LCG Status & Progress

openlab Board of Sponsors

25th April 2008

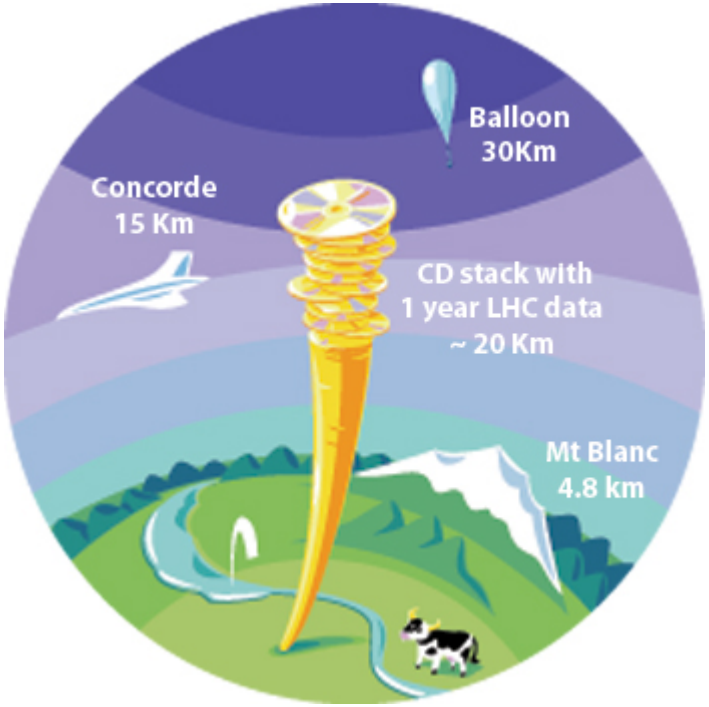
Ian Bird
LCG Project Leader

CERN openlab

The LHC Data Challenge

CERN IT Department

- The accelerator will be completed in 2008 and run for 10-15 years
- Experiments will produce about **15 Million Gigabytes** of data each year (about 20 million CDs)
- LHC data analysis requires a computing power equivalent to **~100,000 of today's fastest PC processors**
- Requires many cooperating computer centres, as CERN can only provide ~20% of the capacity



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Solution: the Grid

CERN IT Department

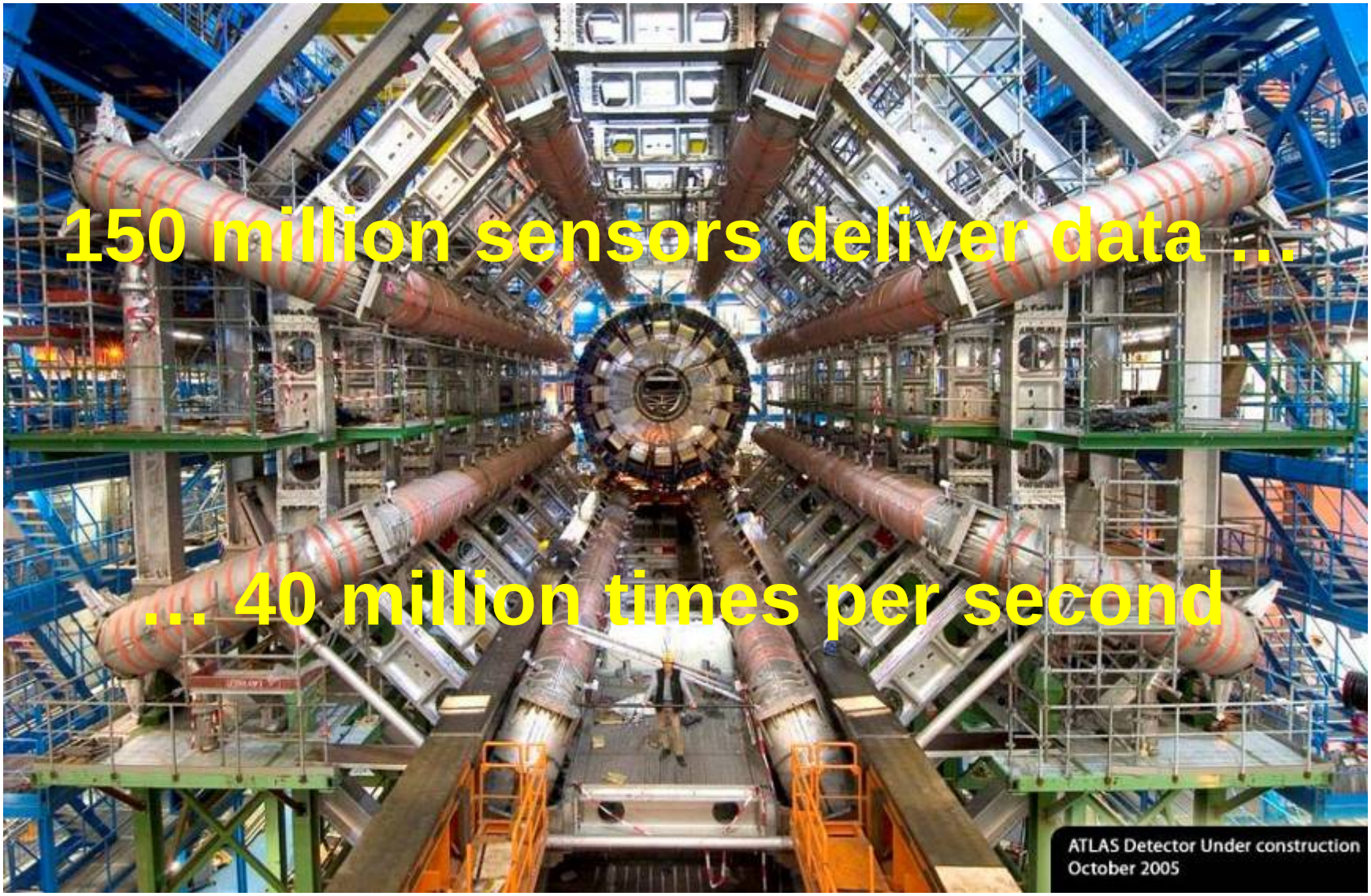
- Use the Grid to unite computing resources of particle physics institutions around the world

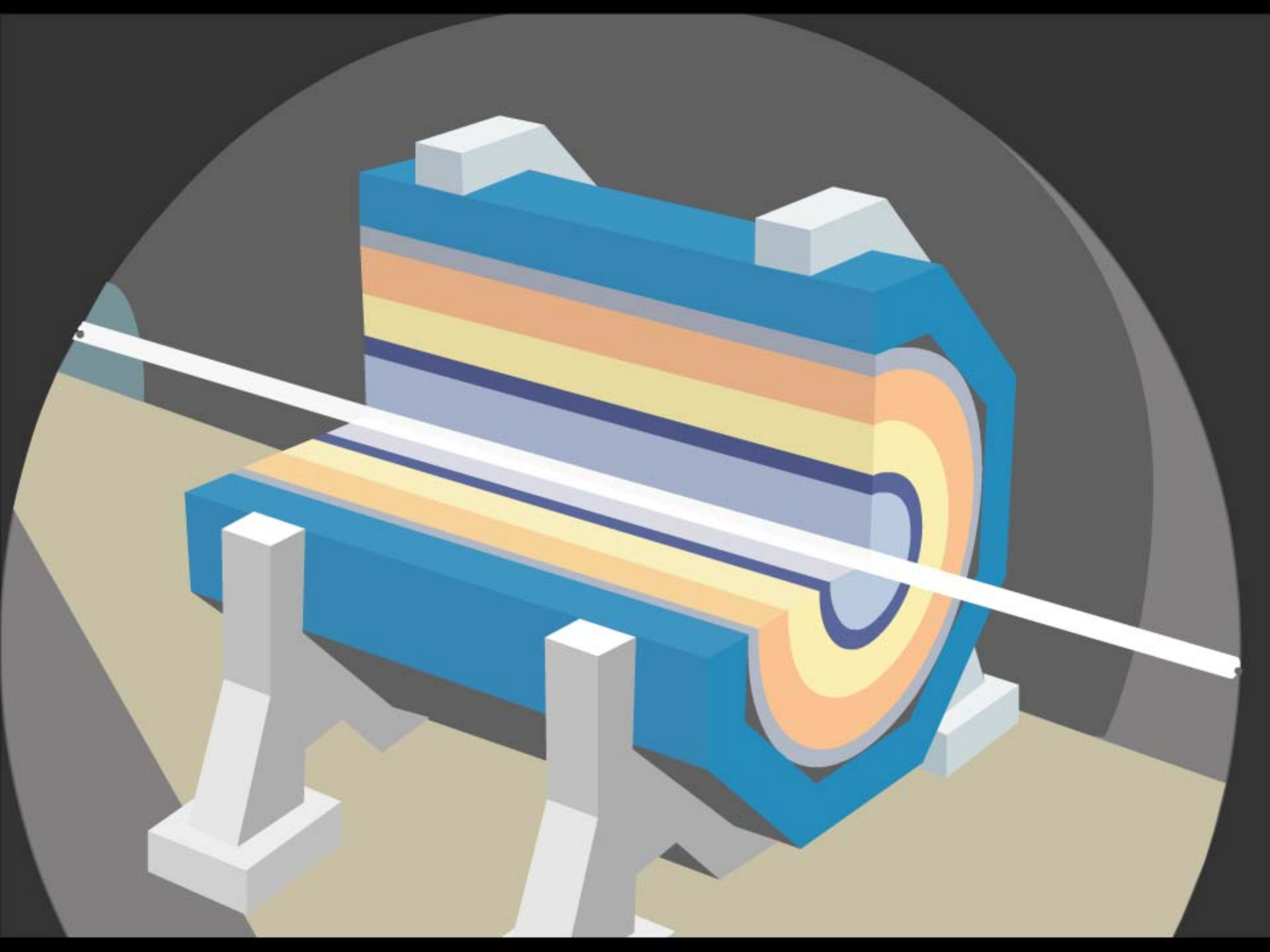
The **World Wide Web** provides seamless access to information that is stored in many millions of different geographical locations

