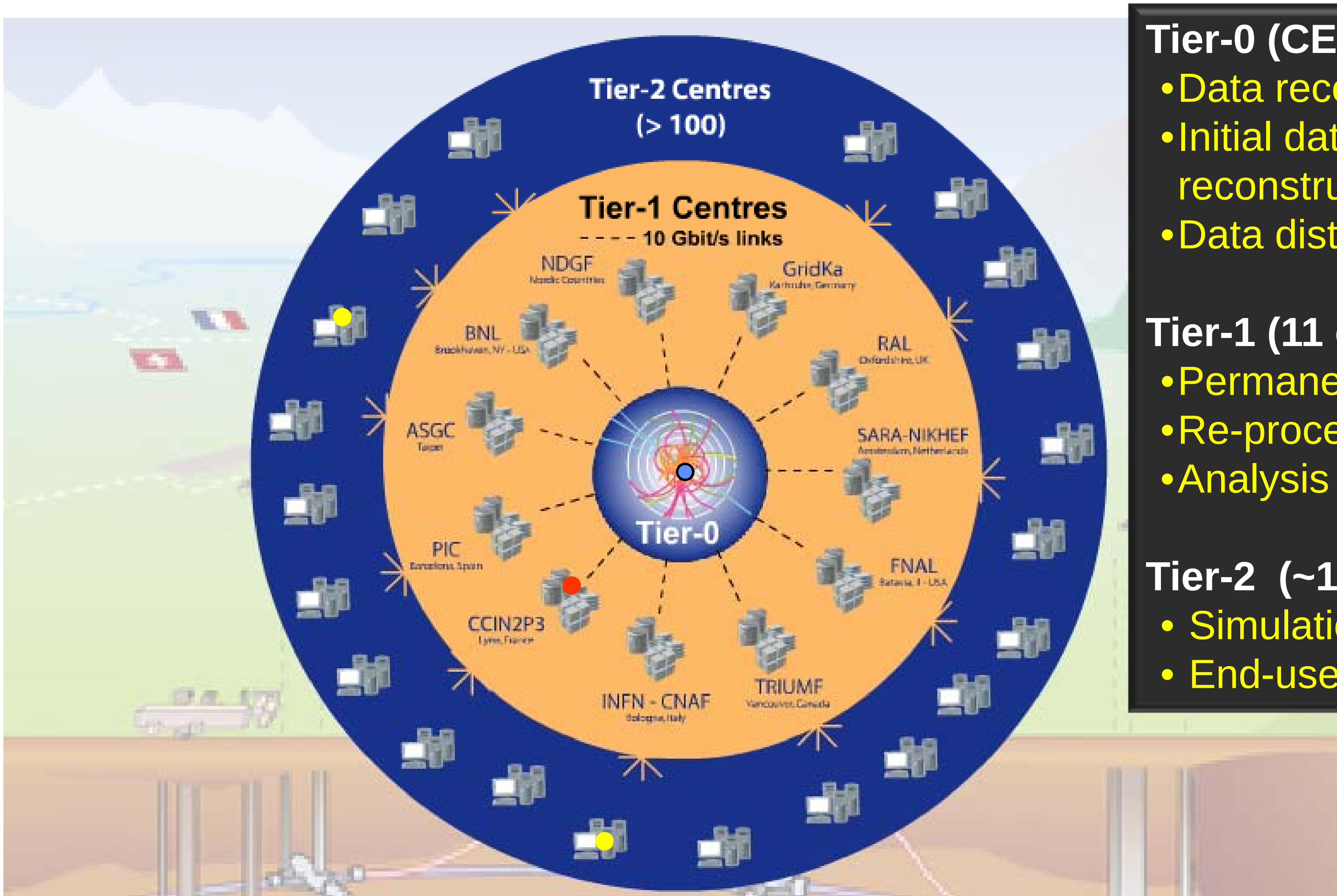
- The resources are distributed – for funding and sociological reasons

- Our task is to make use of the resources available to us – no matter where they are located
  - We know it would be simpler to put all the resources in 1 or 2 large centres
  - This is not an option ... today

Tier 0 at CERN: Acquisition, First pass processing
Storage & Distribution



Tier 0 – Tier 1 – Tier 2

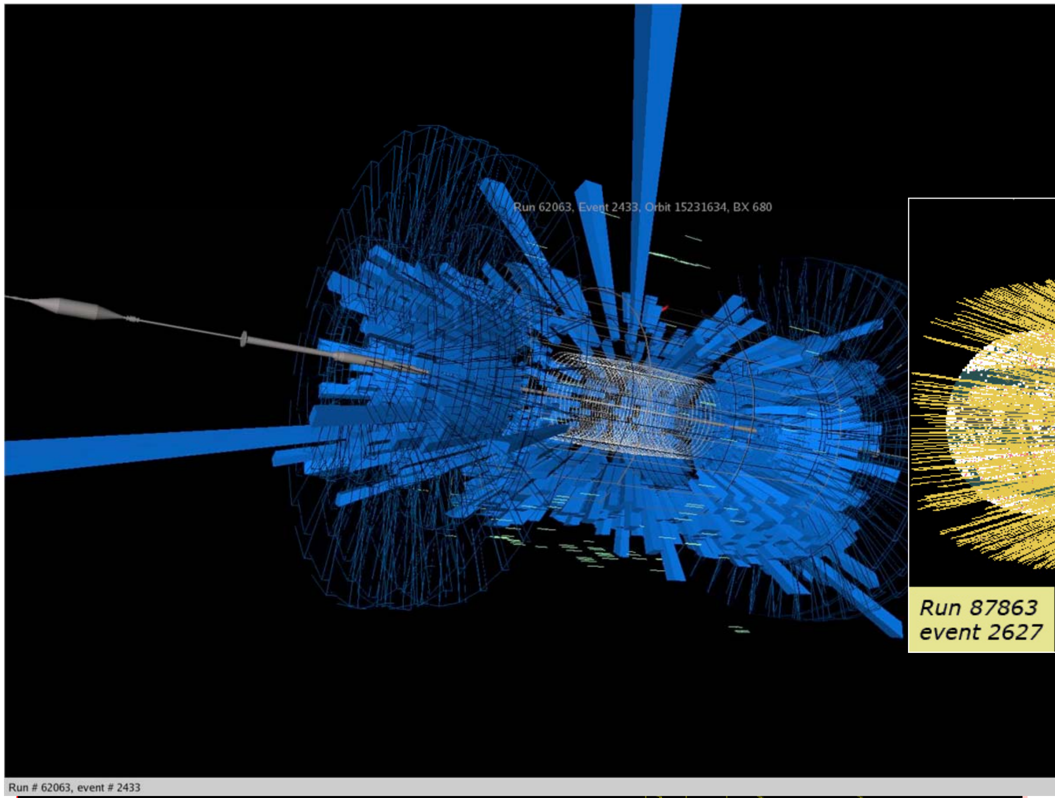


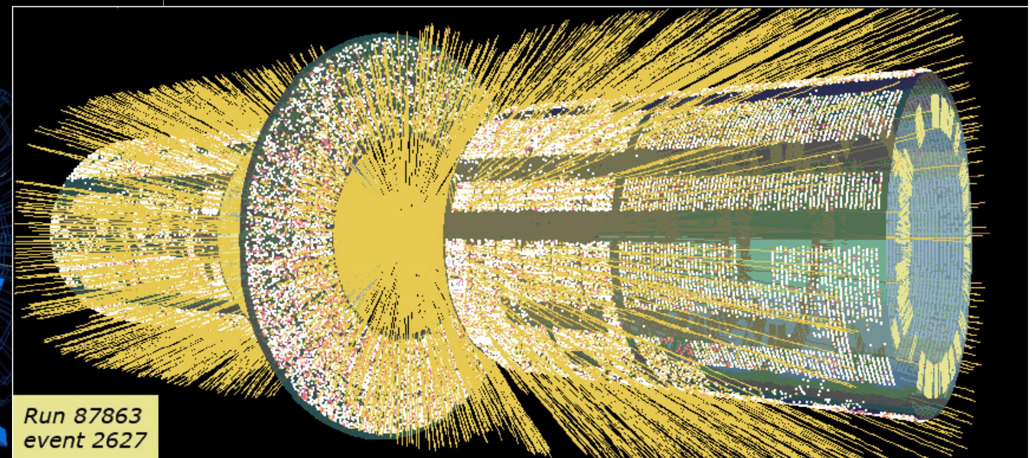
Tier-0 (CERN):
- Data recording
- Initial data reconstruction
- Data distribution

