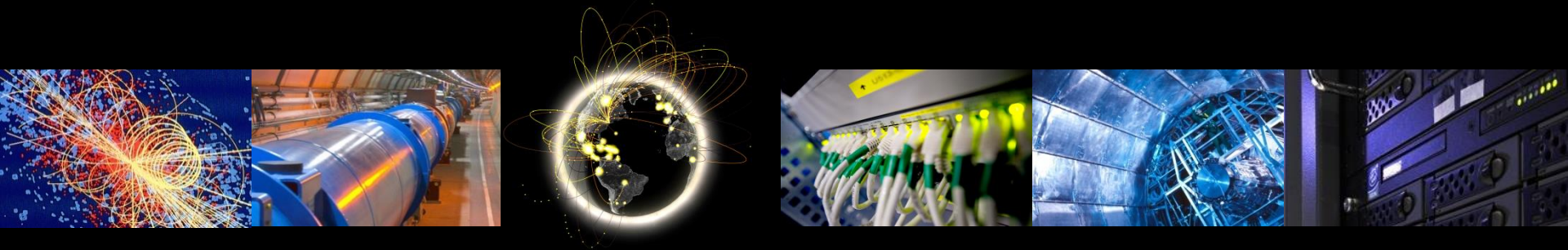


# Computing on the grid and in the cloud

Laurence Field

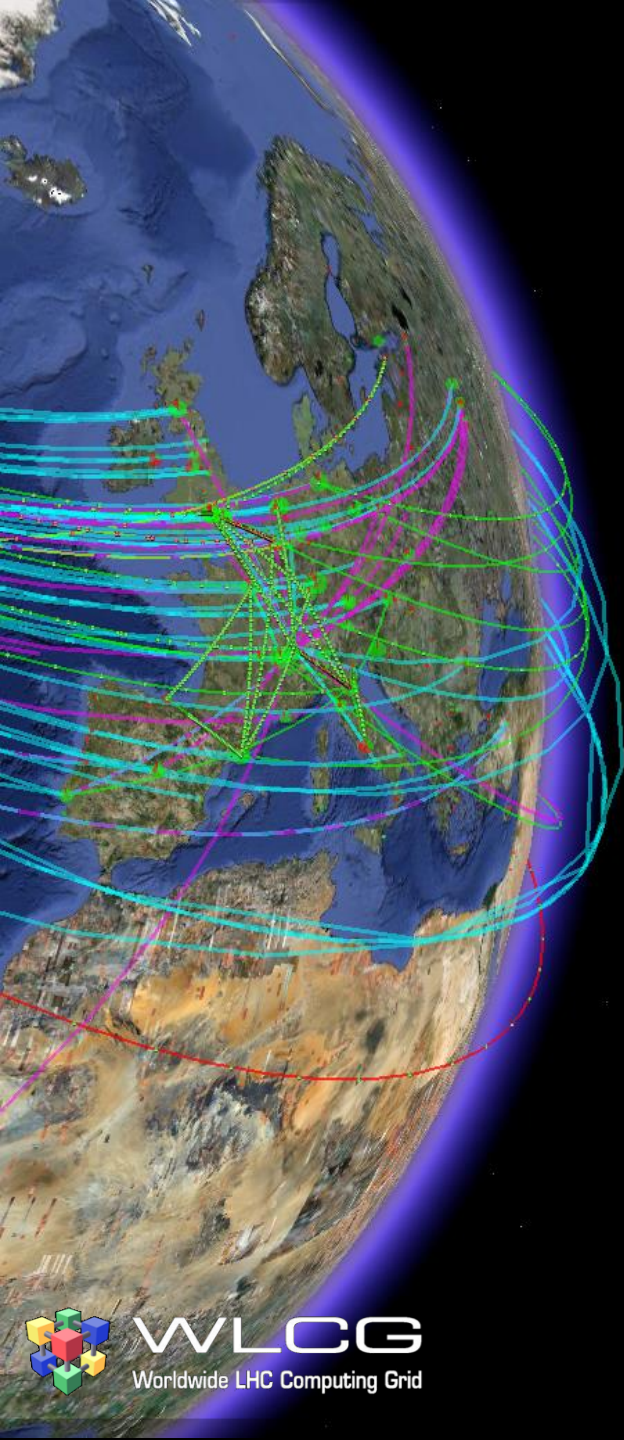
Support for Distributed  
Computing Group



# Overview

- The computational problem
- The computing challenge
- Grid computing
- The WLCG
- Operational experience
- Future perspectives