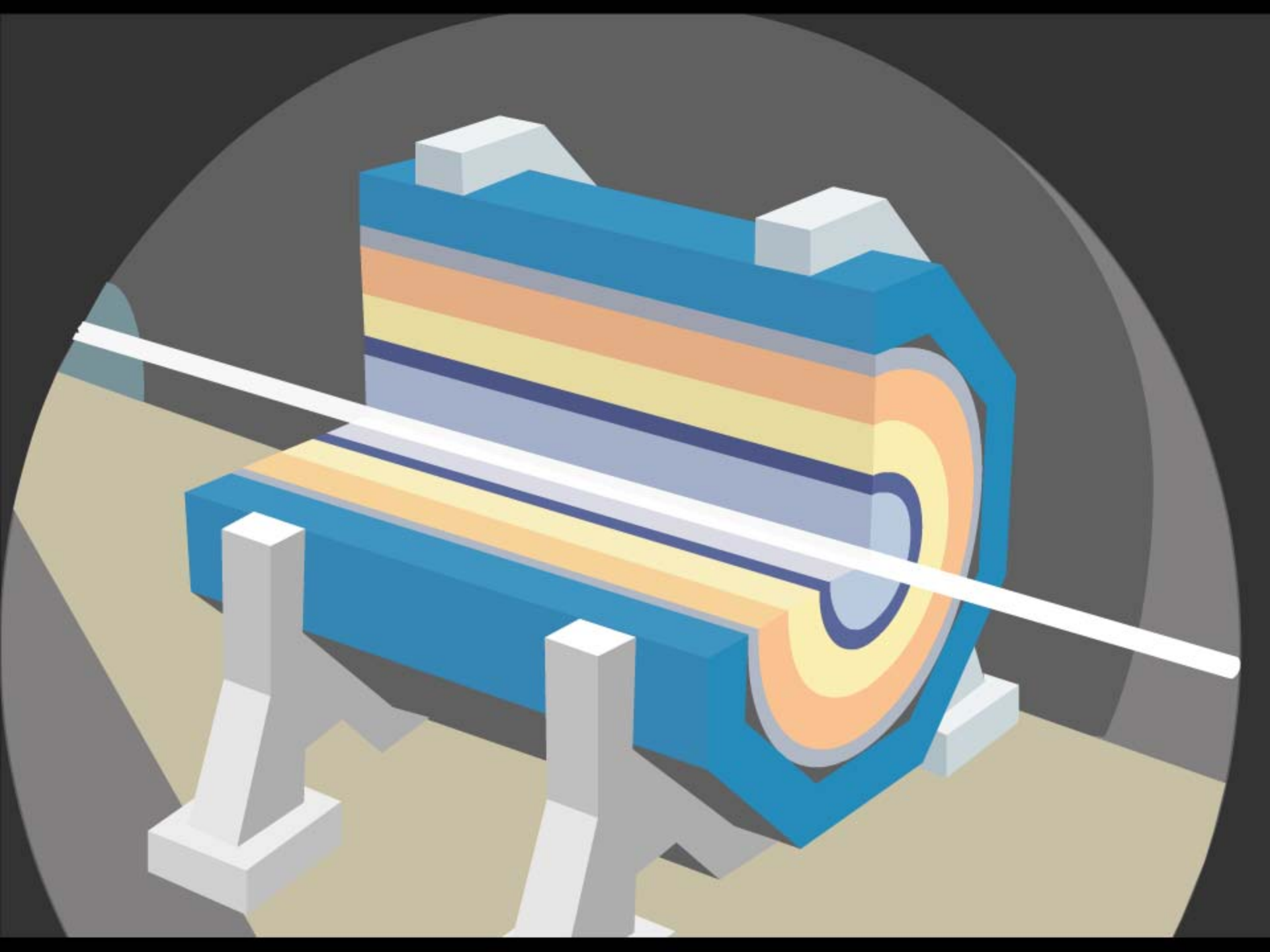
The **Grid** is an infrastructure that provides seamless access to computing power and data storage capacity distributed over the globe

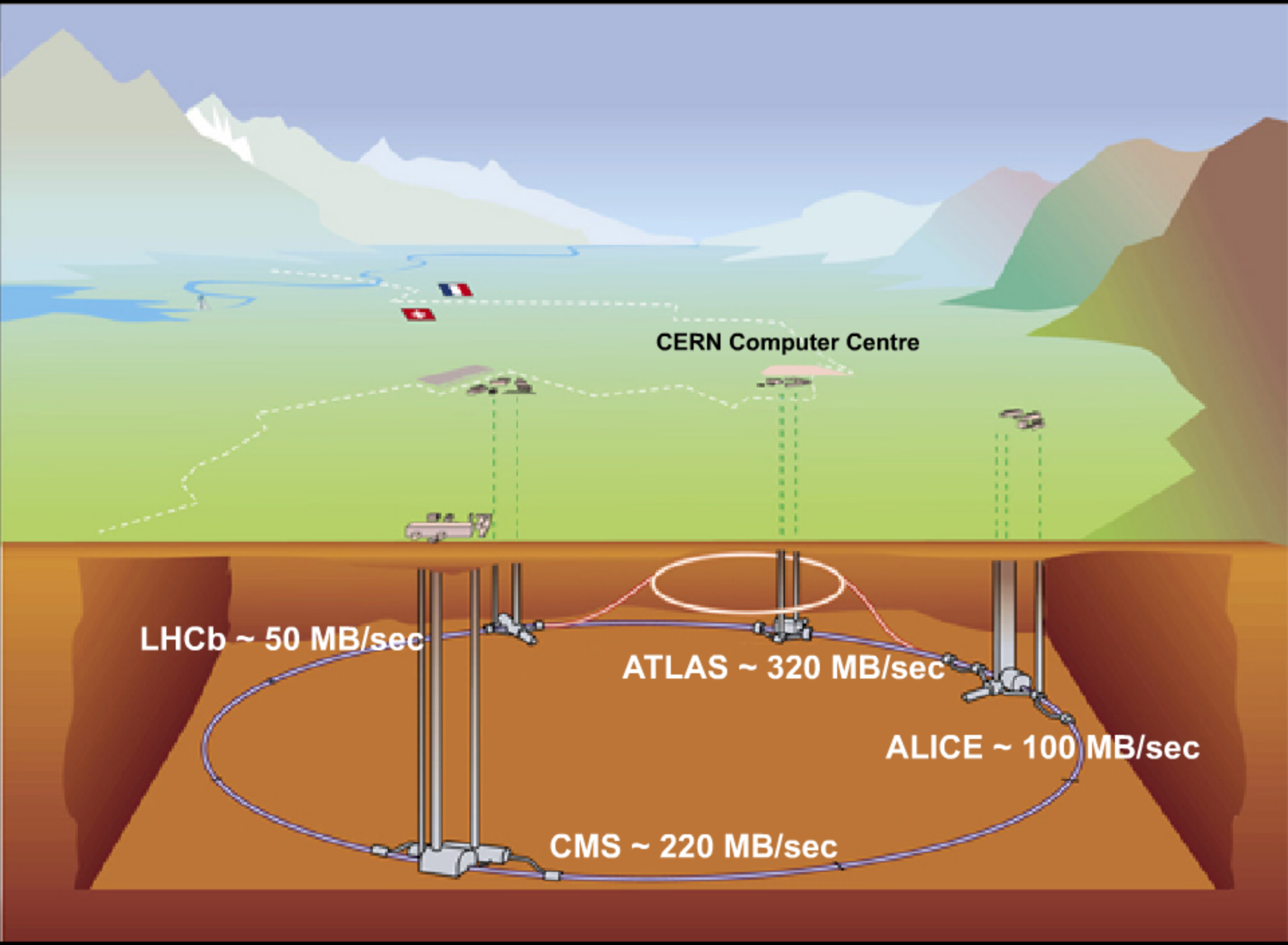
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