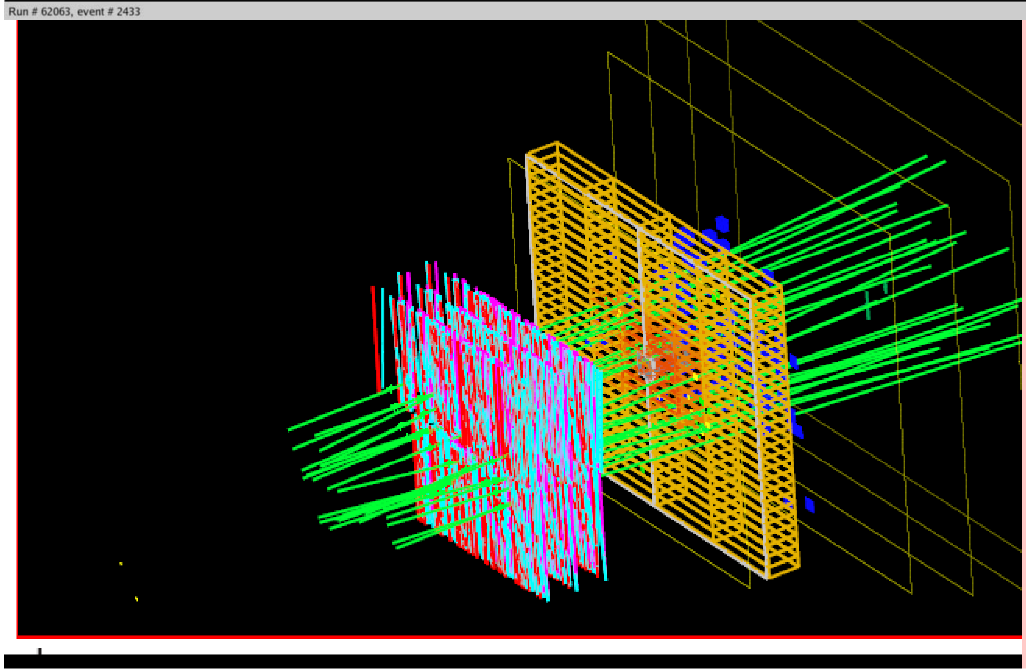
Tier-1 (11 centres):
- Permanent storage
- Re-processing
- Analysis

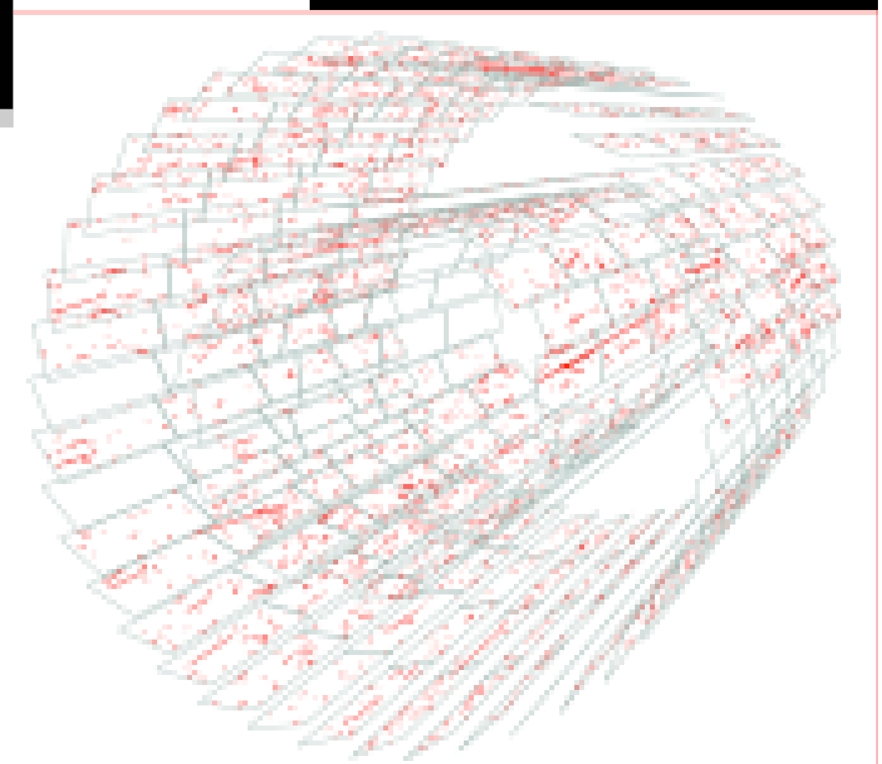
Tier-2 (~130 centres):
- Simulation
- End-user analysis