# The Computational Problem





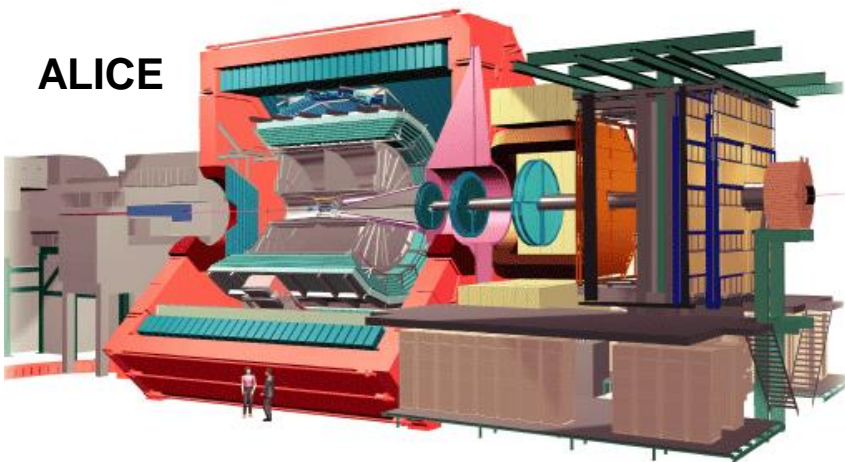
Delivering collisions at 40MHz



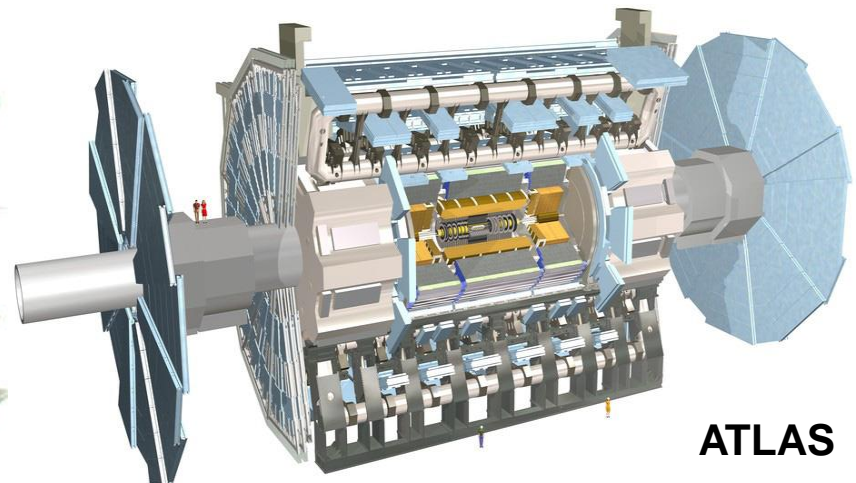


# The Detectors

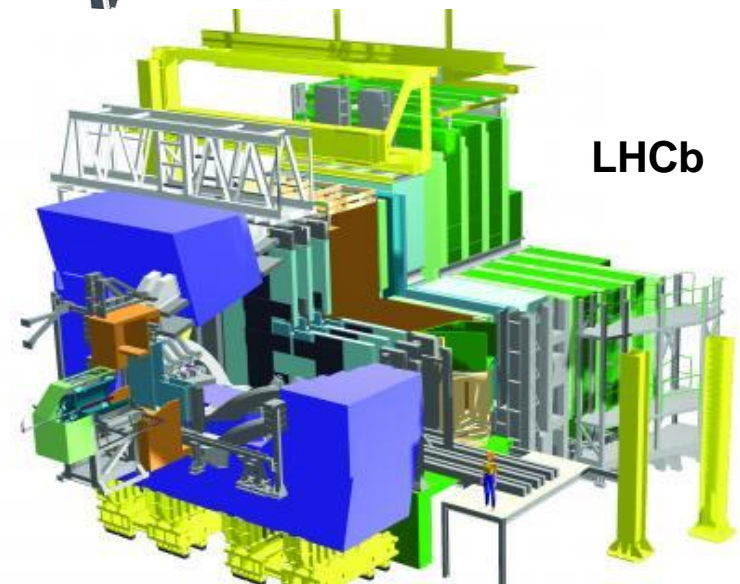
**ALICE**



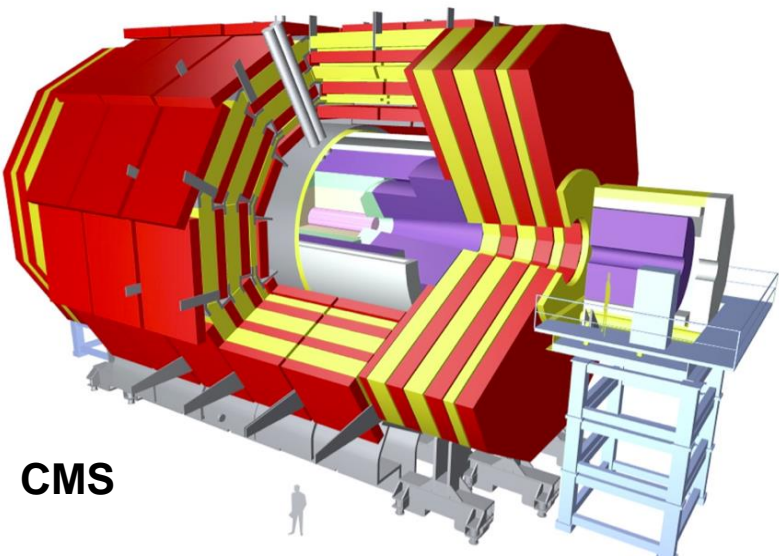
**ATLAS**



**LHCb**



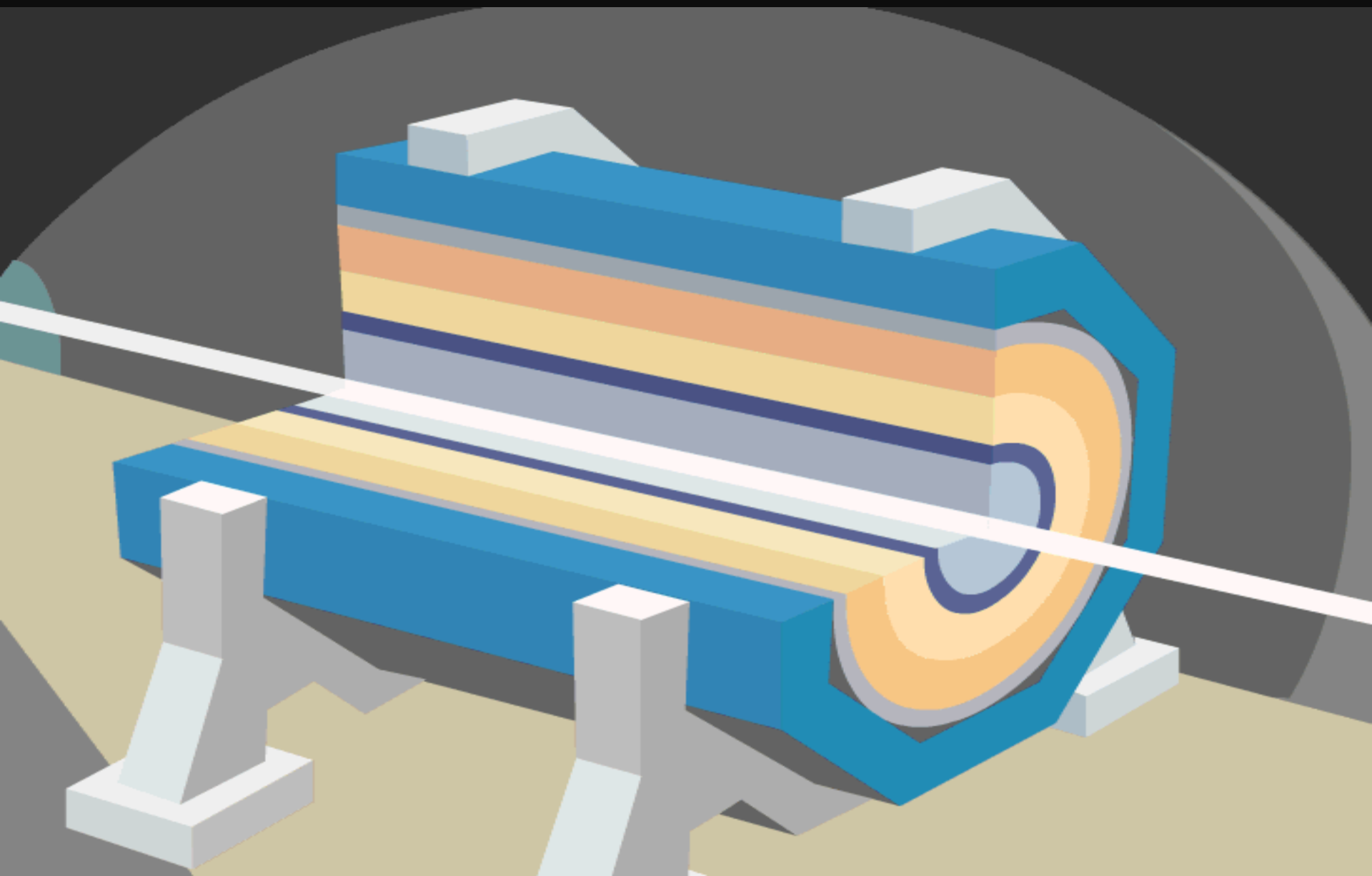
**CMS**



150 million sensors

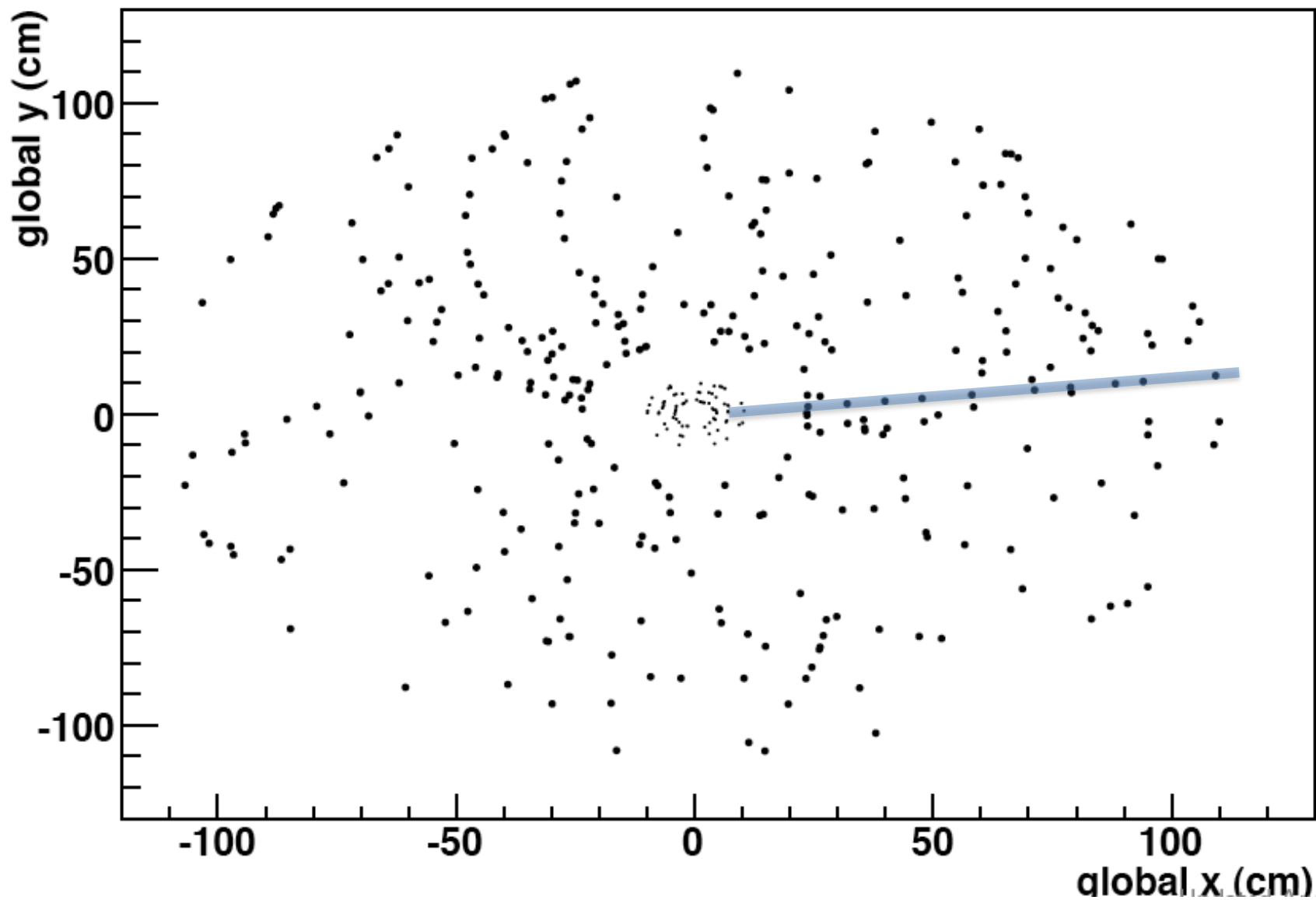


# A Collision





# Raw Data





# Data Acquisition

~ 300.000 MB/s  
from all sub-detectors

1 GB/s  
Raw Data

*Trigger and data acquisition*

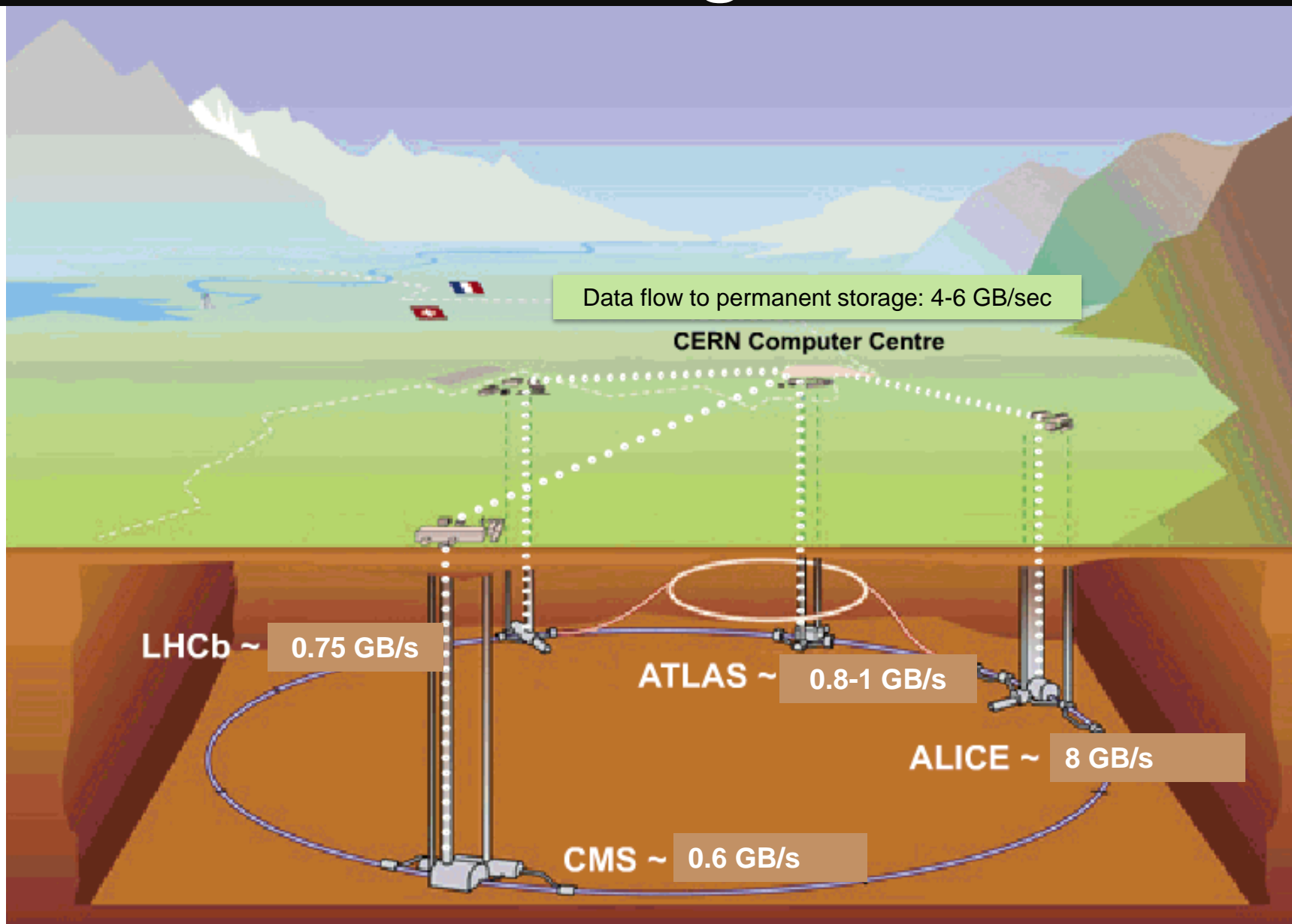


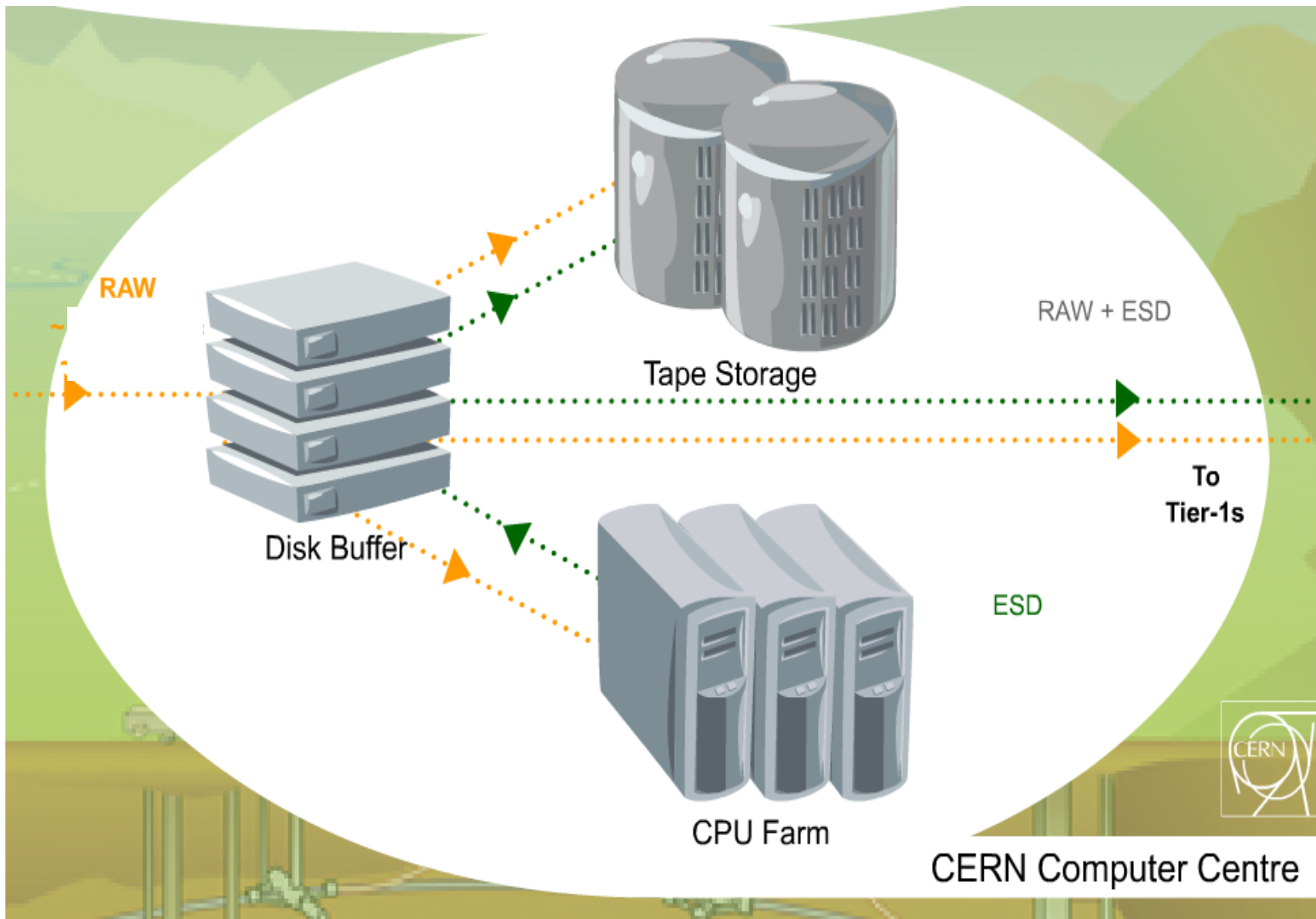
*Event filter computer farm*

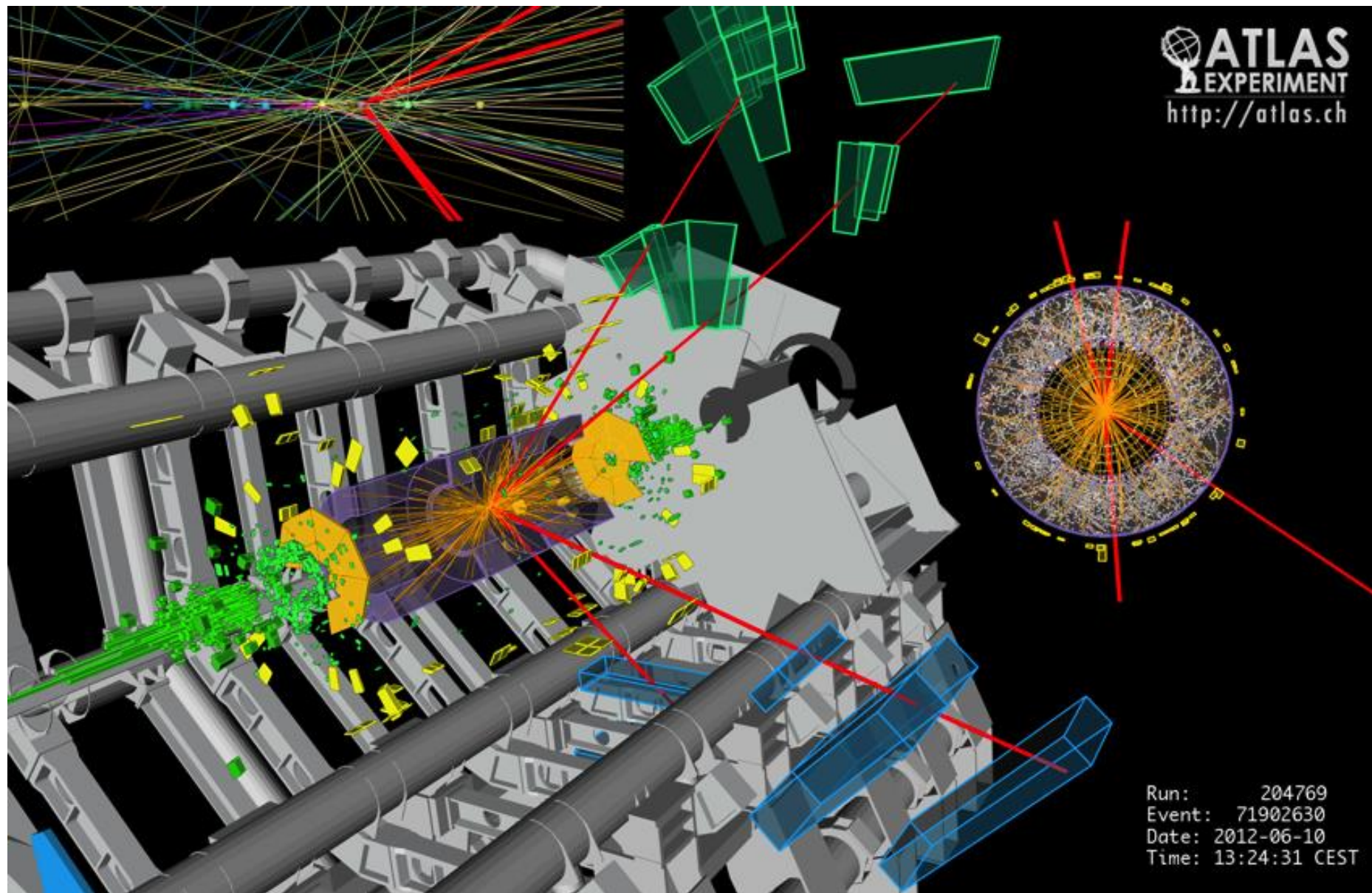




# Data Mining



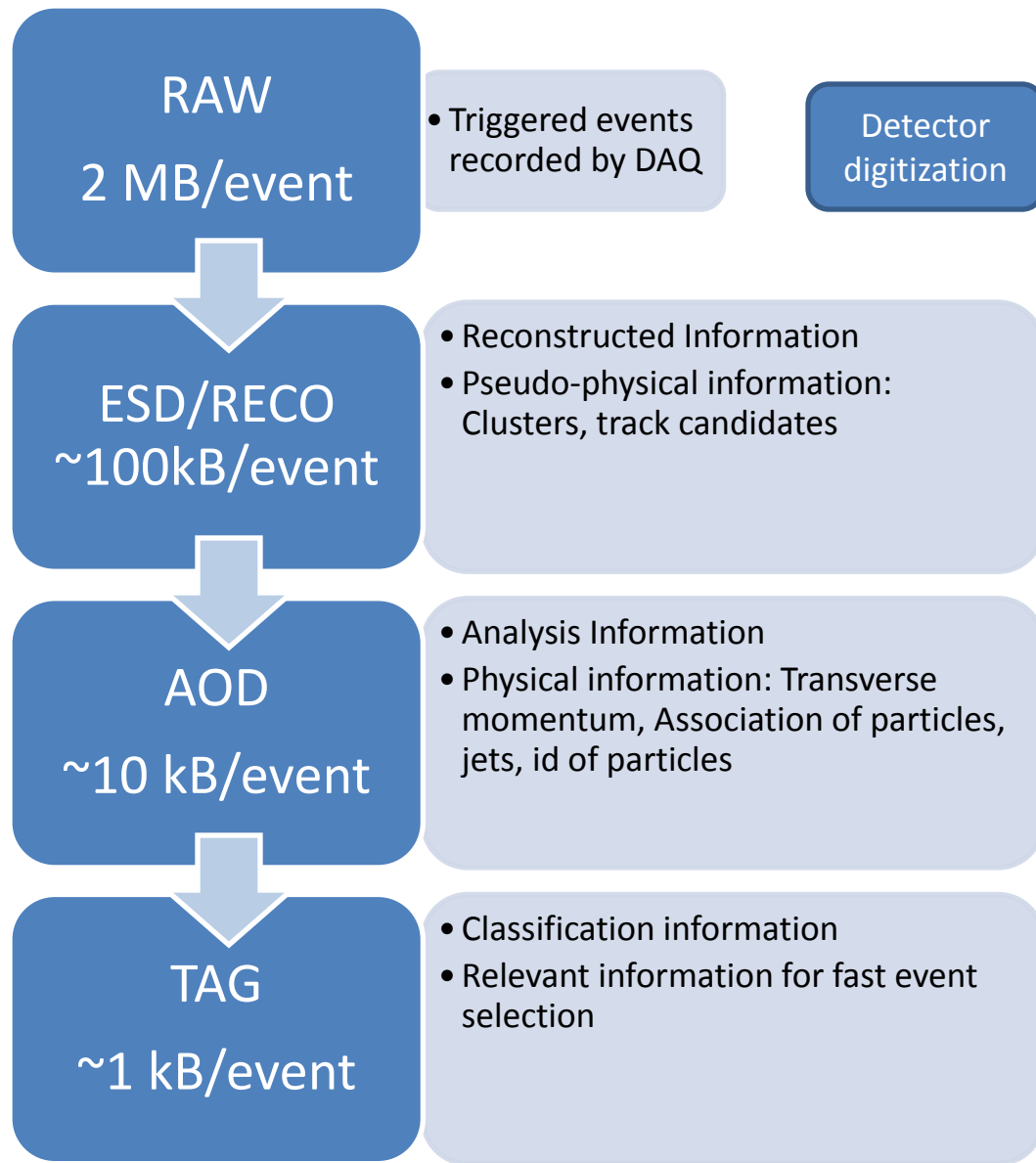




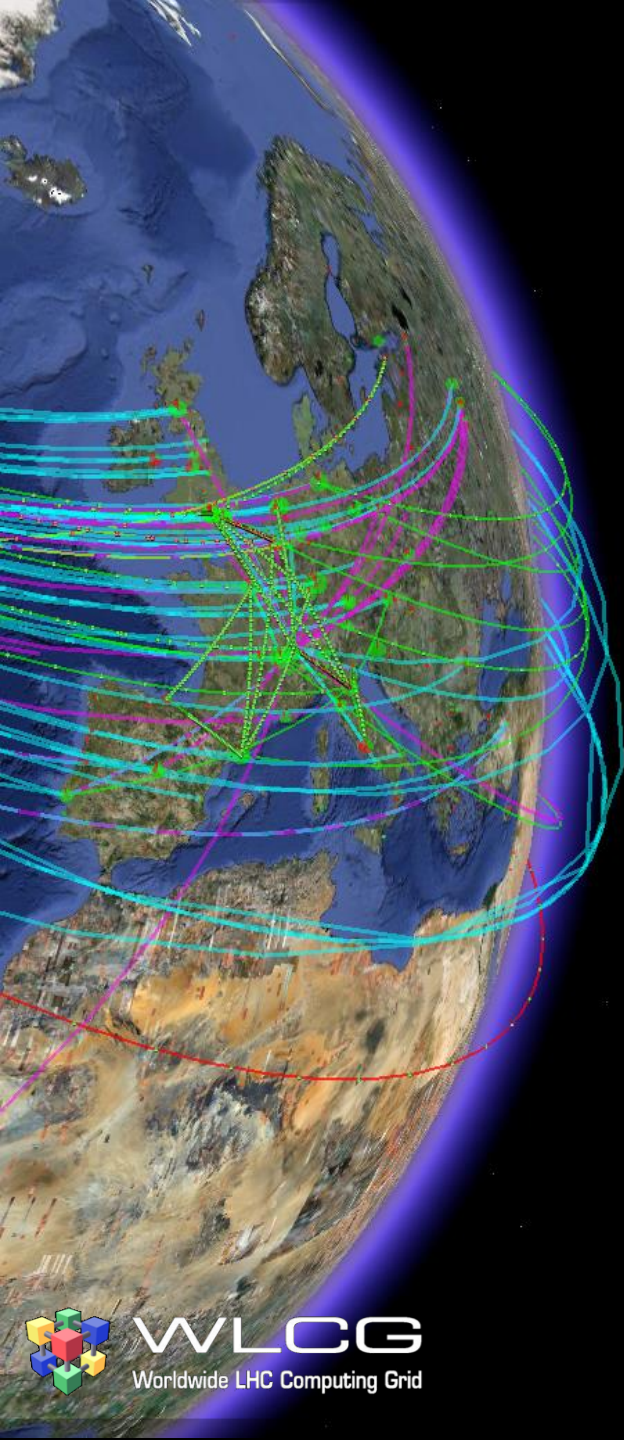


# Data and Algorithms

- Data are organized as Events