View of the ATLAS detector (under construction)



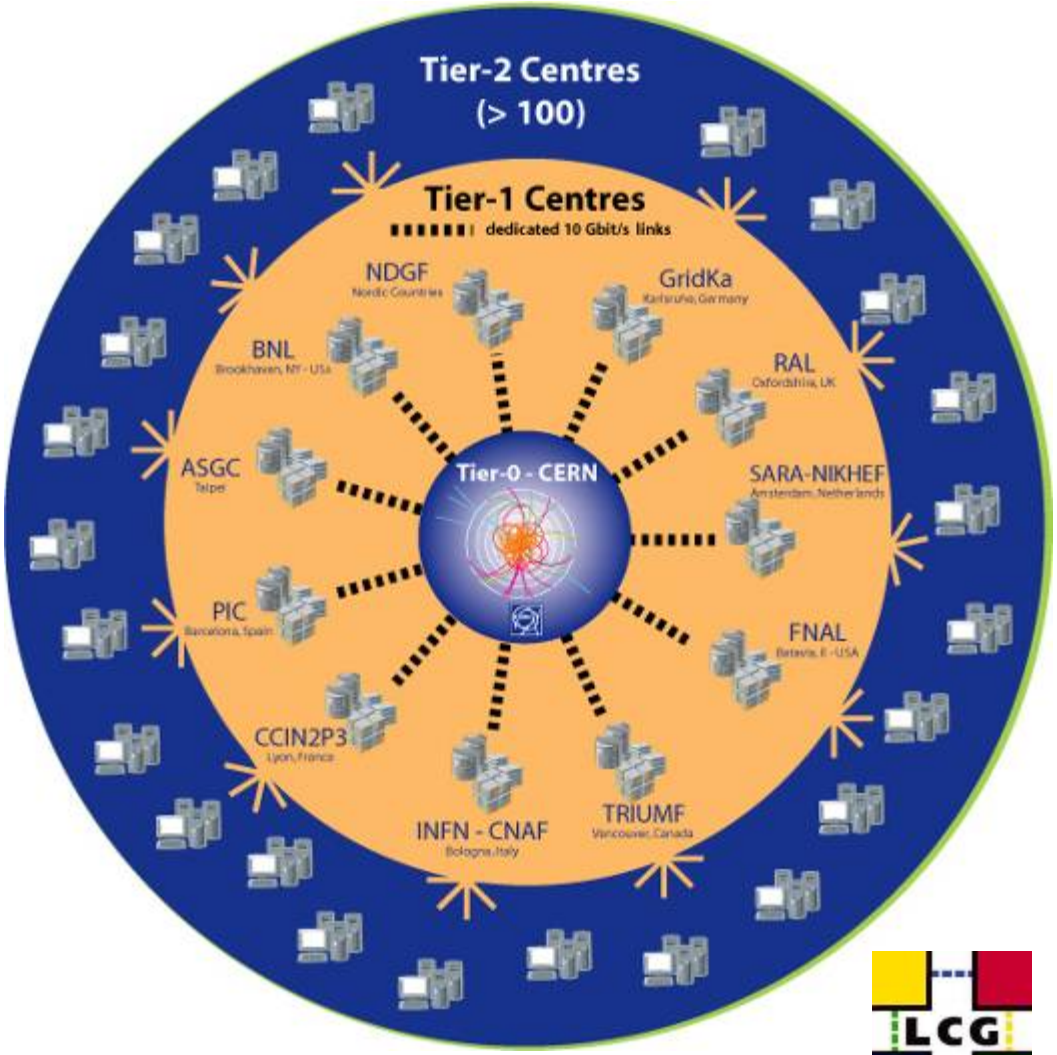





CERN Computer Centre
LHCb ~ 50 MB/sec
ATLAS ~ 320 MB/sec
ALICE ~ 100 MB/sec
CMS ~ 220 MB/sec

LHC Computing Grid project (LCG)

- More than 140 computing centres
- 12 large centres for primary data management: CERN (Tier-0) and eleven Tier-1s
- 38 federations of smaller Tier-2 centres
- 35 countries involved



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

WLCG Collaboration

CERN IT Department

- The Collaboration
  - 4 LHC experiments
  - ~140 computing centres
  - 12 large centres (Tier-0, Tier-1)
  - 38 federations of smaller "Tier-2" centres
  - ~35 countries
- Memorandum of Understanding
  - Agreed in October 2005, now being signed
- Resources
  - Focuses on the needs of the four LHC experiments
  - Commits resources
    - each October for the coming year
    - 5-year forward look
  - Agrees on standards and procedures
- Relies on EGEE and OSG (and other regional efforts)