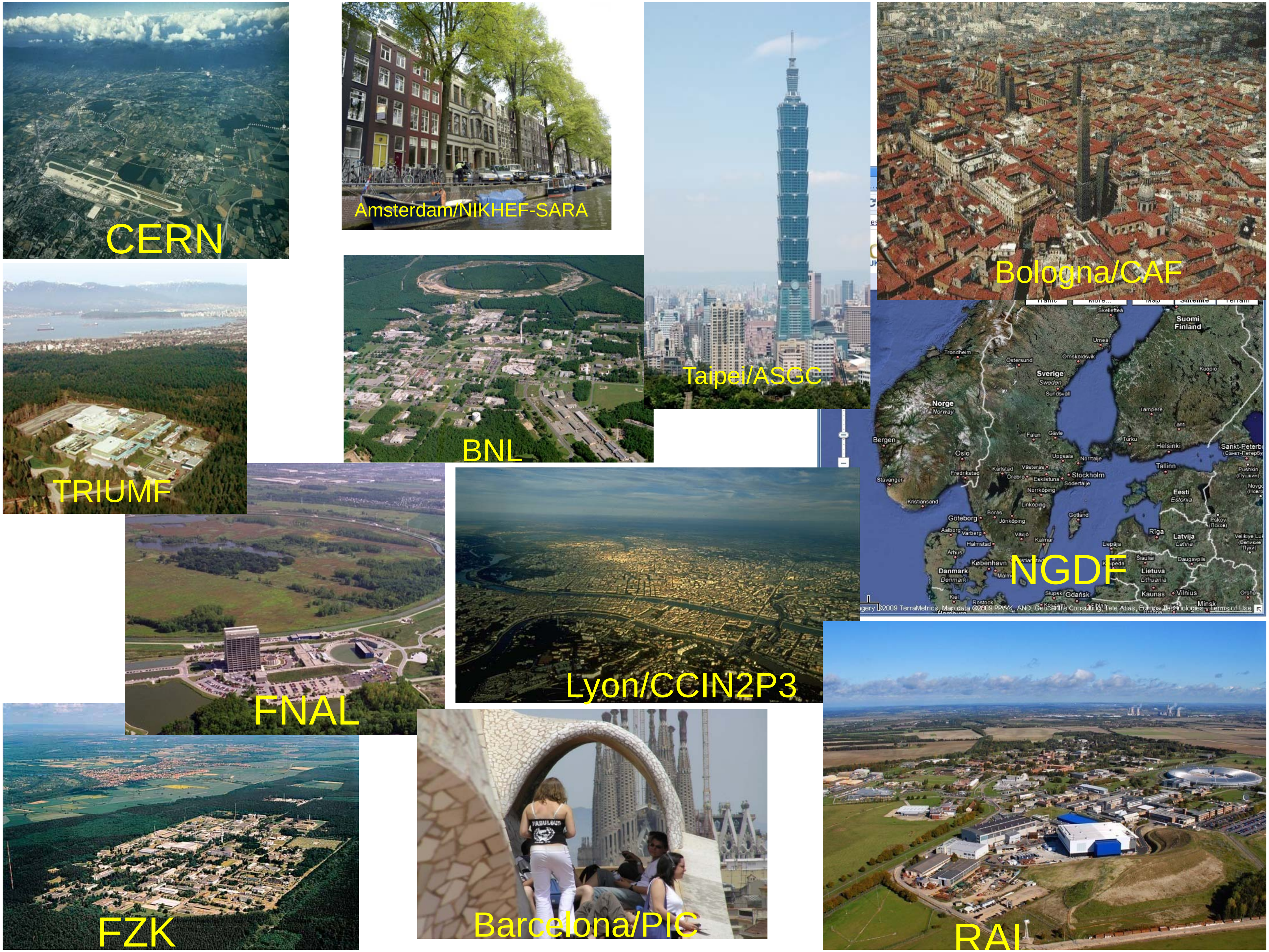
First events





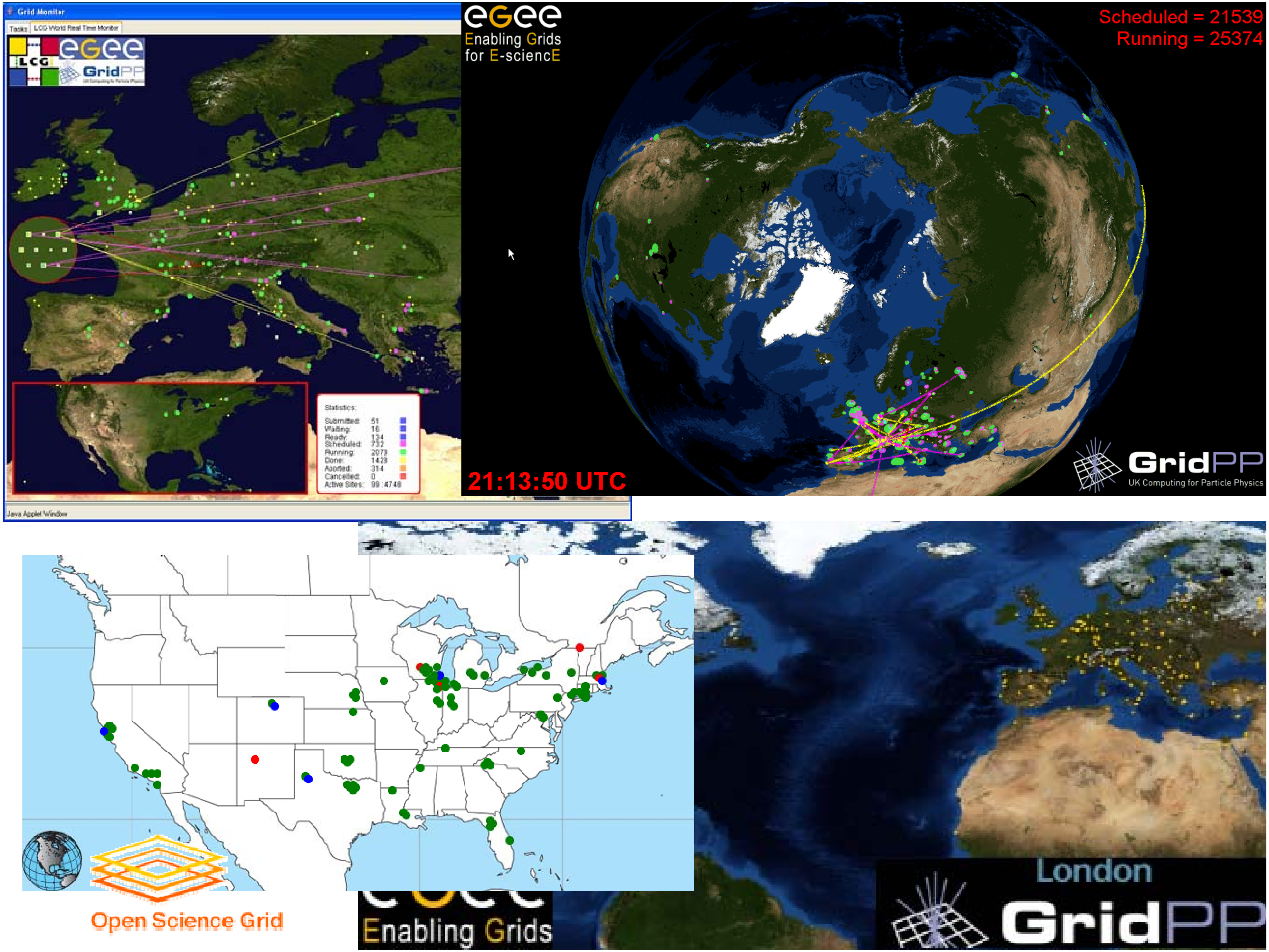






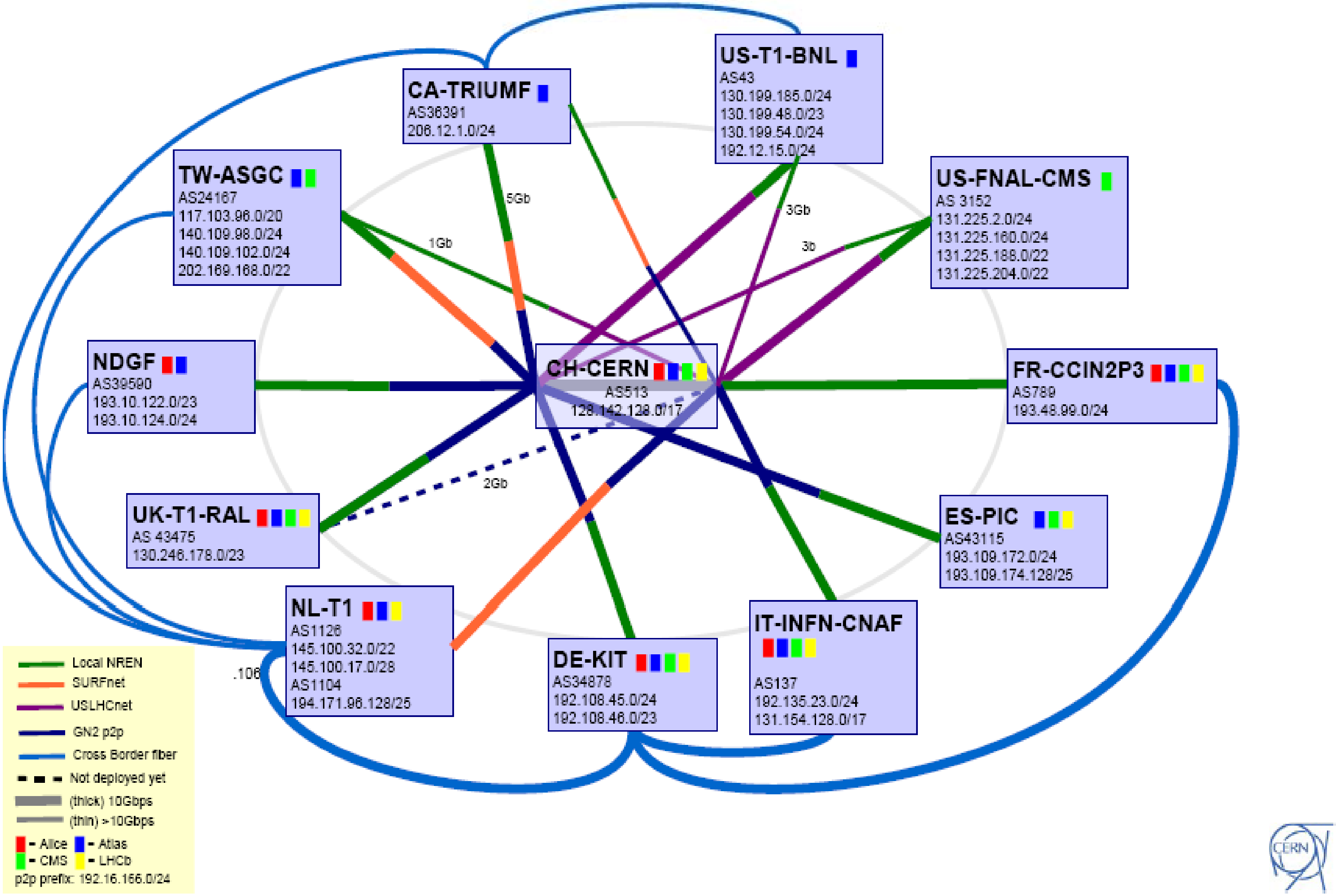
CERN
Amsterdam/NIKHEF-SARA
Taipei/ASGC
Bologna/CAF
TRIUMF
BNL
NGDF
FNAL
Lyon/CCIN2P3
FZK
Barcelona/PIC
RAL

CERN
Taipei/ASGC
Bologna/CAF
TRIUMF
BNL
NGDF
FNAL
Lyon/CCIN2P3
FZK
Barcelona/PIC
RAL