  - Particle collisions
- Event processing algorithms
  - Selection/Filtering
  - Reconstruction
  - Simulation (generation)
  - Analysis
- Embarrassingly parallel
  - Events are independent
    - Process one event at a time
- High Throughput Computing



# The Computing Challenge



# Computational Workflow

Online trigger  
and filtering

Selection &  
reconstruction

Offline Reconstruction

100%

Raw data

Event  
summary  
data

10%

Event  
reprocessing

Event  
simulation

Offline Simulation  
w/GEANT4

Processed  
Data (Active tapes)

Batch  
physics  
analysis

1%

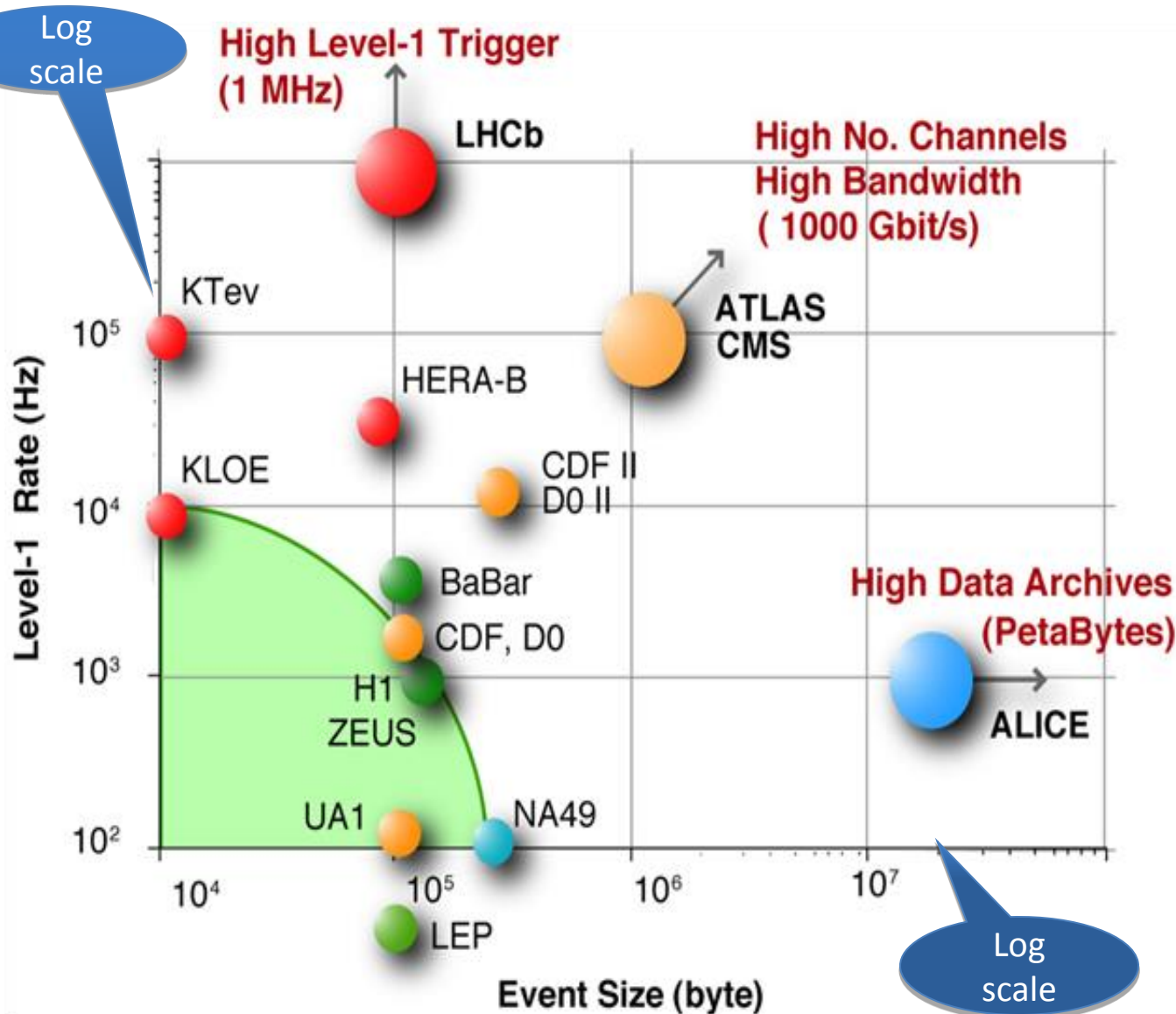
Offline Analysis w/ROOT

Analysis objects  
(extracted by physics topic)

Interactive  
analysis



# Data Volume



- 25PB per year + simulation
- Preservation
  - for 25+ years
- Processing
  - 340k cores