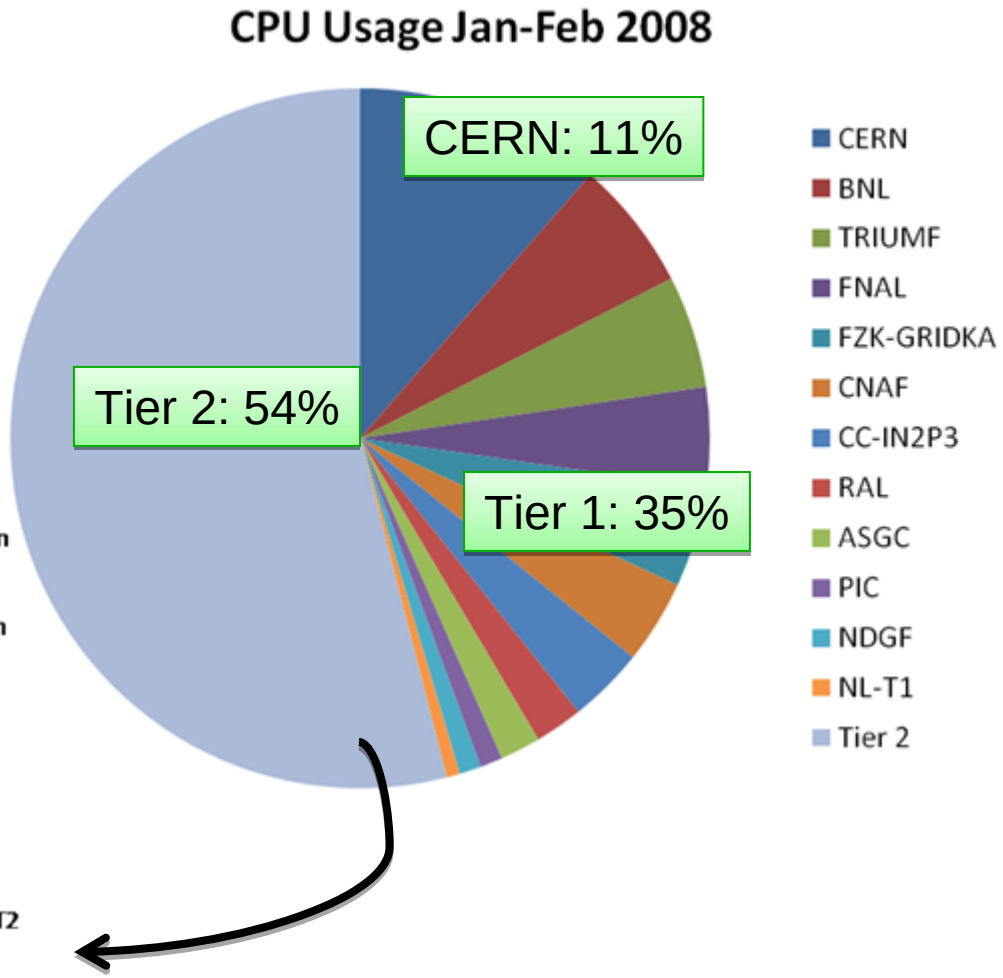
Open Science Grid

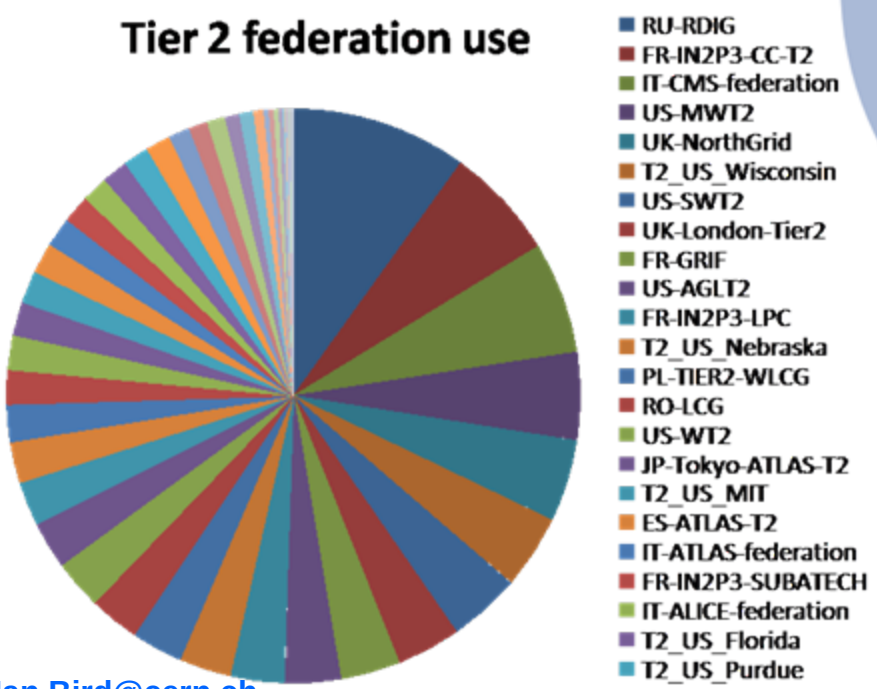
eGee Enabling Grids for E-sciencE

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

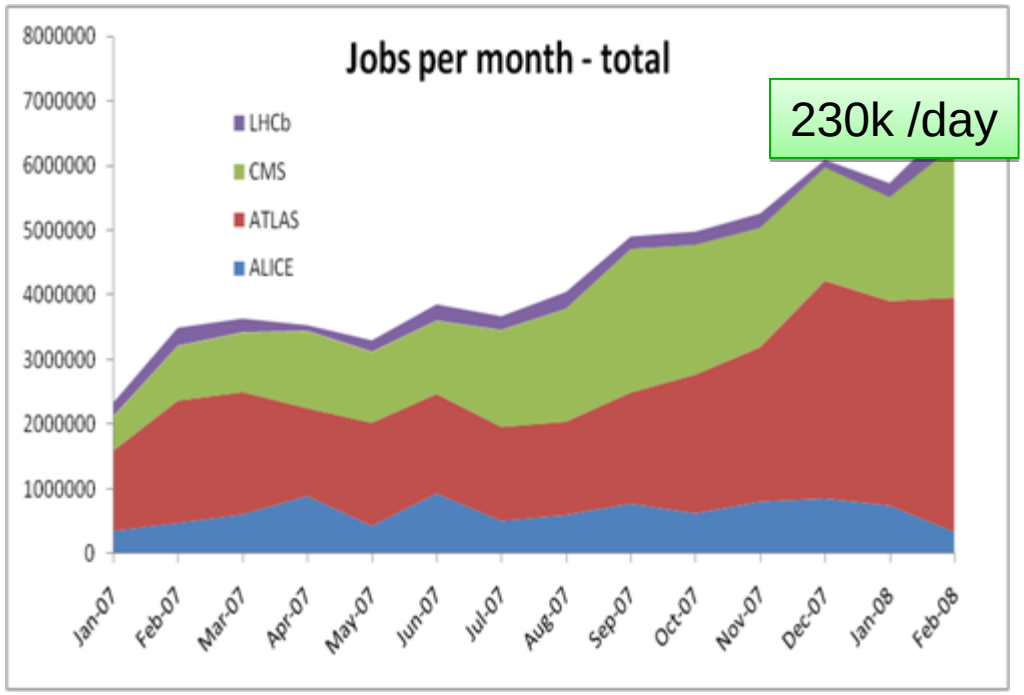
Recent grid use

- Across all grid infrastructures
- Preparation for, and execution of CCRC'08 phase 1
  - Move of simulations to Tier 2s





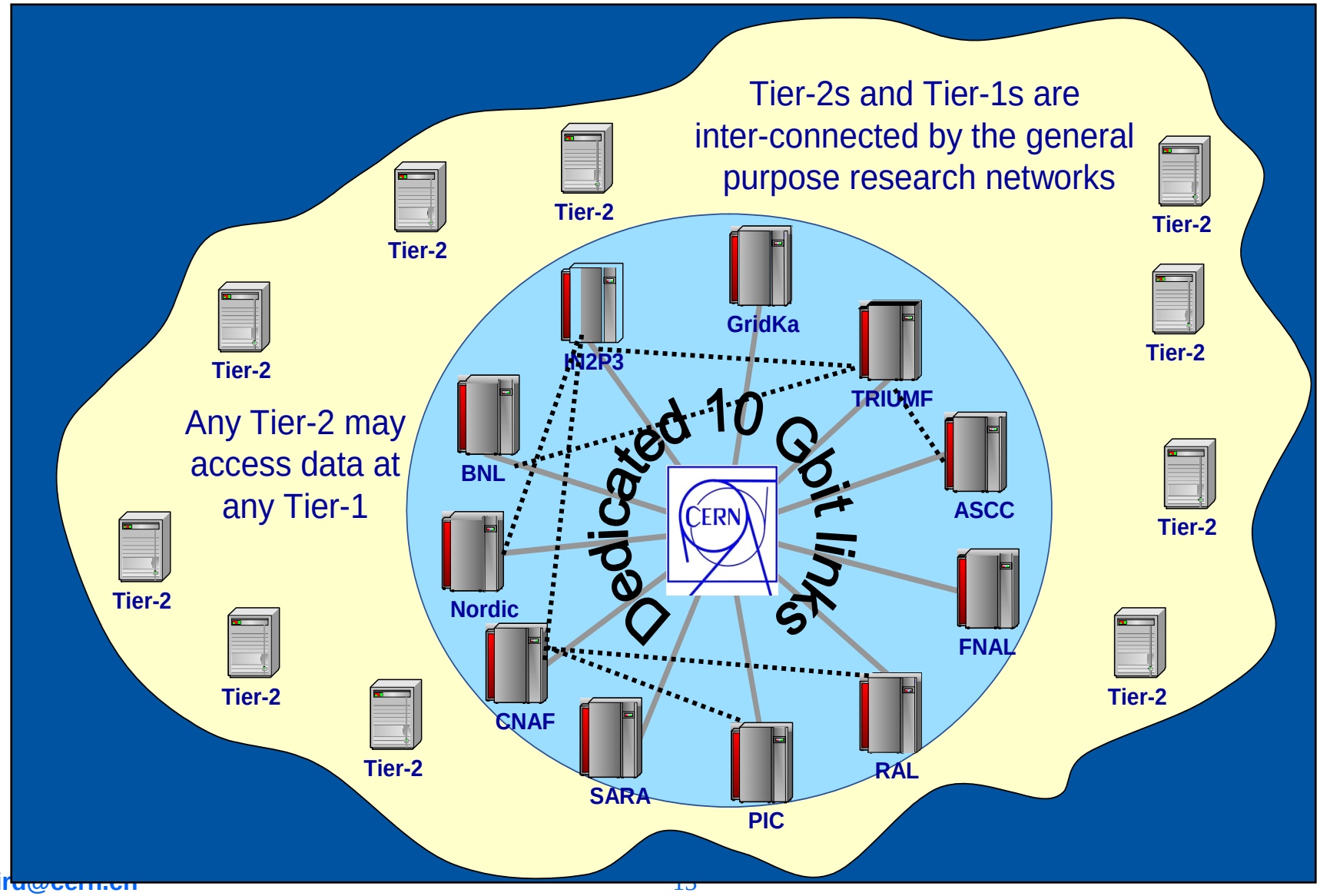
Recent grid activity



- WLCG ran ~ 44 M jobs in 2007 – workload has continued to increase – now at ~ 250k jobs/day
- Distribution of work across Tier0/Tier1/Tier 2 really illustrates the importance of the grid system
  - Tier 2 contribution is around 50%; > 85% is external to CERN