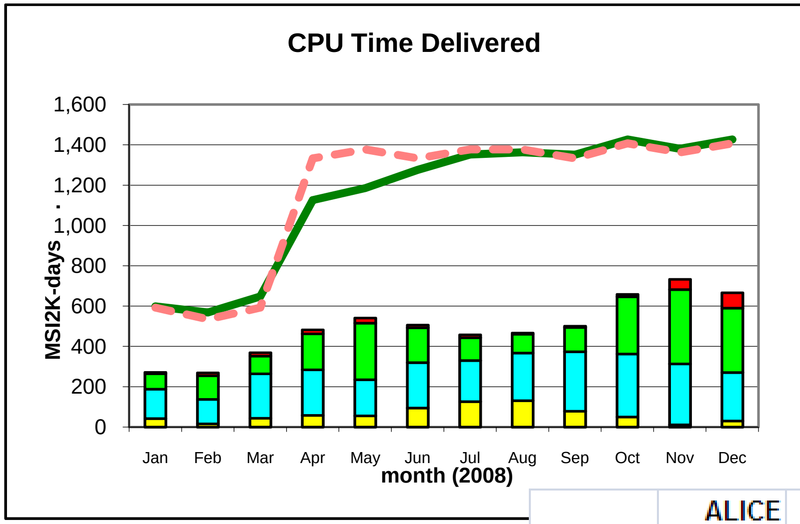

Grid Monitor
Tasks
LCG World Real Time Monitor
Statistics:
Submitted: 51
Waiting: 16
Ready: 134
Scheduled: 732
Running: 2073
Done: 1423
Aborted: 314
Cancelled: 0
Java Applet Window
egee
Enabling Grids
for E-sciencE
Scheduled = 21539
Running = 25374
21:13:50 UTC
GridPP
UK Computing for Particle Physics
Open Science Grid
Enabling Grids
London
GridPP

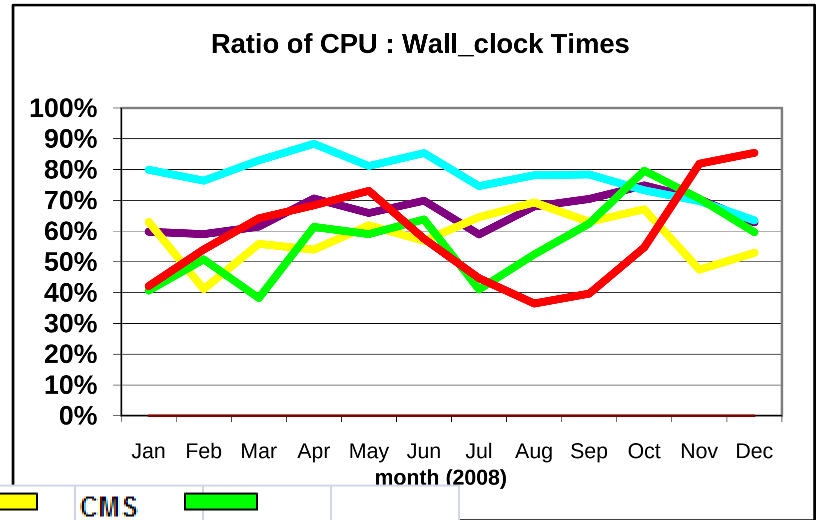
LHCOPN status

CERN IT Department

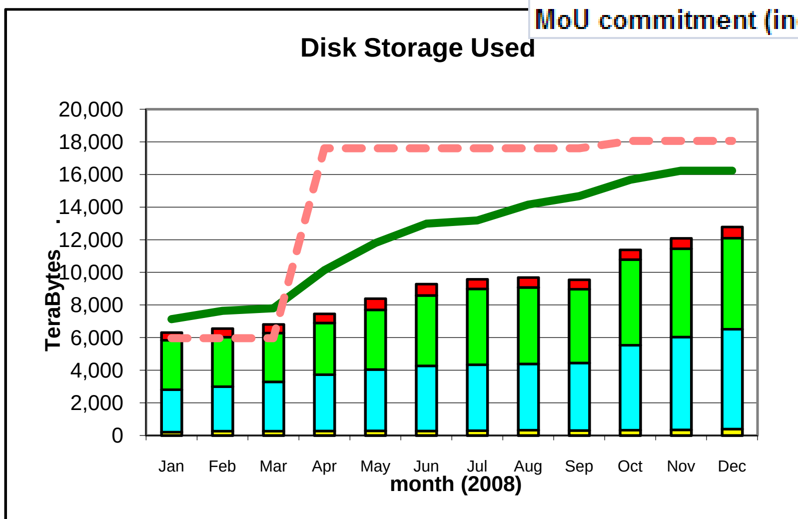


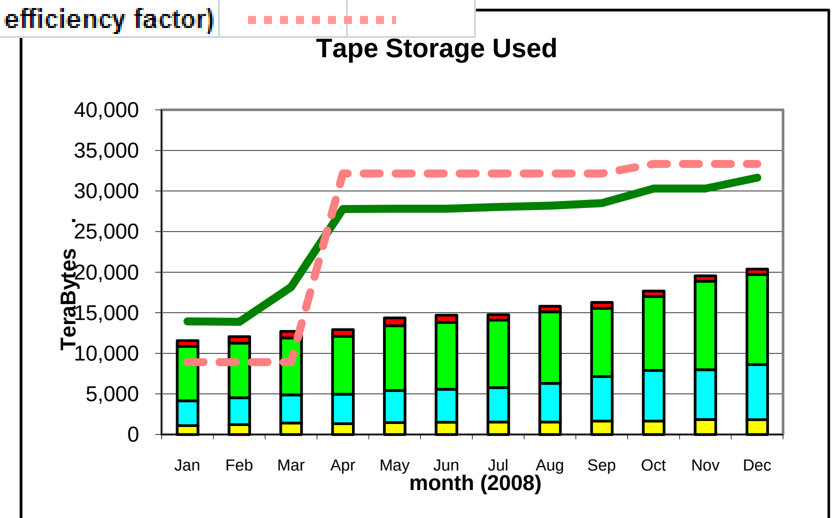
CERN + Tier 1 accounting - 2008





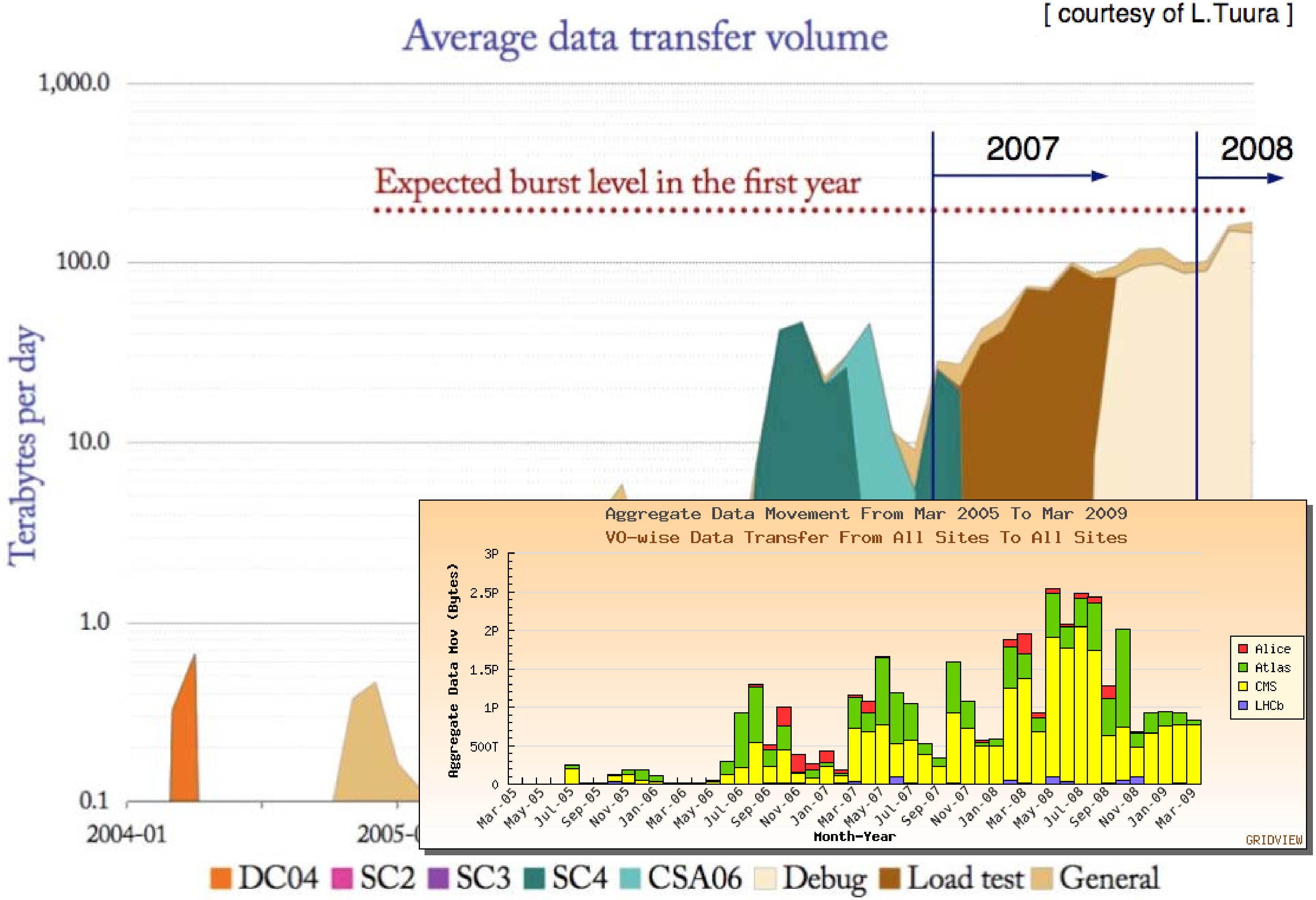
ALICE	(yellow)	CMS	(green)
ATLAS	(cyan)	LHCb	(red)
installed capacity (inc. efficiency factor)	(solid green line)		
MoU commitment (inc. efficiency factor)	(dashed pink line)		