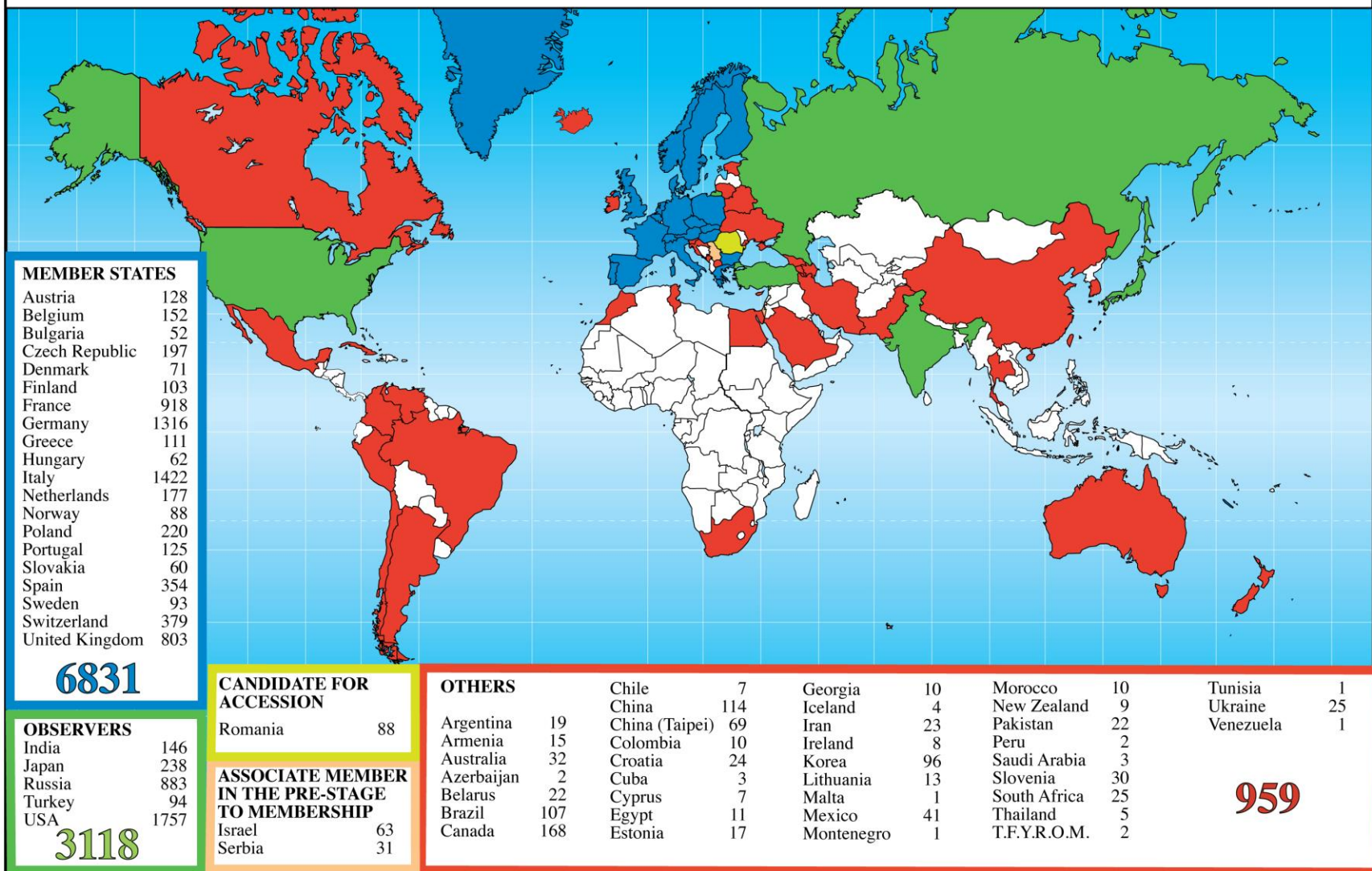


- 1 PB
  - Detector data rate
  - 240m DVD tower
- 25PB
  - Run 1 yearly output
  - 6km DVD Tower
- 100PB
  - CERN data centre
  - 24km DVD tower
- 140PB
  - ATLAS dataset
  - 33.6km DVD tower





## Distribution of All CERN Users by Location of Institute on 14 January 2013





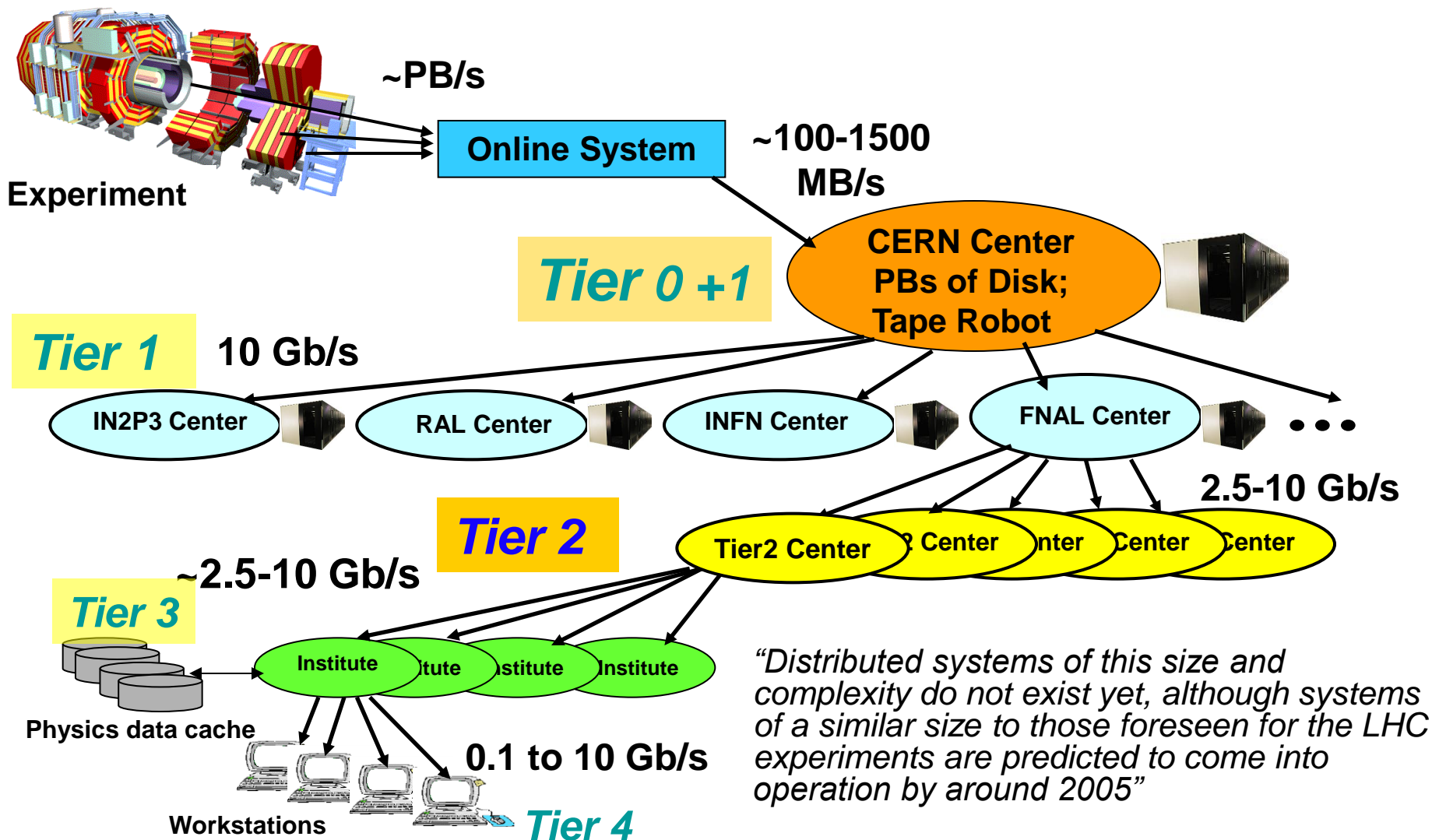
# Distributed HTC

- Technical and political/financial reasons
  - No single centre could provide ALL the computing
    - Buildings, Power, Cooling, Cost, ...
  - The community is distributed
    - Computing already available at many institutes
      - Funding for computing is also distributed
- How do you distributed HTC?
  - With big data
  - With hundreds of computing centres
  - With a global user community
  - It is 1998
  - And data is coming!



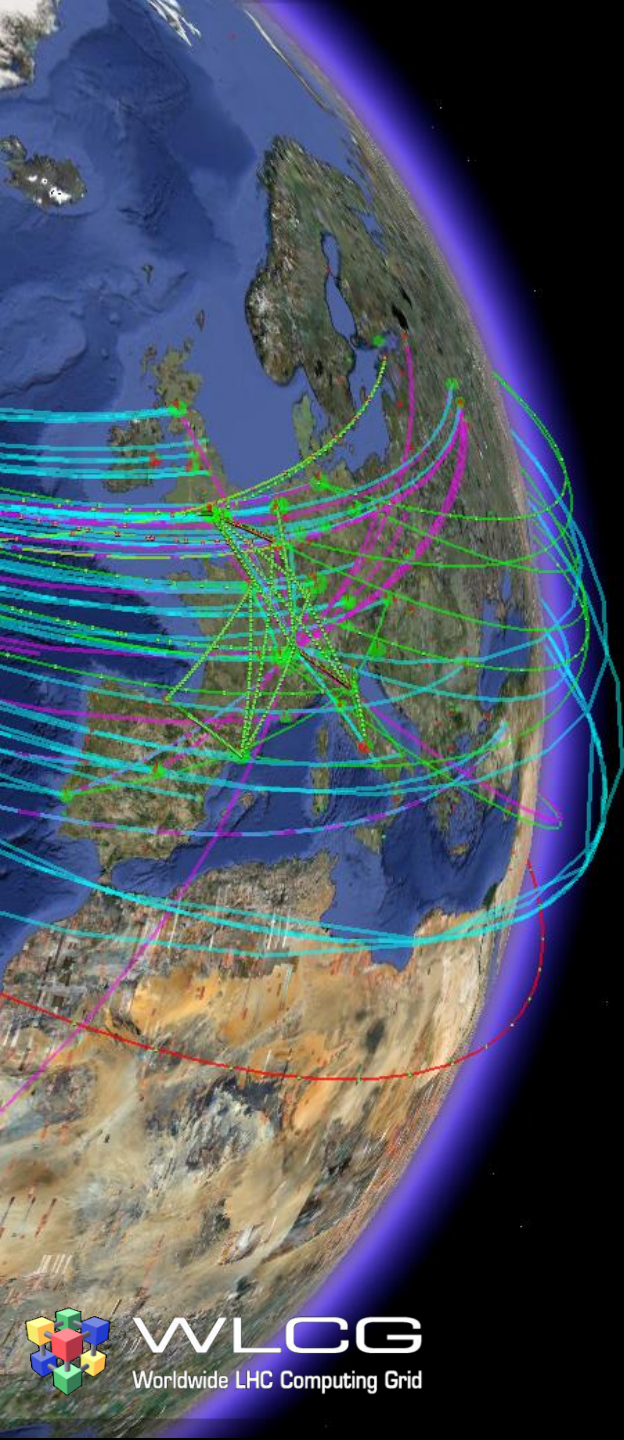
# The MONARC Model - 1999

Models of Networked Analysis at Regional Centres



# The Grid

- *“Coordinated resource sharing and problem –solving in dynamic, multi-institutional virtual organizations”*





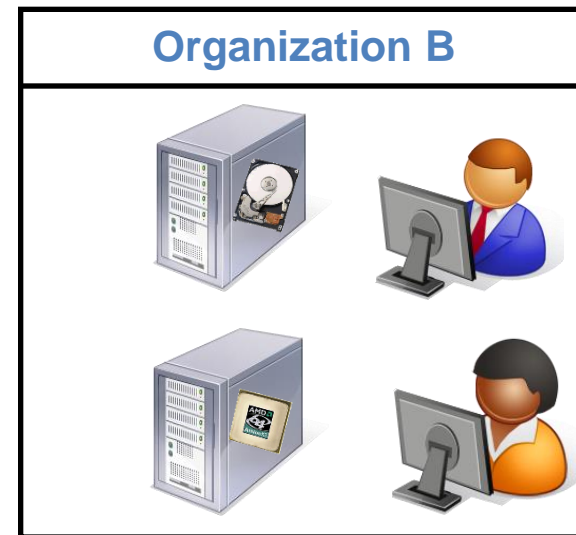
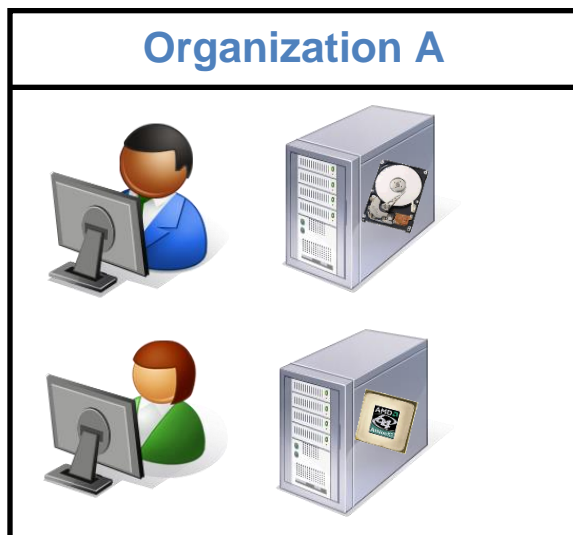
# The Origin Of Grid Computing

- Metacomputing
  - Information Wide Area Year (IWAY) - 1995
    - Attempt to link 17 supercomputing centres in the U.S.
      - As a *seamless* resource
        - » As easy as using a single computer
  - A Metacomputing Infrastructure Toolkit - 1996
    - Heterogeneity, administrative domains, scale
      - Low-level mechanisms for high-level services
  - The National Technology Grid – 1997
    - Aimed to deploy metacomputing systems across the U.S.
    - Provide routine application support
      - Previously metacomputing required heroic efforts
    - Analogous to the Electrical Power Grid
      - Aims to *seamlessly* deliver computing power as a resource similar to how electrical power is delivered over the electrical power grid