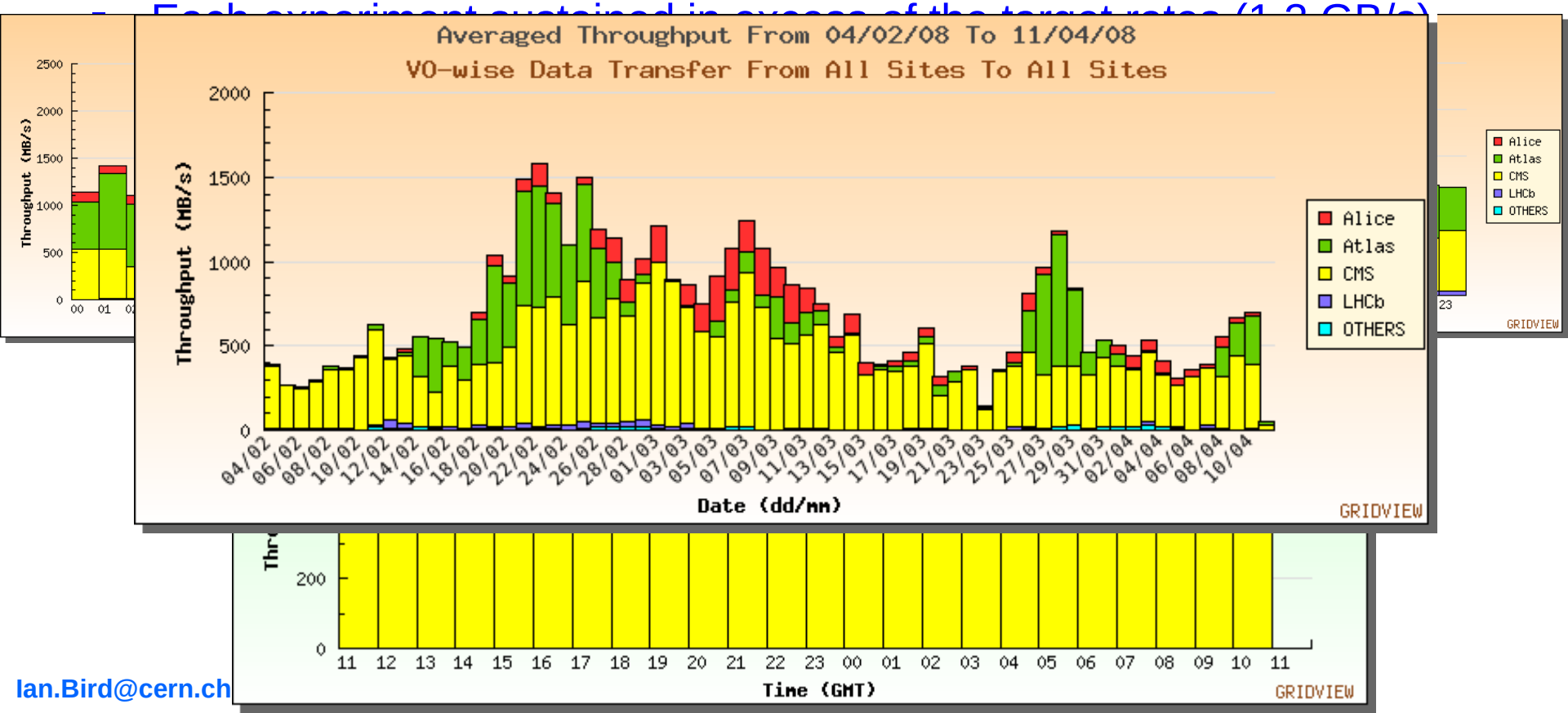
- These workloads (reported across all WLCG centres) are at the level anticipated for 2008 data taking

LHC OPN


Tier-2s and Tier-1s are inter-connected by the general purpose research networks
Any Tier-2 may access data at any Tier-1
Dedicated 10 Gbit links
CERN
GridKa
IN2P3
TRIUMF
BNL
ASCC
Nordic
FNAL
CNAF
RAL
SARA
PIC
Tier-2

Data transfer

- Data distribution from CERN to Tier-1 sites
  - The target rate was achieved in 2006 under test conditions
  - Autumn 2007 & CCRC'08 under more realistic experiment testing, reaching & sustaining target rate with ATLAS and CMS active
- Each experiment sustained in excess of the target rates (1.2 GB/s)



Ian.Bird@cern.ch

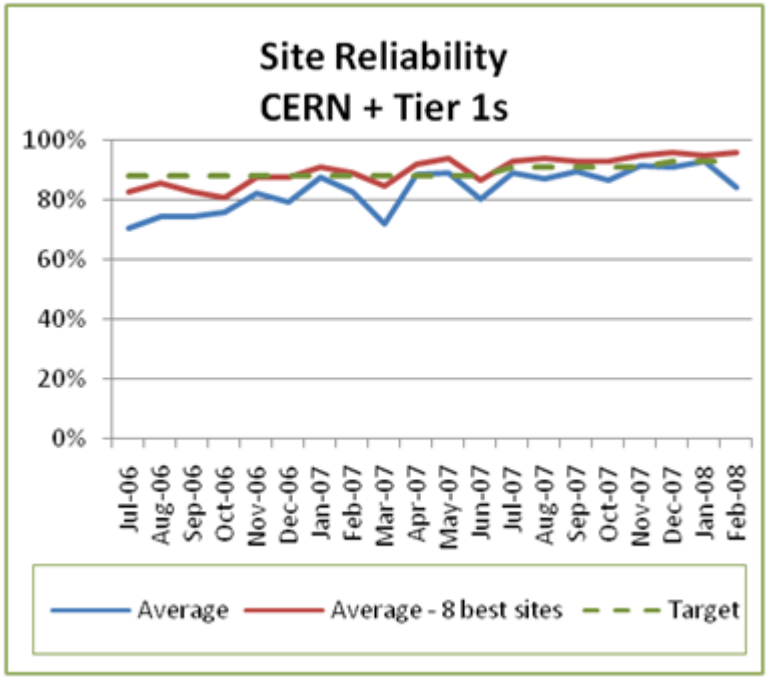
Combined Computing Readiness Challenge – CCRC'08

- Objective was to show that we can run together (4 experiments, all sites) at 2008 production scale:
  - All functions, from DAQ ⇨Tier 0 ⇔Tier 1s ⇔Tier 2s
- Two challenge phases were foreseen:
  1. **Feb:** not all 2008 resources in place – still adapting to new versions of some services (e.g. SRM) & experiment s/w
  2. **May:** all 2008 resources in place – full 2008 workload, all aspects of experiments' production chains
- Agreed on specific targets and metrics – helped integrate different aspects of the service
  - ❑ Explicit "**scaling factors**" set by the experiments for each functional block (e.g. data rates, # jobs, etc.)
  - ❑ Targets for "**critical services**" defined by experiments – essential for production, with analysis of impact of service degradation / interruption
  - ❑ WLCG "**MoU targets**" – services to be provided by sites, target availability, time to intervene / resolve problems …

CCRC'08 (Feb) – results

- Preparation:
  - Focus on understanding missing and / or weak aspects of the service and in identifying pragmatic solutions
  - Main outstanding problems in the middleware were fixed (just) in time and many sites upgraded to these versions
  - The deployment, configuration and usage of SRM v2.2 went better than had predicted, with a noticeable improvement during the month
- Despite the high workload, we also demonstrated (most importantly) that we can support this work with the available manpower, although essentially no remaining effort for longer-term work
- If we can do the same in May – when the bar is placed much higher – we will be in a good position for this year's data taking
- However, there are certainly significant concerns around the available manpower at all sites – not only today, but also in the longer term, when funding is unclear