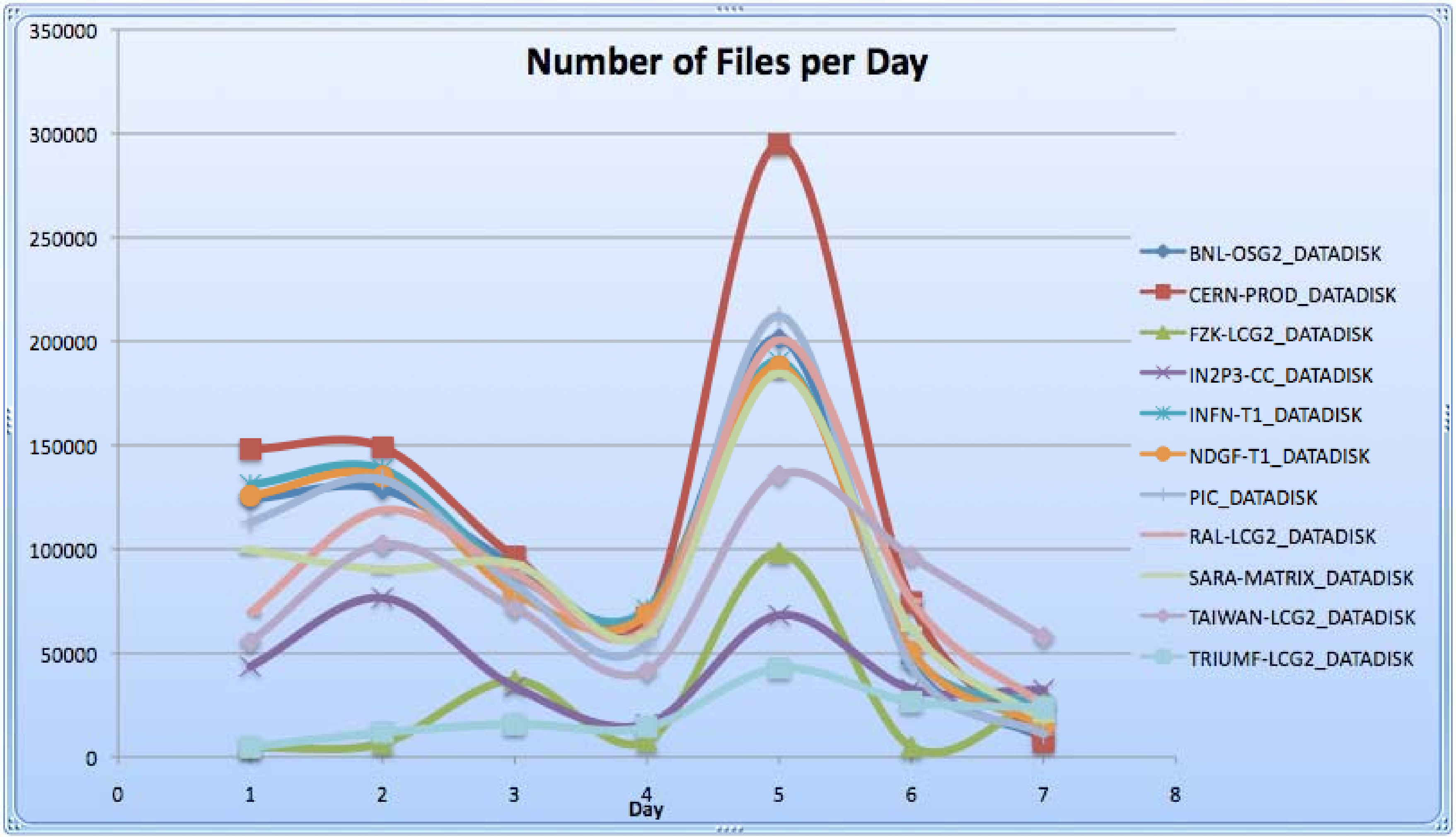


CMS Data Transfer History

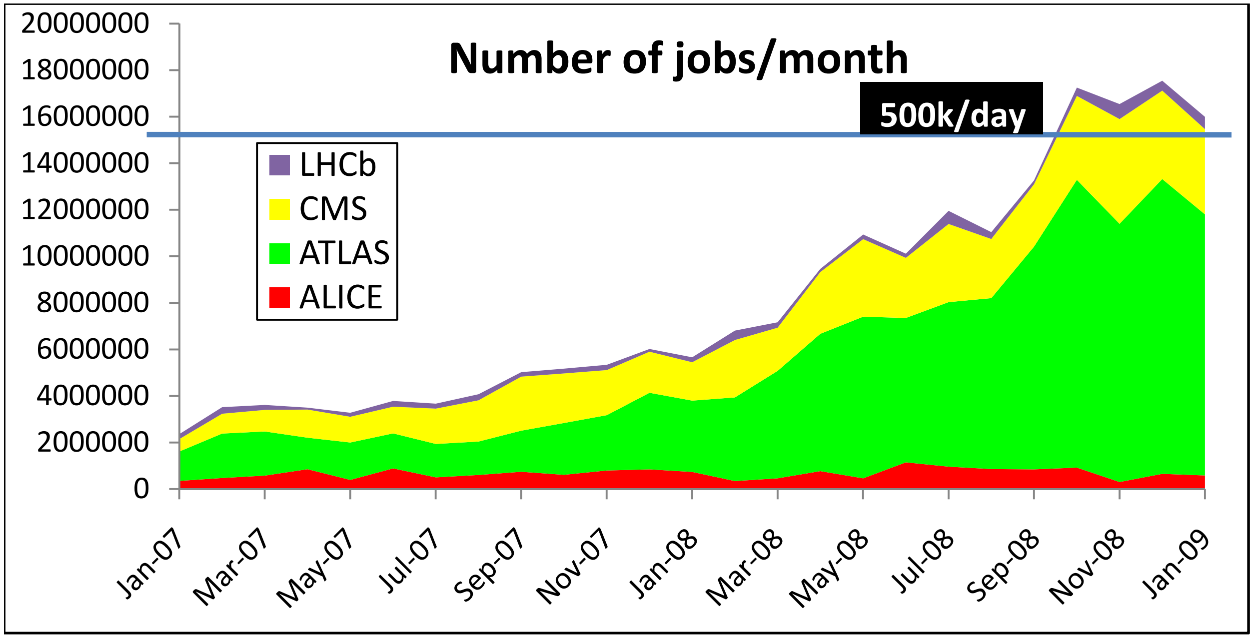
[courtesy of L.Tuura]



10M files Test @ ATLAS



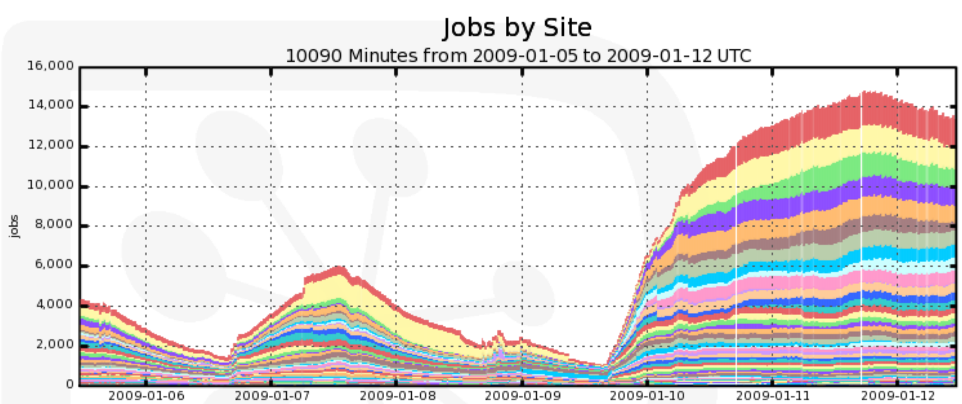
(From S. Campana)