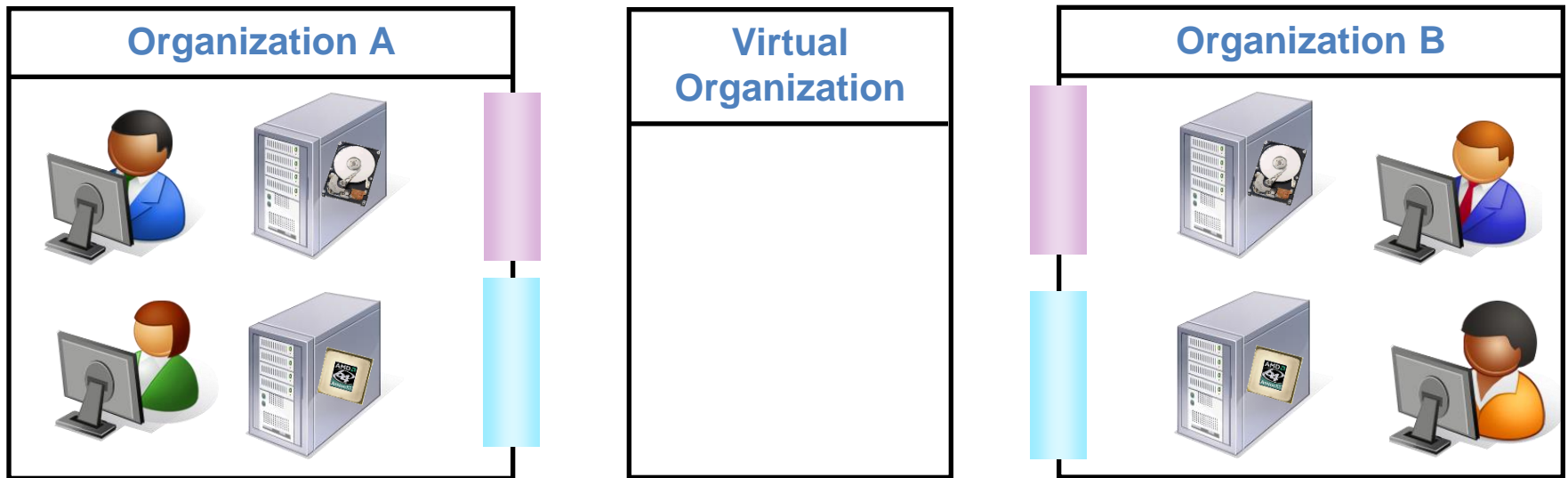




# What Is The Problem?



- Organization A and B are administrative domains
  - Independent policies, systems and authentication mechanisms
- Users have local access to their local system using local methods
- Users from A wish to collaborate with users from B
  - Pool the resources
  - Split tasks by specialty
  - Share common frameworks



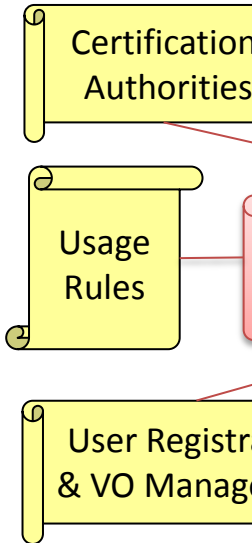
- The Users from A and B create a Virtual Organization
  - Users have a unique identify but also the identity of the VO
- Organizations A and B support the Virtual Organization
  - Place “grid” interfaces at the organizational boundary
  - These map the generic “grid” functions/information/credentials
    - To the local security functions/information/credentials
- Multi-institutional e-Science Infrastructures

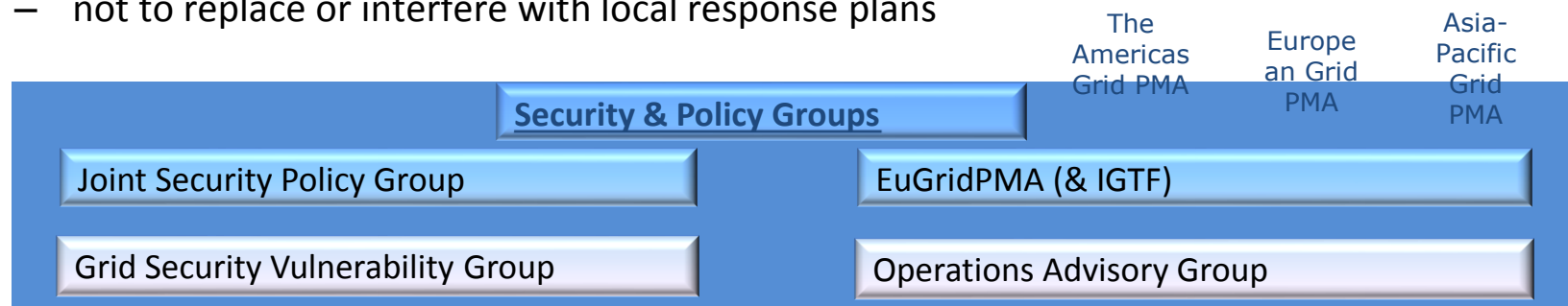
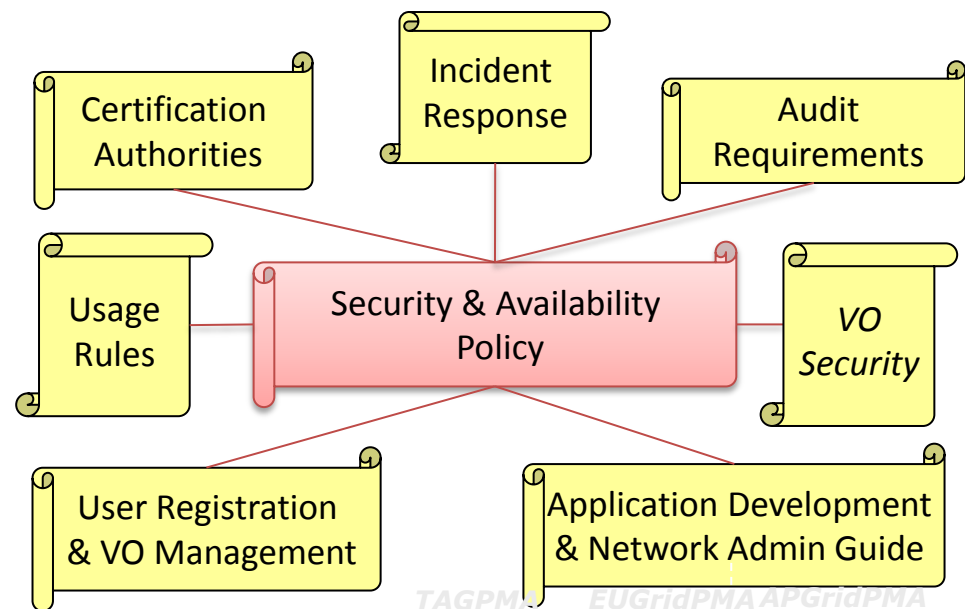


# A Security Architecture

- User authentication
  - Pre-configuration within an organization
  - Not possible for large number of users and resources
- Delegation of trust concept
  - Org A trusts a user from Org B because Org A has relationship with Org B
- Security policy to enable single sign on spanning multiple admin domains
  - Interoperability with local policies in dynamic environments
- Virtual Organization
  - A multi-institutional collaboration
- Key concept, multiple trust domains
  - Individual operations confined to a single trust domain
    - And subject to local policy
      - local authorization decision for access control
- A mapping from a global to local subject exists
  - Mutual authentication required for operations between trust domains

# Security & Policy

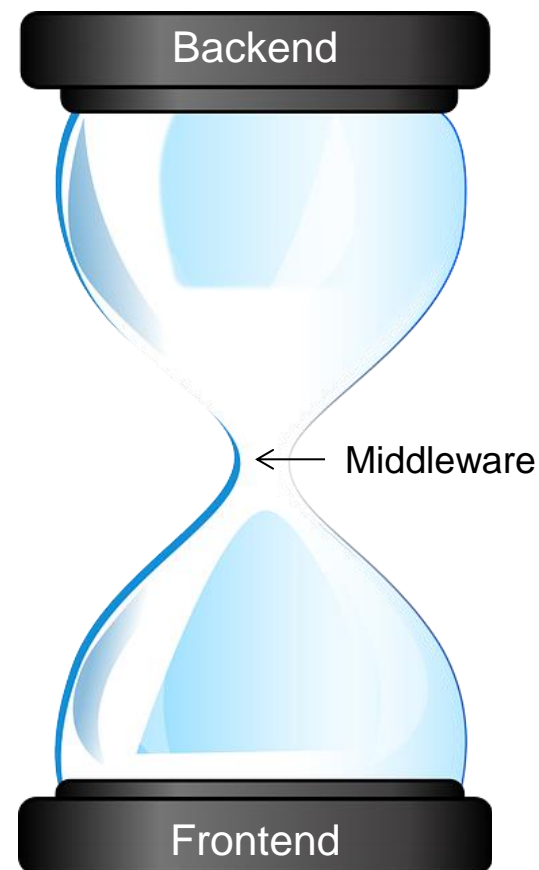
- Collaborative policy development
  - Joint Security Policy Group
  - Certification Authorities
    - EUGridPMA → IGTF, etc.
  - Grid Acceptable Use Policy (AUP)
    - common, general and simple AUP
    - for all VO members
    - using many Grid infrastructures
      - EGI, OSG, NGIs, ...
  - Incident Handling and Response
    - defines basic communications paths
    - defines requirements (MUSTs) for IR
    - not to replace or interfere with local response plans
- 
- ```
graph TD; A[Certification Authorities] --> B[Usage Rules]; B --> C[User Registration & VO Management];
```





# The Hourglass Model

- Three tiered model
  - Middle tier mediates
    - Sophisticated back-end services
    - Potential simple front end services
- Protocol-based architecture
  - Built upon public key-based Grid Security Infrastructure
    - Extend the Transport Layer Security protocols
- Grid Services - 2002
  - Leveraging concepts from the Web service community
  - Network-enable entities that provide some capability
- Integrate across multiple organizations
  - Lack of centralized control
    - Probably missing the federation concept
  - Geographical distribution
  - Different policy environments
    - International issues





# Grid Computing

- A Grid is the hardware and software infrastructure
  - That supports access to computational capabilities
- Five classes of applications were defined
  - Distributed supercomputing
  - High-throughput computing
  - On-demand computing
  - Data-intensive computing
  - Collaborative computing
- Key aspect
  - Sharing of resources across administrative domains
    - Not clear if the technical and political cost would outweigh the benefits
      - Especially when crossing institutional boundaries
- Sharing is governed by policy
  - What, who, conditions in which it occurs

# WLCG

- An International collaboration to distribute and analyse LHC data
- Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists
- CHEP 2000
  - Grid computing discussed
    - Distributed resources
    - Trust model
  - Extending
    - To data intensive tasks
    - To a global scale



CERN



US-BNL



Amsterdam/NIKHEF-SARA



Taipei/ASGC



Bologna/CNAF



Ca-  
TRIUMF



US-FNAL



De-FZK

## WLCG Collaboration Status

Tier 0; 13 Tier 1s; 72 Tier 2 federations  
(156 Tier 2 sites)

Today we have 58 MoU signatories, nearly 40 countries:

Australia, Austria, Belgium, Brazil, Canada, China, Czech Rep,  
Denmark, Estonia, Finland, France, Germany, Greece, Hungary, India,  
Israel, Italy, Japan, Latin America, Netherlands, Norway, Pakistan,  
Poland, Portugal, Rep. Korea, Romania, Russia, Slovakia, Slovenia,  
Spain, Sweden, Switzerland, Taipei, Turkey, UK, Ukraine, USA.