Tier 0/Tier 1 Site reliability



- Target:
  - Sites 91% & 93% from December
  - 8 best: 93% and 95% from December
- See QR for full status

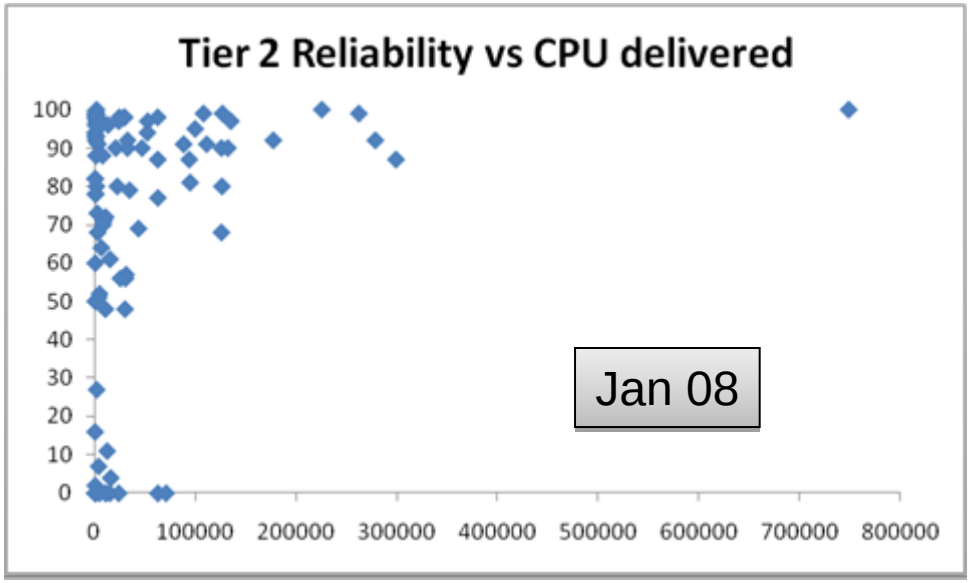
	Sep 07	Oct 07	Nov 07	Dec 07	Jan 08	Feb 08
All	89%	86%	92%	87%	89%	84%
8 best	93%	93%	95%	95%	95%	96%
Above target (+>90% target)	7 + 2	5 + 4	9 + 2	6 + 4	7 + 3	7 + 3

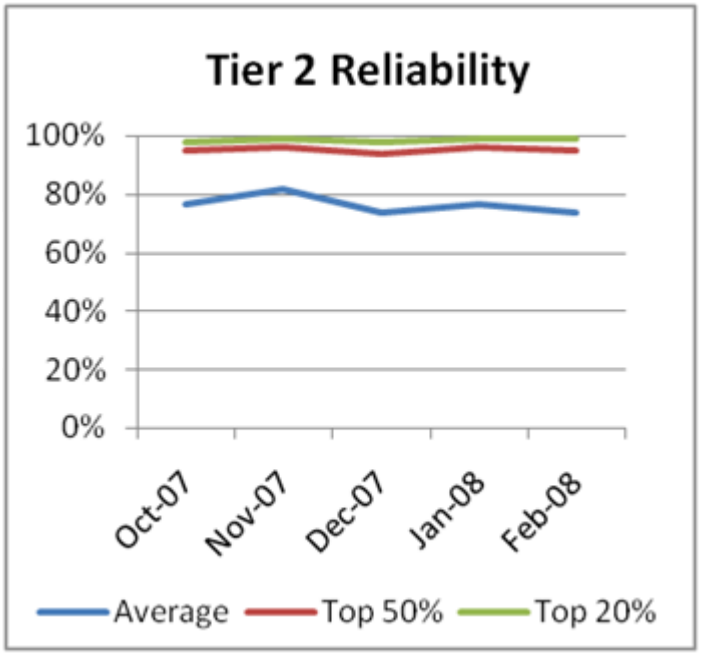
Tier 2 Reliabilities

- Reliabilities published regularly since October

Overall	Top 50%	Top 20%	Sites
76%	95%	99%	89→100

- In February 47 sites had > 90% reliability





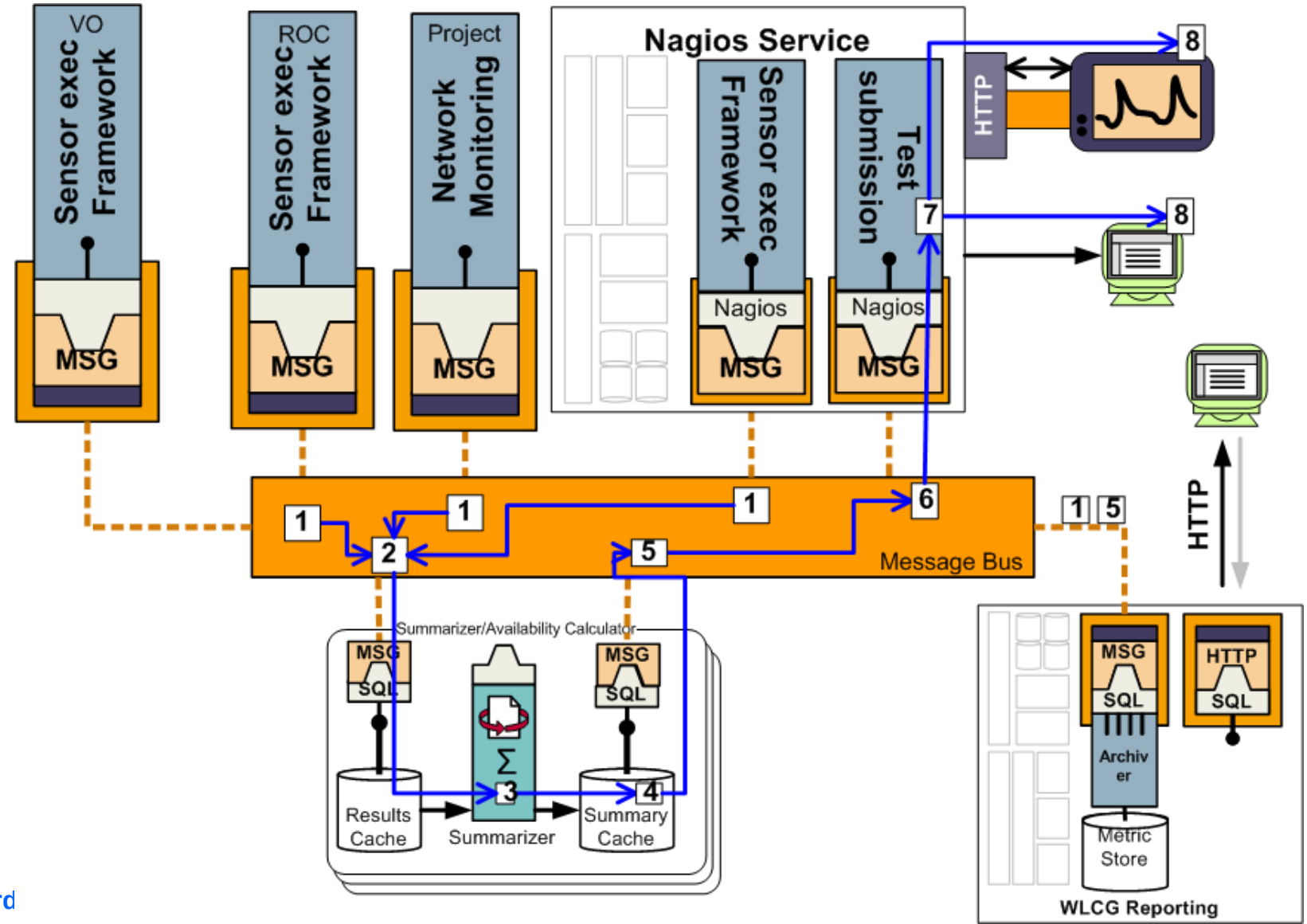
- For the Tier 2 sites reporting:

Sites	Top 50%	Top 20%	Sites> 90%
%CPU	72%	40%	70%

Reliability…

- Site reliability/availability affects resource delivery and usability of a site
  - This has been ***the*** outstanding problem for a long time (slow improvement of reliability)
- Addressed through human oversight (Grid Operator on Duty) ...
  - Teams of experts on duty to flag problems and follow up with sites
    - In place since late 2004; effort intensive
    - Instrumental in stabilisation and gradual improvement of reliability
    - Unsustainable in the long term
- ... and though better monitoring ...
  - Monitoring tools (many!)
  - Aggregating information
  - Understanding (visualising) the data
  - Automation

Integration: Message-based archiving & reporting



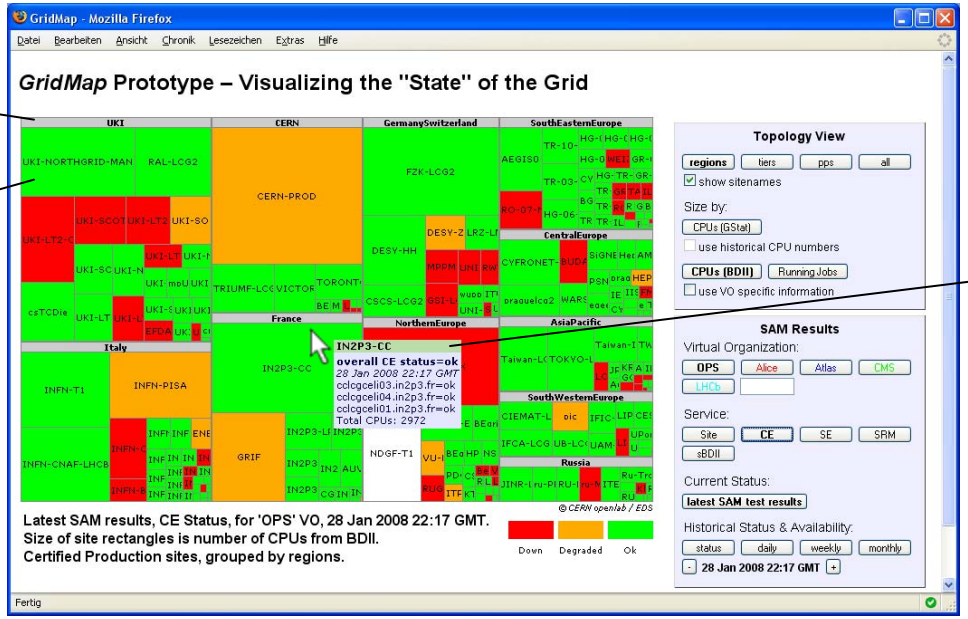
GridMap – Status

- Top-level visualization of distributed systems*
- Prototype deployed since EGEE'07

  1st version Oct 2007, 2nd version Nov 2007, Link: http://gridmap.cern.ch

region

site
- size = #CPUs
- colour = status or availability



details, links to monitoring tool

- GridMaps visually correlate "importance" to status

* Patent Pending

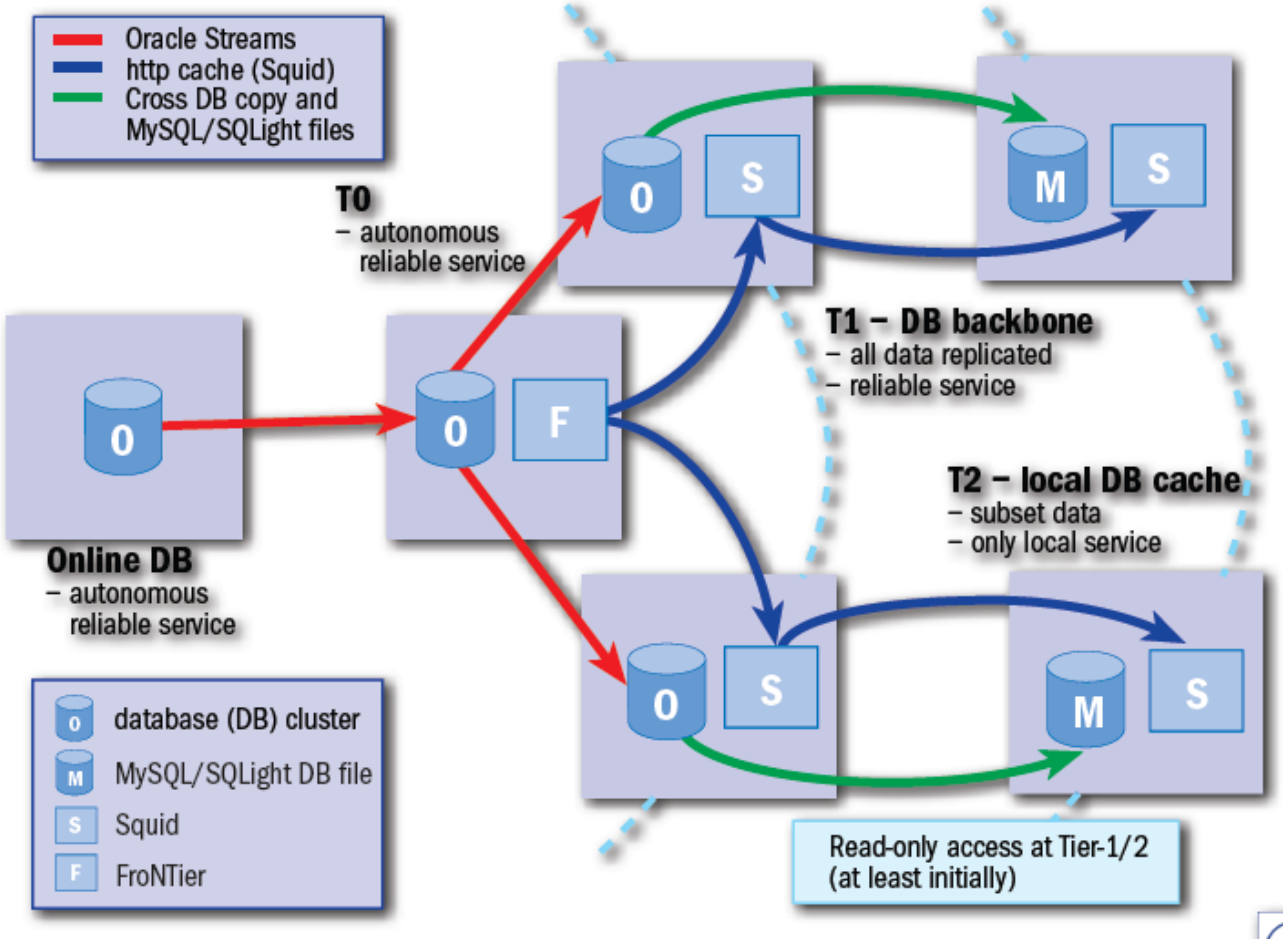
Operations evolution

- Existing model of “central” management – while essential in getting to the point we are at now – is unsustainable in the long run

- Devolve the responsibility for operational oversight to the regions (regional, national operations teams):
  - We now begin to have the understanding and tools to facilitate this
  - Local (site) fabric monitoring should now get grid as well as local alarms – sites can respond directly without needing a central operator to spot a problem and open a ticket
  - Define critical tests (generic and VO-specific) that can generate alarms at a site
    - Tools and monitoring “architecture” can now start to support this

- Central project management tasks will simplify to gathering data relevant to the MoU
  - Accounting, reliability, responsiveness, etc.