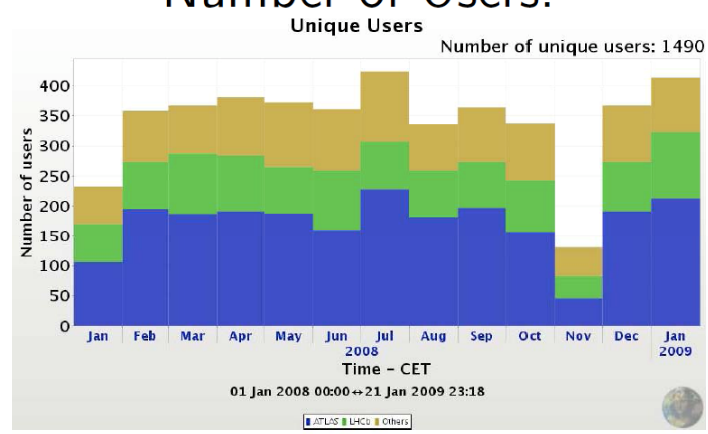
Main outstanding issues related to service/site reliability

	Alice	ATLAS	CMS	LHCb	Total	
Tier-1s	6.24	32.03	30.73	2.50	71.50	34.3%
Tier-2s	9.61	52.23	55.04	20.14	137.02	65.7%
Total	15.85	84.26	85.77	22.64	208.52	

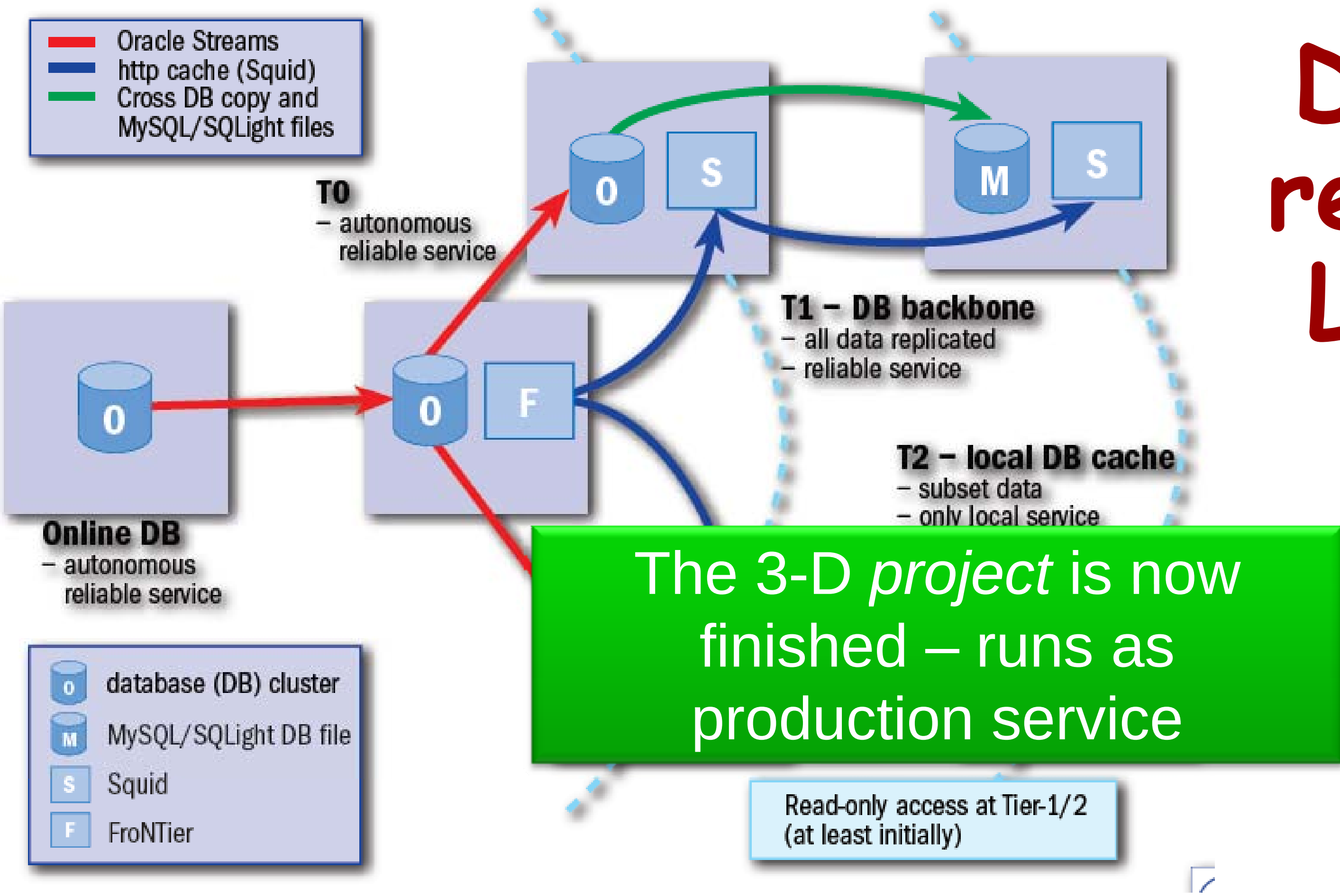
From APEL accounting portal for Aug.'08 to Jan.'09; #s in MSI2k



Number of Users:



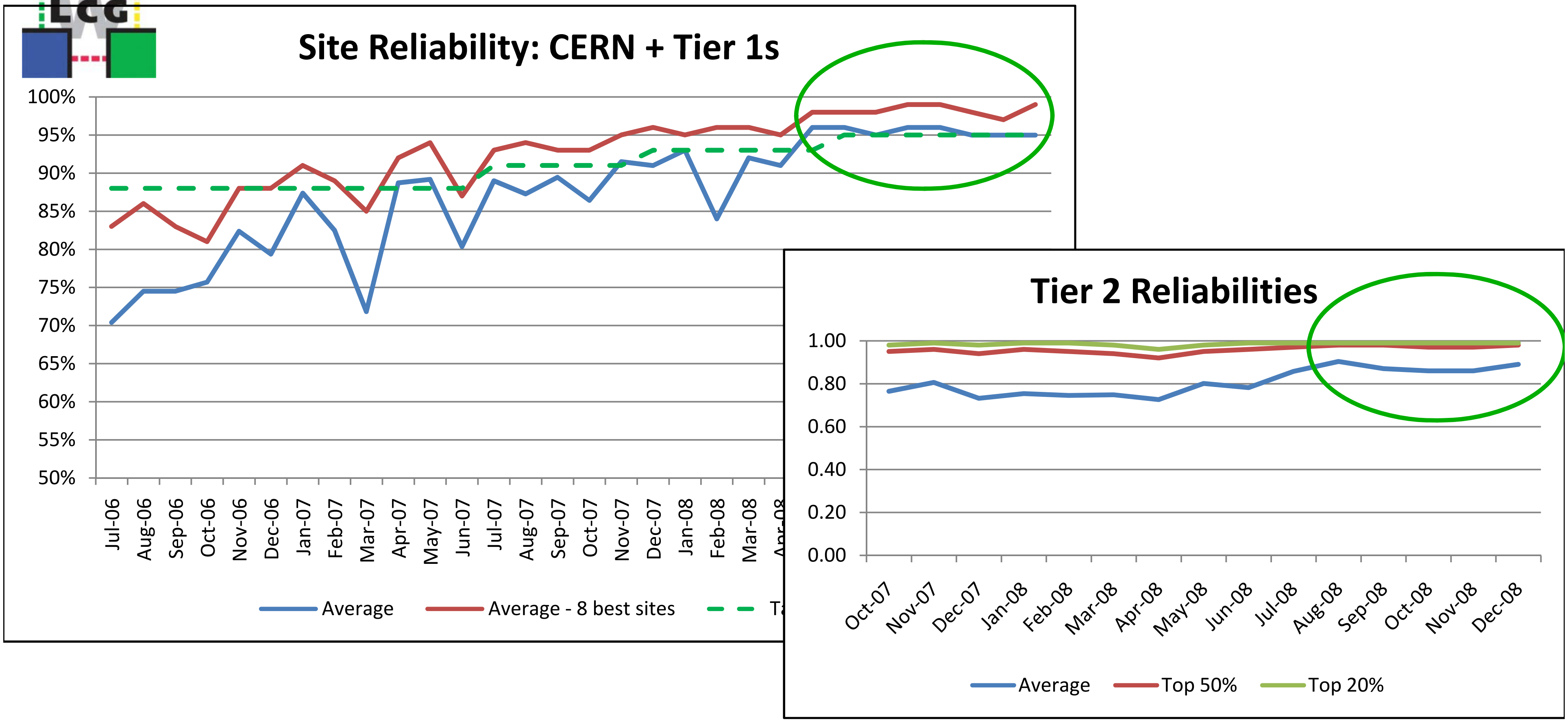
Database replication LCG 3-D



The 3-D *project* is now finished – runs as production service

- In full production
  - Several GB/day user data can be sustained to all Tier 1s
- ~100 DB nodes at CERN and several 10's of nodes at Tier 1 sites
  - Very large distributed database deployment
- Used for several applications
  - Experiment calibration data; replicating (central, read-only) file catalogues