NDGF



Barcelona/PIC



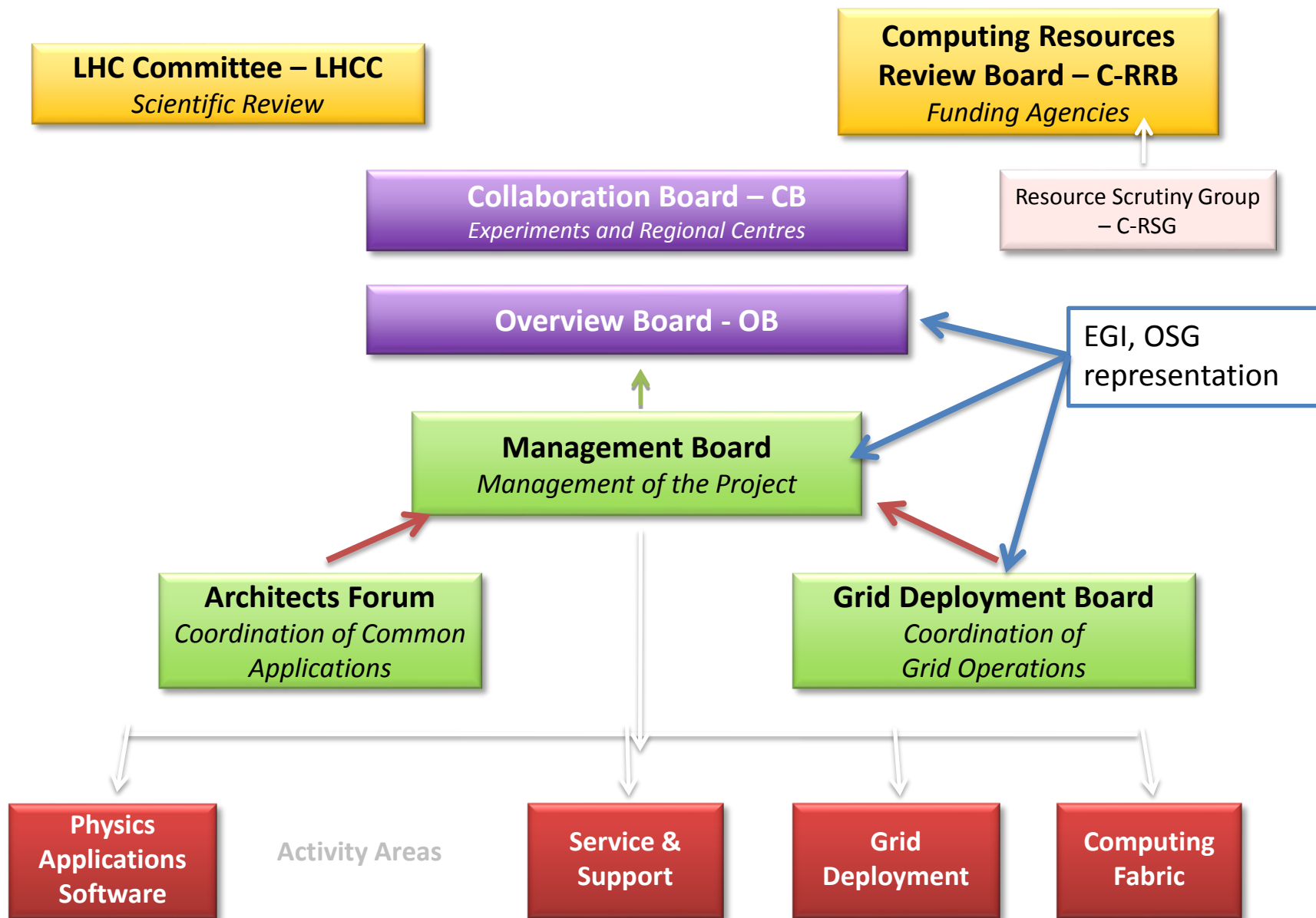
Lyon/CCIN2P3



UK-RAL



# Organisation Structure





# What does WLCG cover?

## Collaboration

Coordination & management & reporting

Coordinate resources & funding

Coordination with service & technology providers

Common requirements

Memorandum of Understanding

## Framework

Service management

Service coordination

Operational security

Support processes & tools

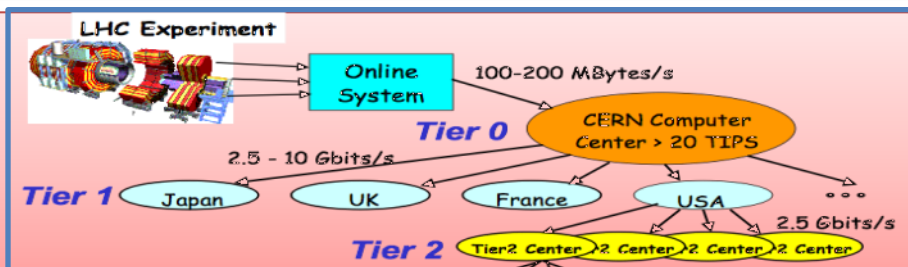
Common tools

Monitoring & Accounting

World-wide trust federation  
for CA's and VO's

Complete Policy framework

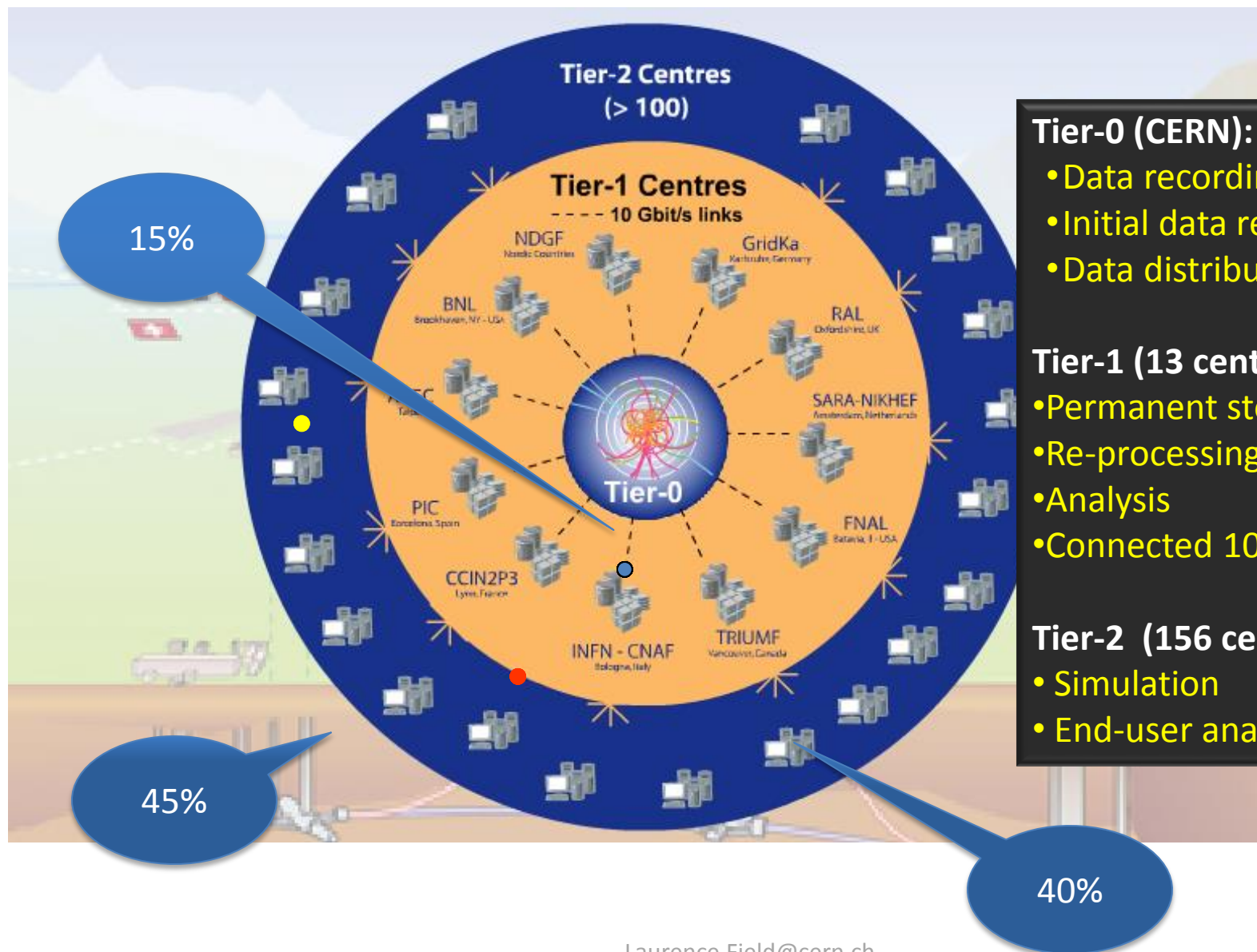
## Distributed Computing services



Physical resources: CPU, Disk, Tape, Networks



# A Tiered Architecture



## Tier-0 (CERN): (15%)

- Data recording
- Initial data reconstruction
- Data distribution

## Tier-1 (13 centres): (40%)

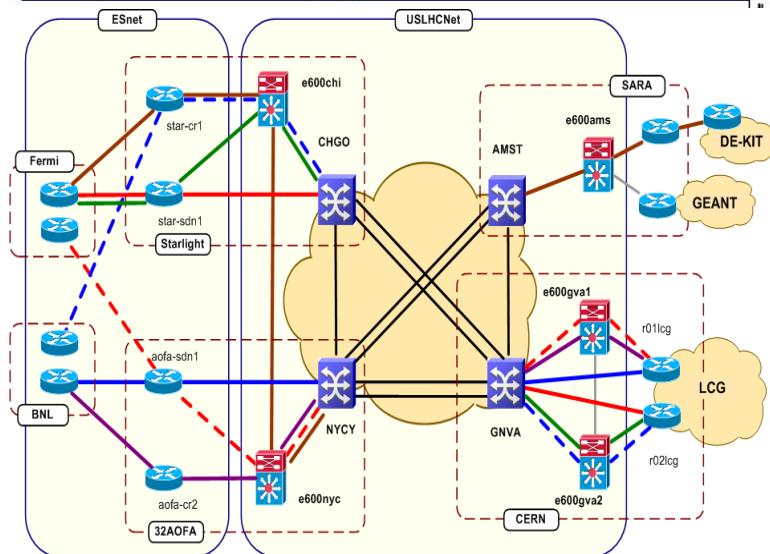
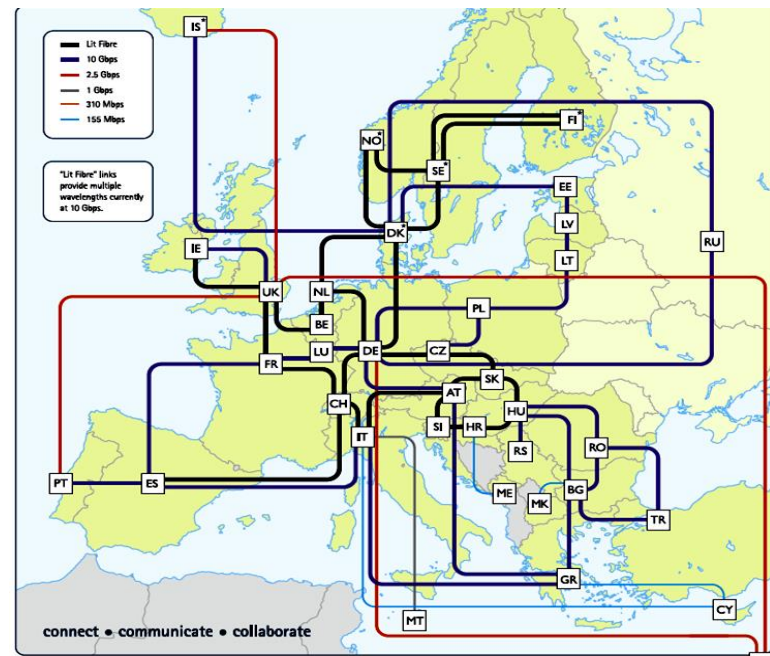
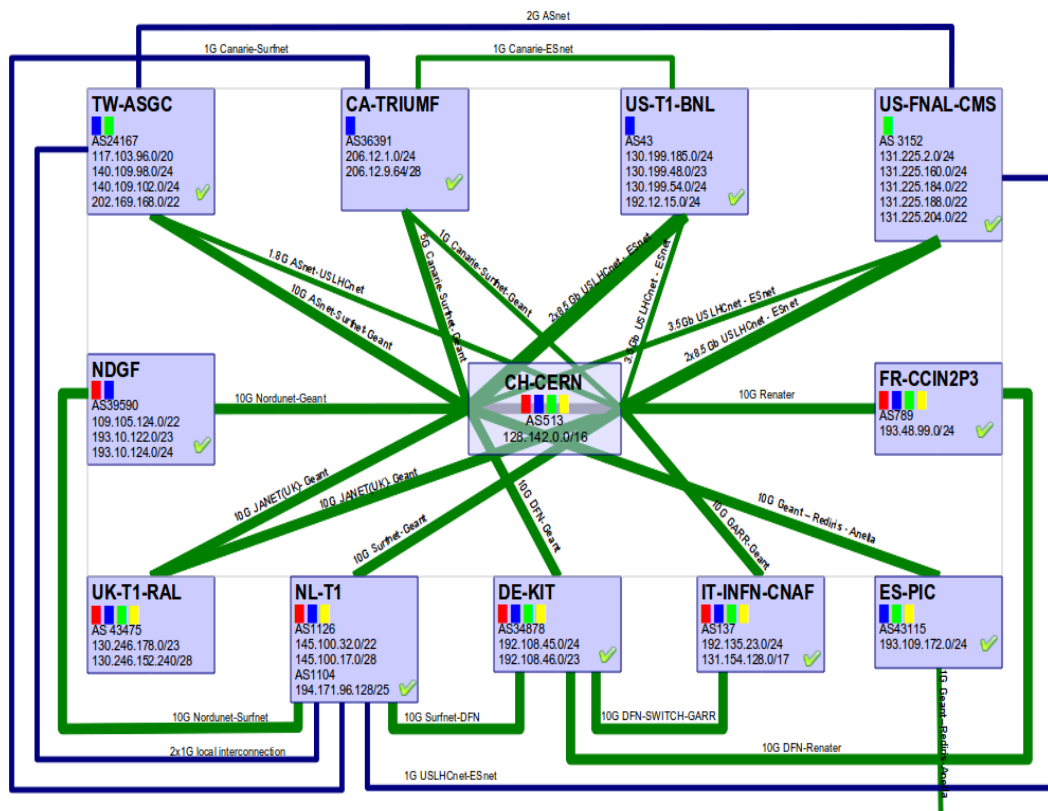
- Permanent storage
- Re-processing
- Analysis
- Connected 10 Gb fibres