Database replication



- 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

Applications Area – new projects

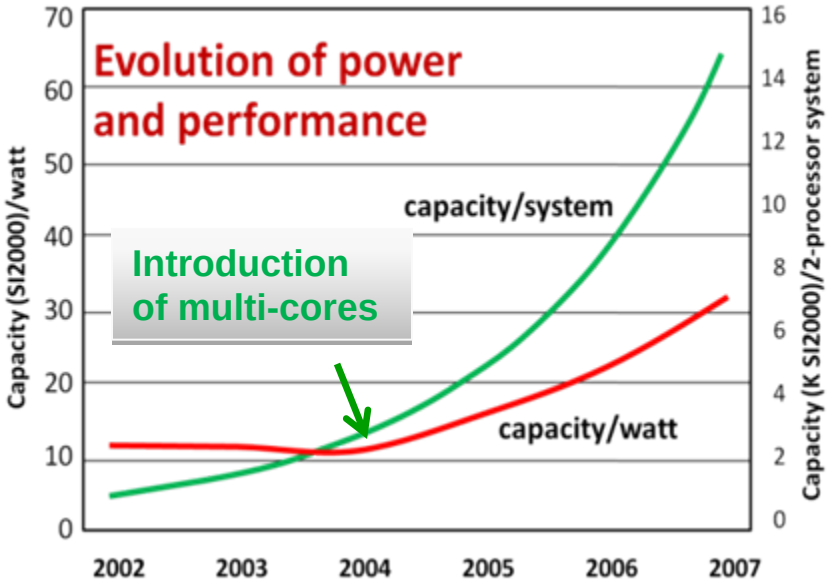
- Parallelization of software frameworks to exploit multi-core processors
  - Adaptation of experiment software to new generations of multi-core processors – essential for efficient utilisation of resources
  - Investigate current and future multi-core architectures
    - Evaluate tools to measure performance
    - Develop a measurement and analysis methodology
  - Measure and analyze performance of current LHC physics application software on multi-core architectures
    - Identify bottlenecks
    - Prototype solutions at the level of system and core libraries
  - Investigate solutions to parallelize current LHC physics software at application framework level
    - Identify reusable design patterns and implementation technologies to achieve parallelization
  - Investigate solutions to parallelize algorithms used in current LHC physics application software
    - Identify reusable design patterns and implementation technologies to achieve effective high granularity parallelization

New projects – 2

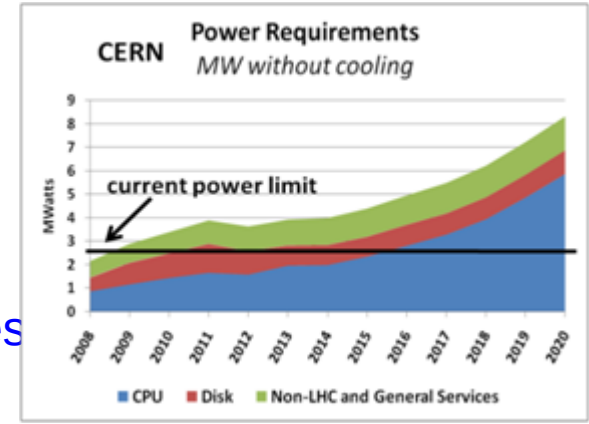
- Portable analysis environment using virtualization technology
  - Study how to simplify the deployment of the complex software environments to distributed (grid) resources
  - Evaluation of the available virtualization technologies
    - Understand and validate technologies by checking their performance, usability and platform constraints
    - Evaluation of the tools to build and manage ‘virtual appliances’
  - Deployment of a read-only distributed file system with aggressive caching schema
    - Essential to avoid any pre-installation of layered software
    - Validate performance, scalability and usability
  - Collect requirements from experiments and confront them with available technologies
    - Suggest optimal choice for given use case
  - Provide prototypes of data analysis virtual appliances for at least 2 experiments
    - Asses their suitability for providing portable and easy to install data analysis environments
    - Assist experiments in adapting their software process to this platform

Power and infrastructure

- Expect power requirements to grow with capacity of CPU
  - This is not a smooth process: depends on new approaches and market-driven strategies (hard to predict) e.g. improvement in cores/chip is slowing; power supplies etc. already >90% efficient
  - No expectation to get back to earlier capacity/power growth rate



- e.g. Existing CERN Computer Centre will run out of power in 2010
  - Current usable capacity is 2.5MW
  - Given the present situation Tier 0 capacity will stagnate in 2010

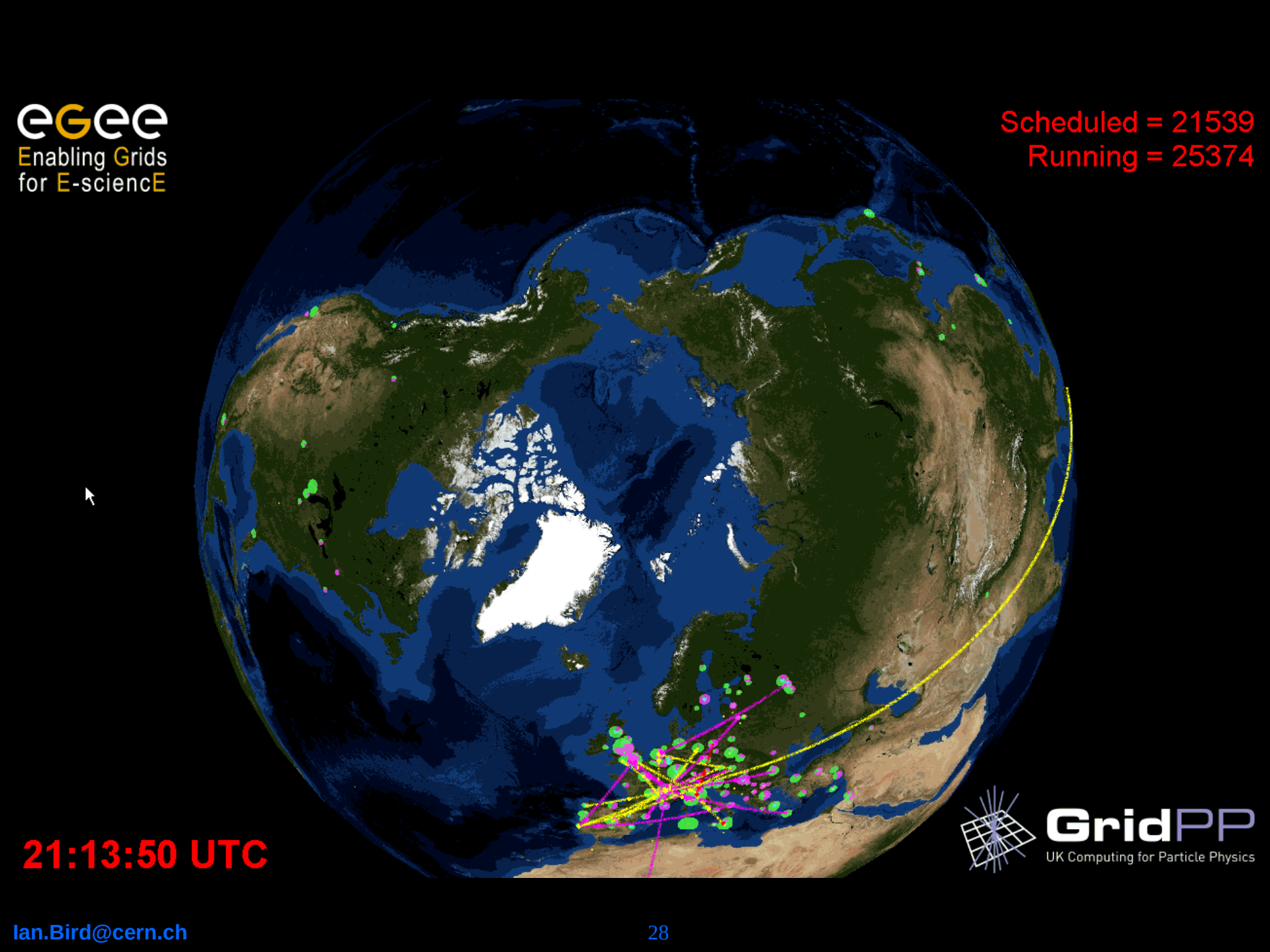


- Major investments are needed for new Computer Centre infrastructure at CERN and major Tier 1 centres
  - IN2P3, RAL, FNAL, BNL, SLAC already have plans
  - IHEPCCC report to ICFA at DESY in Feb '08

Summary

- We have an operating production quality grid infrastructure that:
  - Is in continuous use by all 4 experiments (and many other applications);
  - Is still growing in size – sites, resources (and still to finish ramp up for LHC start-up);
  - Demonstrates interoperability (and interoperation!) between 3 different grid infrastructures (EGEE, OSG, Nordugrid);
  - Is becoming more and more reliable;
  - **Is ready for LHC start up**

- For the future we must:
  - Learn how to reduce the effort required for operation;
  - Tackle upcoming issues of infrastructure (e.g. Power, cooling);
  - Manage migration of underlying infrastructures to longer term models;
  - Be ready to adapt the WLCG service to new ways of doing distributed computing.


eGee
Enabling Grids
for E-sciencE
Scheduled = 21539
Running = 25374
21:13:50 UTC
GridPP
UK Computing for Particle Physics