Reliabilities

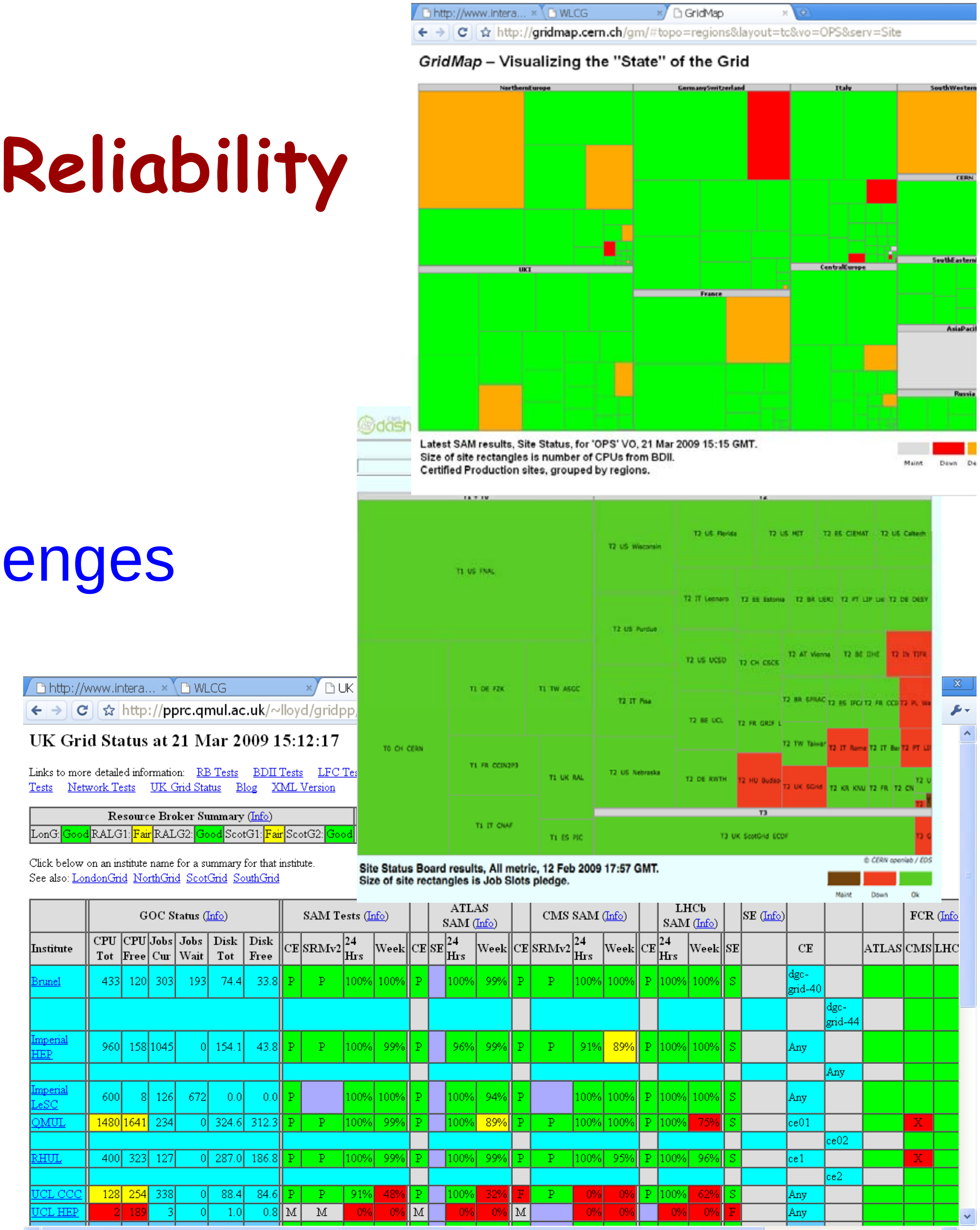


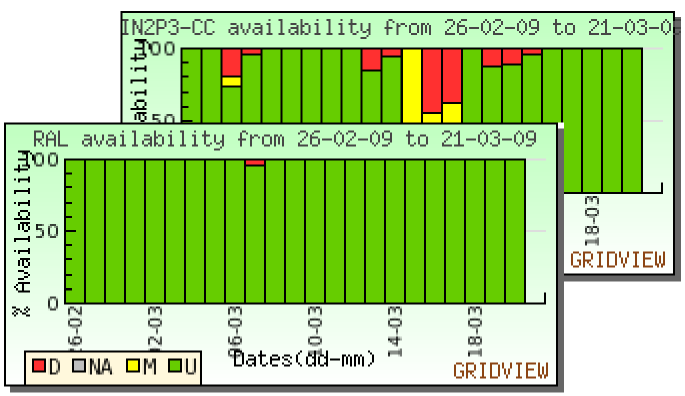
Improvement during CCRC and later is encouraging
-Tests do not show full picture – e.g. Hide experiment-specific issues,
- “OR” of service instances probably too simplistic

a) publish VO-specific tests regularly;
b) rethink algorithm for combining service instances