## Tier-2 (156 centres): (45%)

- Simulation
- End-user analysis



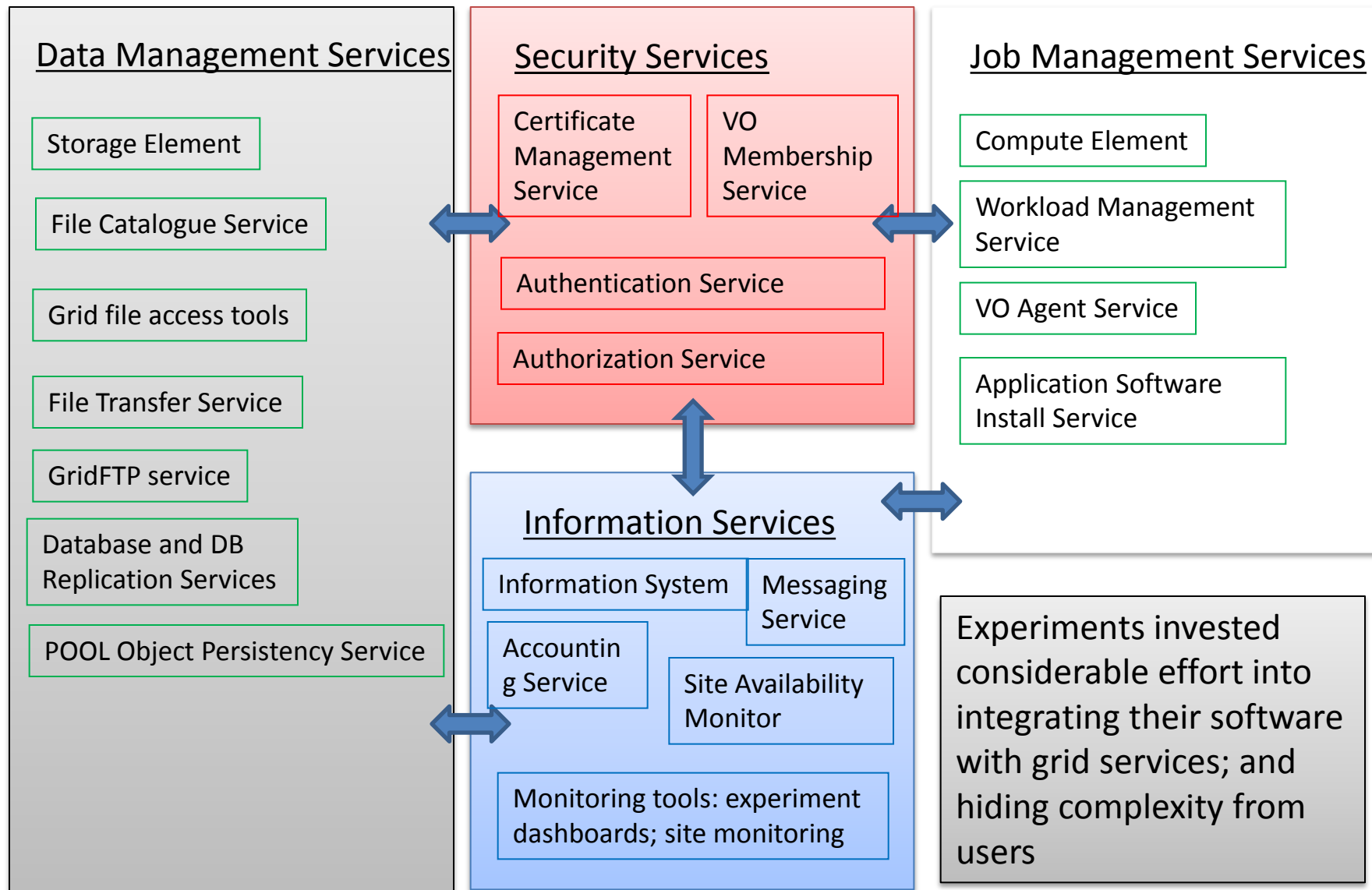
## LHCOPN



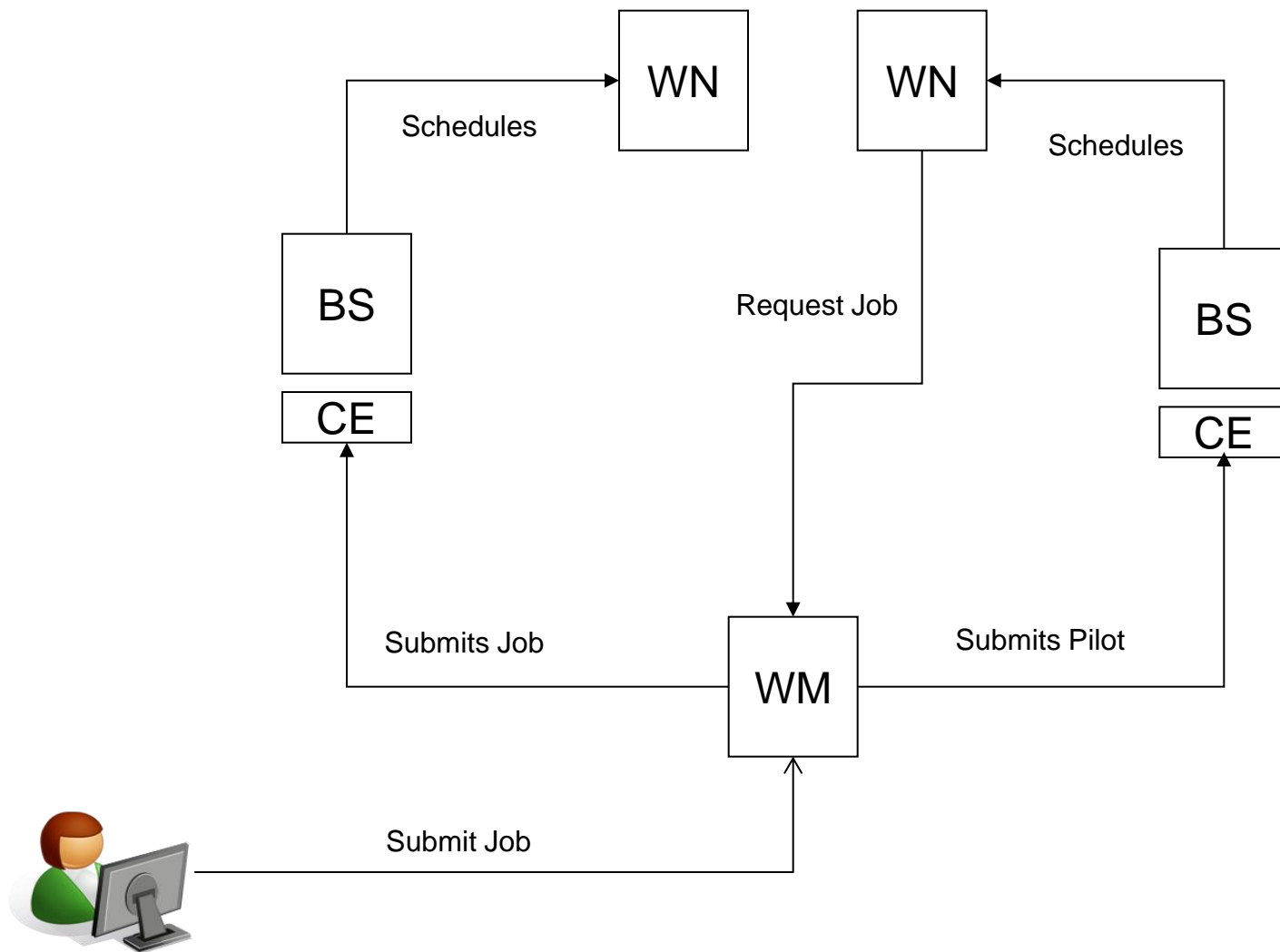
- Relies upon
  - OPN, GEANT, US-LHCNet
  - NRENs & other national & international providers



# Original Grid Services



# Metascheduling and Pilots



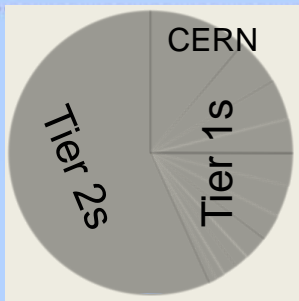
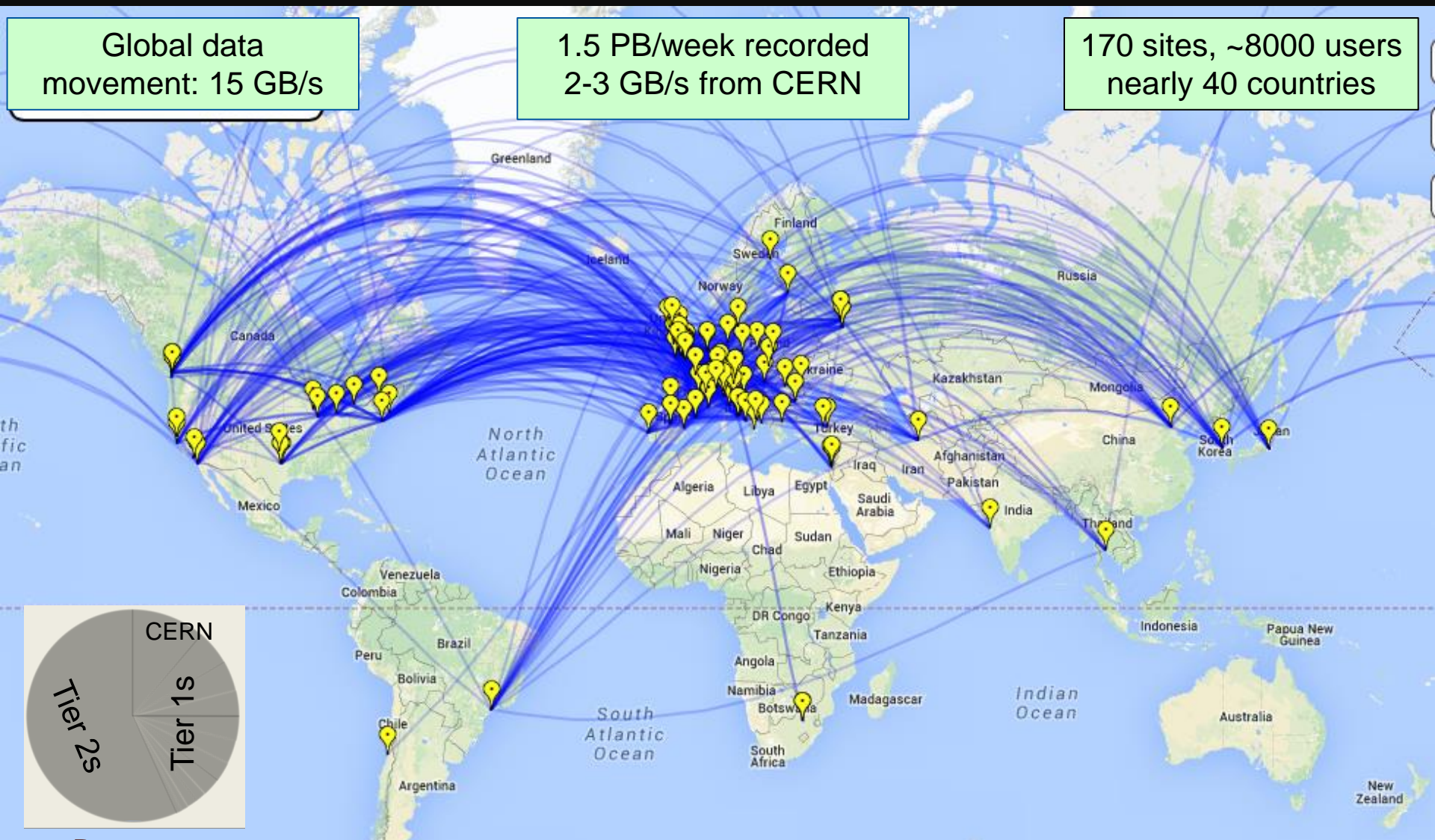


# WLCG Infrastructure

Global data  
movement: 15 GB/s

1.5 PB/week recorded  
2-3 GB/s from CERN

170 sites, ~8000 users  
nearly 40 countries



Resource  
distribution

2 M jobs / day

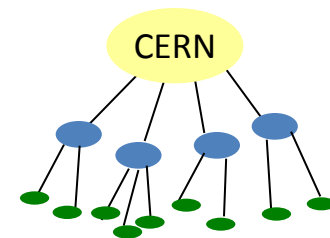
250 000 CPU days/day

200PB Storage



# The Brief History of WLCG

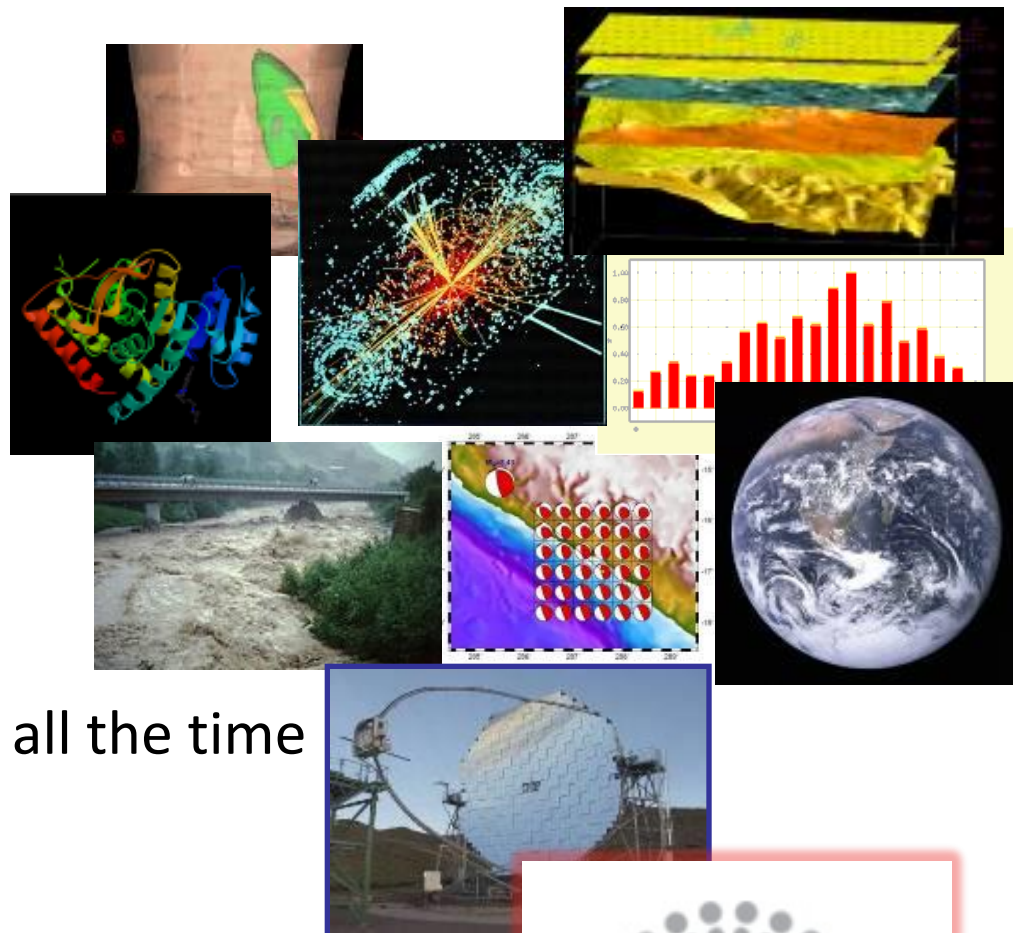
- 1999 - MONARC project
  - Defined the initial hierarchical architecture
- 2000 - Growing interest in Grid technology
  - HEP community main driver in launching the DataGrid project
- 2001-2004 - EU DataGrid project
  - Middleware & testbed for an operational grid
- 2002-2005 - LHC Computing Grid
  - Deploying the results of DataGrid for LHC experiments
- 2004-2006 - EU EGEE project phase 1
  - A shared production infrastructure building upon the LCG
- 2006-2008 - EU EGEE project phase 2
  - Focus on scale, stability Interoperations/Interoperability
- 2008-2010 - EU EGEE project phase 3
  - Efficient operations with less central coordination
- 2010 - 201x EGI and EMI
  - Sustainability