Improving Reliability

- Testing
- Task forces/challenges
- Monitoring
  - Appropriate
  - Followed up

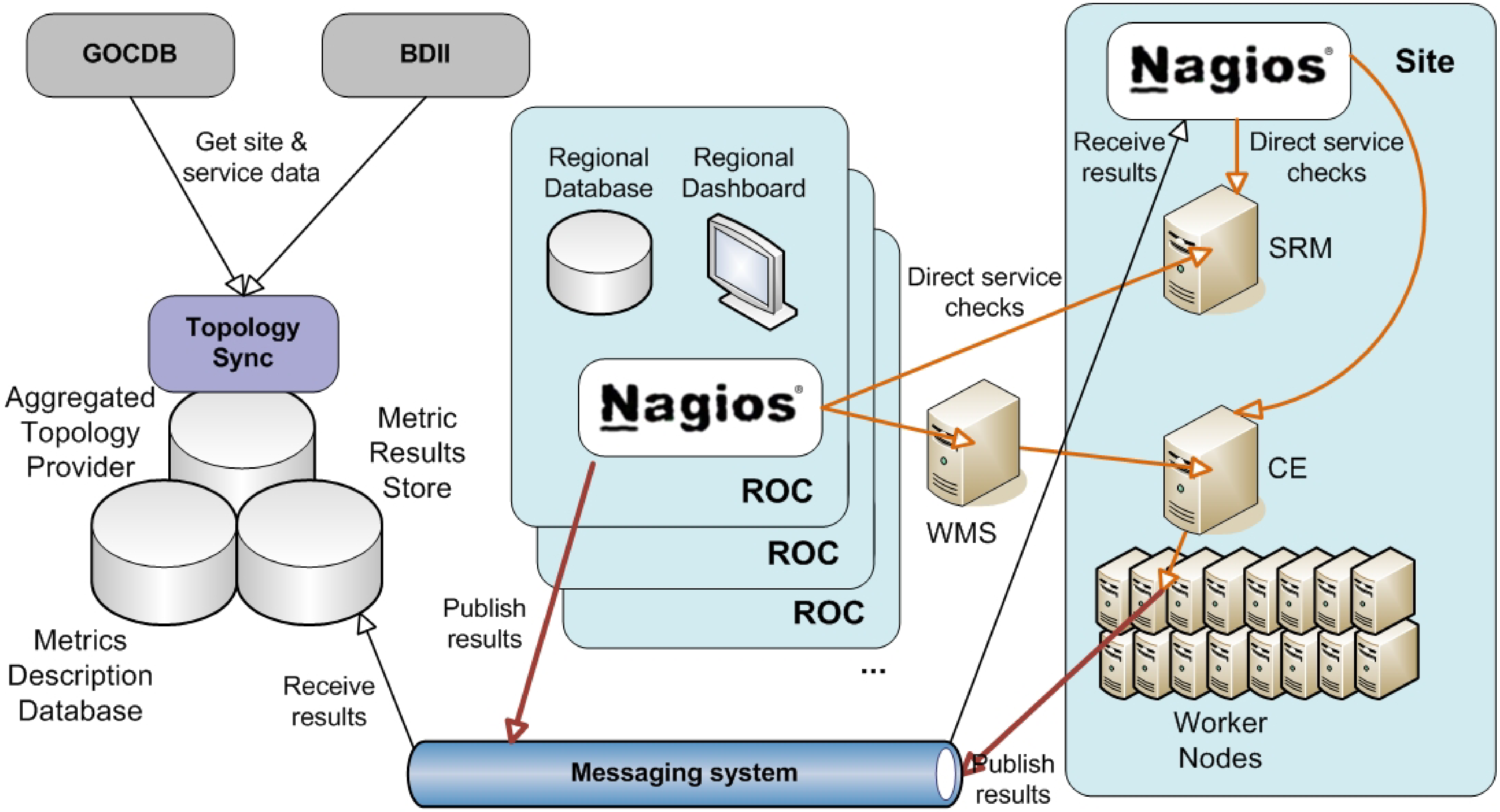




challenges

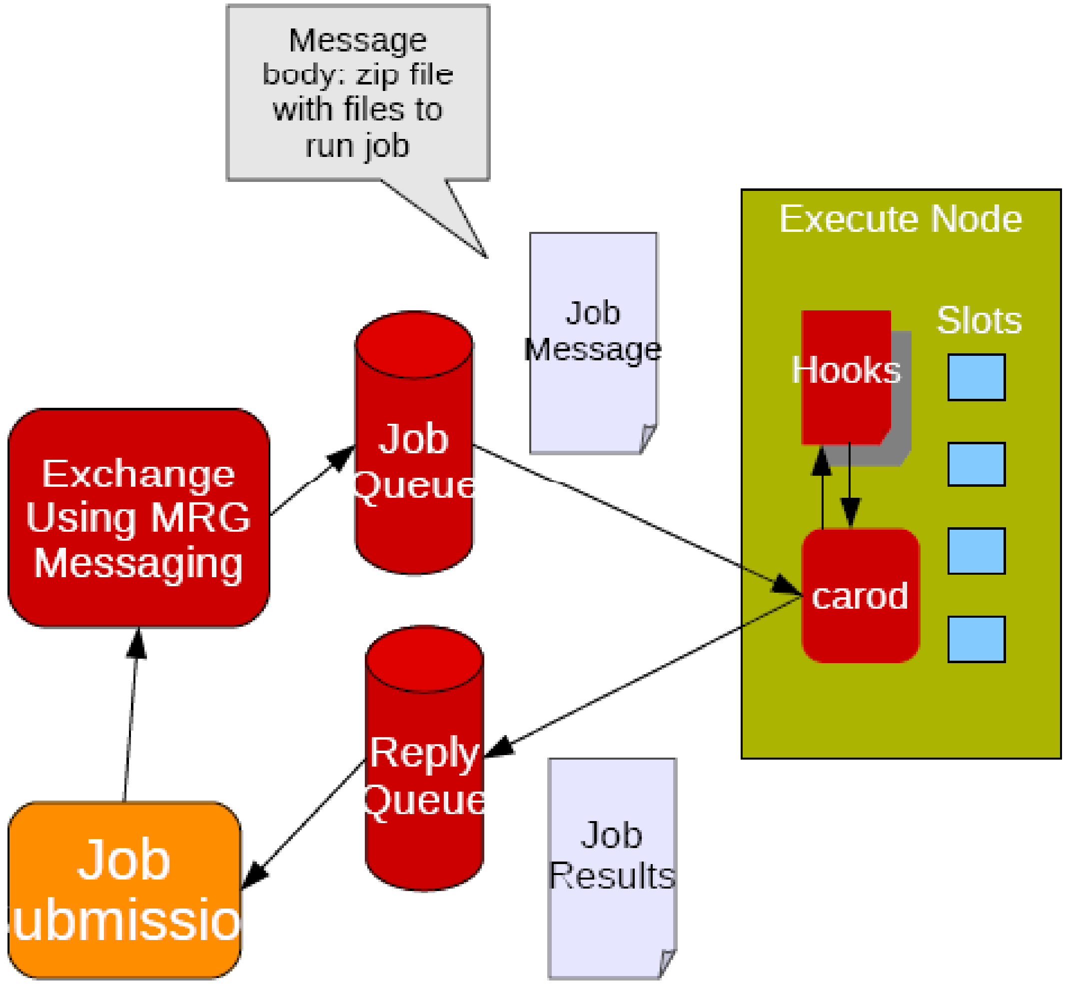
- Multi(many)-core:
  - Better memory use efficiency
  - Co-scheduling many similar processes onto a single box
  - Parallelizing (multi-thread, MPI, ...) the applications
- New technologies:
  - Clouds + virtualisation
    - Use as “overflow” resources for peak periods – demonstrated
    - Running our facilities as “clouds” – use of virtualisation, management tools, etc.
    - Buying resources directly ... needs education of the funding agencies
    - Lessons: simplicity of interfaces, usage patterns,...
    - Virtualisation – helps improve service reliability; simplifies facility management (tbc) and leaves apps to deal with dependencies
    - Grid → grid of “cloud-like” objects
  - Filesystems
    - Lustre, Hadoop, NFS4.1, etc
    - Can we use these to improve our service reliability? Usability?
  - Messaging systems
    - Use for integrating systems – Web services across languages etc did not deliver

What are the limitations & possible solutions?



Messaging Software Ecosystem Examples

- MRG Grid provides low latency scheduling via messaging
  - Useful pattern for other systems
- MRG/Qpid provides features people often build on top of messaging
  - XML Exchange, LVQ, Ring Queue, TTL, Federation, Management, etc.
- Open Source projects are building on AMQP Messaging
  - OpenIPA project is using AMQP Messaging for management and monitoring of Identity, Policy, Audit systems
  - LibVirt project is using AMQP messaging for management and monitoring
  - Wireshark supports AMQP



Challenges cont...

- Simplification of data management
  - Clouds don't help much here
  - Abstraction –SRM – has added complexity
  - How much is required? How can we simplify?
  - What are the lessons to learn?
  - Database access – grid authn/authz would help – ...
- New Tier 0 centre
  - We will run out of power, new centre planned, will it be ready when we need it???
- Moving from EGEE to a European sustainable grid infrastructure
  - Whilst maintaining a solid service

EGEE → EGI+NGIs

- EGI blueprint published in December; endorsed in January by 20 NGIs
- March – policy board has selected Amsterdam as the location of EGI.org (body resp. for managing EGI)
- Initiation of transition process to create an EGI council
  - MoU to be prepared as an interim measure to identify NGIs prepared to commit as described in the blueprint (start with Letters of Intent)
  - Anticipate 1st council meeting in May
- Task force to be established for preparation of EGI proposals, for EC calls anticipated to close in November
- EGEE has outlined a fairly detailed transition plan for the final year of the project
  - But can only go so far

EGI and WLCG

- WLCG cannot take the risk of assuming EGI will be in place at the end of EGEE-III
- We plan to ensure that services provided to use to day by EGEE are assured by our Tier 1 sites
  - Support the formation of a gLite consortium to support the middleware
- In parallel we work with EGI_DS and EC to try and ensure that EGI and the NGIs will deliver what we need

Conclusions

- We have built a working system that will be used for first data taking
  - But it has taken a lot longer than anticipated ... and was a lot harder ... and the reality of grids does not quite match the hype ...
- We now have an opportunity to rethink how we want this to develop in the future
  - Clearer ideas of what is needed
  - And must consider the risks, maintainability, reliability, and complexity
- Change of funding model and new technologies provide opportunities
- Challenges: data management and reliability, reliability, ...
- Should remember ... Our goal is to enable the experiments' computing, not necessarily to develop computer science (unless we have to ...)

WLCG timeline 2009-2010