- A few hundred VOs from several scientific domains

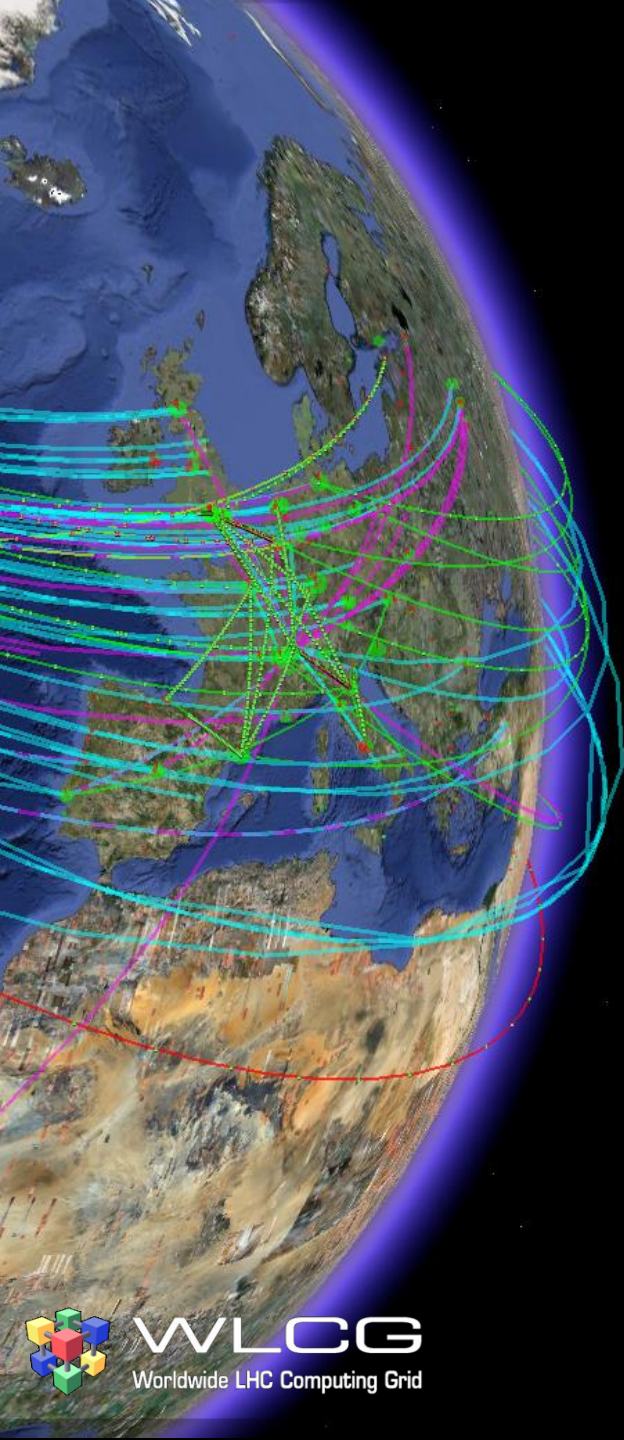
- Astronomy & Astrophysics
- Civil Protection
- Computational Chemistry
- Comp. Fluid Dynamics
- Computer Science/Tools
- Condensed Matter Physics
- Earth Sciences
- Fusion
- High Energy Physics
- Life Sciences
- .....



- Further applications joining all the time
  - Recently fishery ( I-Marine)



# Operations



**WLCG**  
Worldwide LHC Computing Grid

[Laurence.Field@cern.ch](mailto:Laurence.Field@cern.ch)



# Production Grids

- WLCG relies on a production quality infrastructure
  - Used 365 days a year
    - For several years!
  - The system must be fault-tolerant and reliable
    - Can deal with individual sites being down and recover
  - Tier 1s must store the data
    - For at least the lifetime of the LHC (~20 years)
    - Requires active migration to newer media
  - Requires standards of:
    - Availability/reliability
    - Performance
    - Manageability
  - Monitoring and operational tools and procedures
    - As important as the middleware



- Services require
  - Fabric
  - Management
  - Networking
  - Security
  - Monitoring
  - User Support
  - Problem Tracking
  - Accounting
  - Service support
  - SLAs
  - ...
- But now on a global scale
  - Respecting the autonomy of sites
  - Linking the different infrastructures
    - NDGF, EGI, OSG

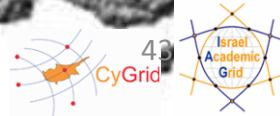


- Not all is provided by WLCG directly
- WLCG links the services
  - Provided by the underlying infrastructures
    - And ensures that they are compatible
- EGI relies on National Grid Infrastructures
  - And some central services
    - User support (GGUS)
    - Accounting (APEL & portal)
- Monitoring the system



# NGIs in Europe

[www.eu-egi.eu](http://www.eu-egi.eu)



[Laurence.Field@cern.ch](mailto:Laurence.Field@cern.ch)

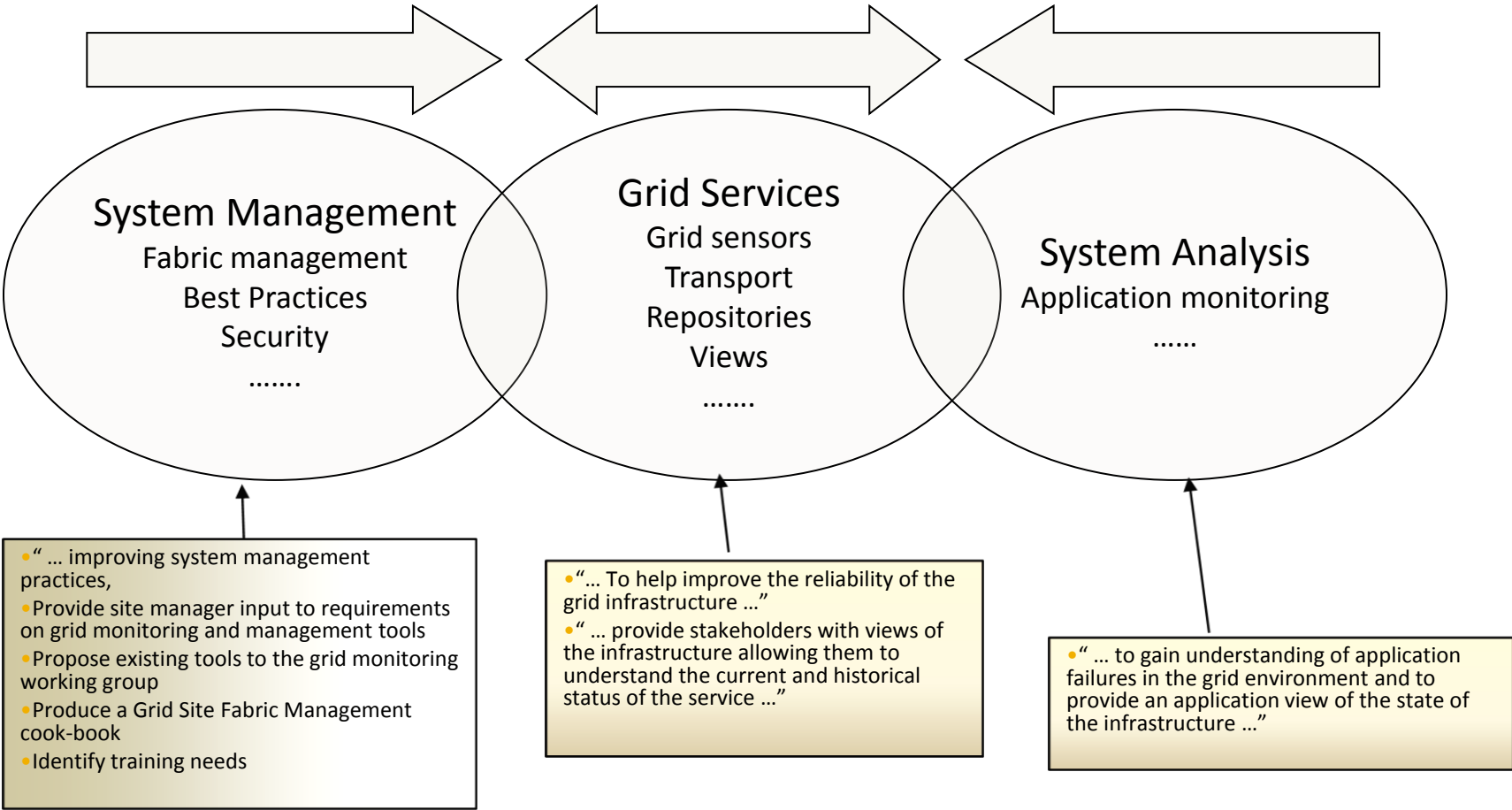


- Daily WLCG Operations Meetings
  - 30 minutes
  - Follow up on current problems
- WLCG T1 Service Coordination meeting
  - Every two weeks
  - Operational Planning
  - Incidents follow-up
- Detailed monitoring of the SLAs



# Grid Monitoring

- The critical activity to achieve reliability

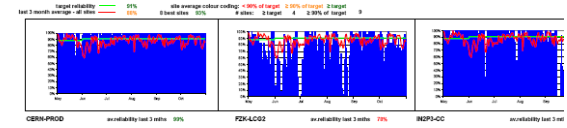


- 




May 2007 - October 2007


Data from SAM monitoring. Plots show *Reliability* calculated as  $\text{Availability} / \text{Scheduled\_Availability}$   
Target reliability is 88% to end May, 91% from June




| Site   | OPS  | ALICE |       |      | ATLAS |      |      | CMS  |     | LHCb |  |
|--------|------|-------|-------|------|-------|------|------|------|-----|------|--|
|        | SAM  | SAM   | AGENT | SAM  | GANGA | PROD | SAM  | CRAB | SAM | PI   |  |
| ASGC   | 93%  | -     | -     | 98%  | 22%   | 82%  | 95%  | 90%  | -   | -    |  |
| BNL    | 91%  | -     | -     | 72%  | 0%    | 0%   | -    | -    | -   | -    |  |
| CERN   | 100% | 97%   | 99%   | 100% | 50%   | 92%  | 100% | 76%  | 96% | 9    |  |
| CNAF   | 80%  | 97%   | 53%   | 85%  | 52%   | 74%  | 100% | 97%  | 66% | 9    |  |
| FNAL   | 89%  | -     | -     | -    | -     | -    | 38%  | 99%  | -   | -    |  |
| FZK    | 91%  | 95%   | 96%   | 62%  | 73%   | 93%  | 99%  | 96%  | 91% | 9    |  |
| IN2P3  | 70%  | 45%   | 89%   | 26%  | 77%   | 79%  | 8%   | 99%  | 97% | 9    |  |
| NDGF   | 97%  | 0%    | 0%    | 76%  | 0%    | 84%  | 0%   | 0%   | -   | -    |  |
| NIKHEF | 92%  | 96%   | 100%  | 92%  | 45%   | 84%  | 53%  | -    | 90% | 19%  |  |
| PIC    | 93%  | -     | -     | 100% | 7%    | 61%  | 100% | 100% | 93% | 88%  |  |
| RAL    | 90%  | 96%   | 99%   | 100% | 15%   | 93%  | 100% | 90%  | 97% | 90%  |  |
| TRIUMF | 95%  | -     | -     | 98%  | 4%    | 94%  | -    | -    | -   | -    |  |



CERN-PROD



BNL-T1



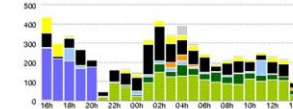
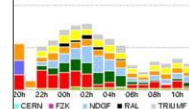
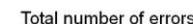
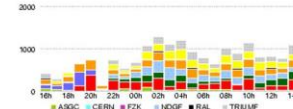
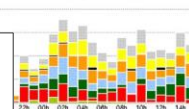
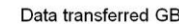
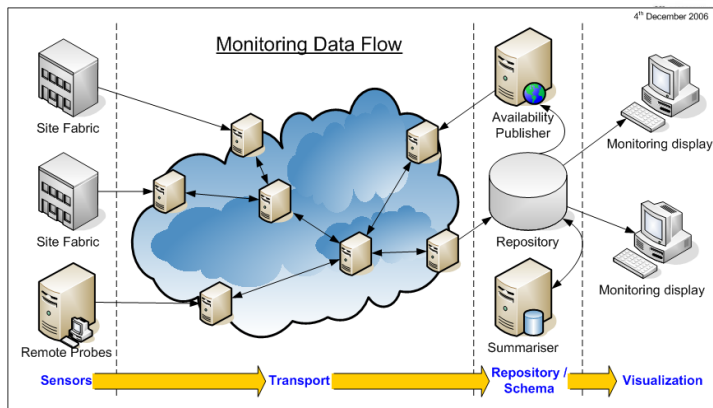
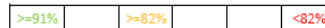
NDGF

SARA-MA

pic

RAL-LC

TRIUMF

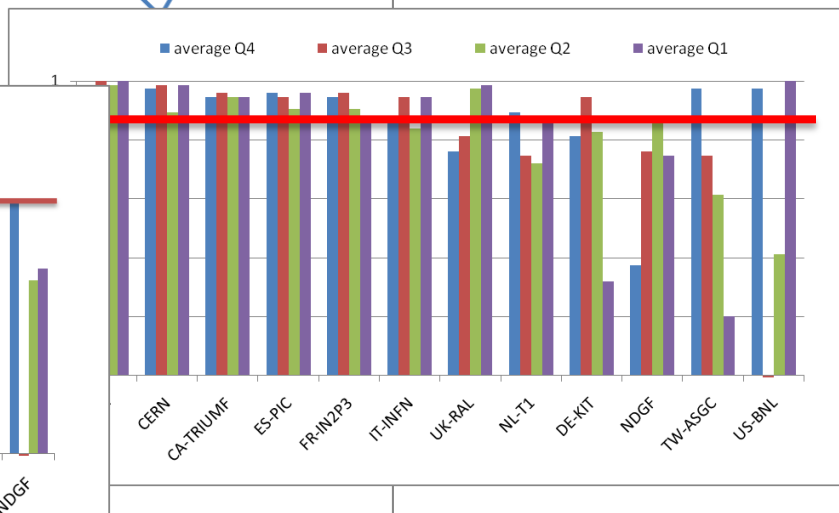
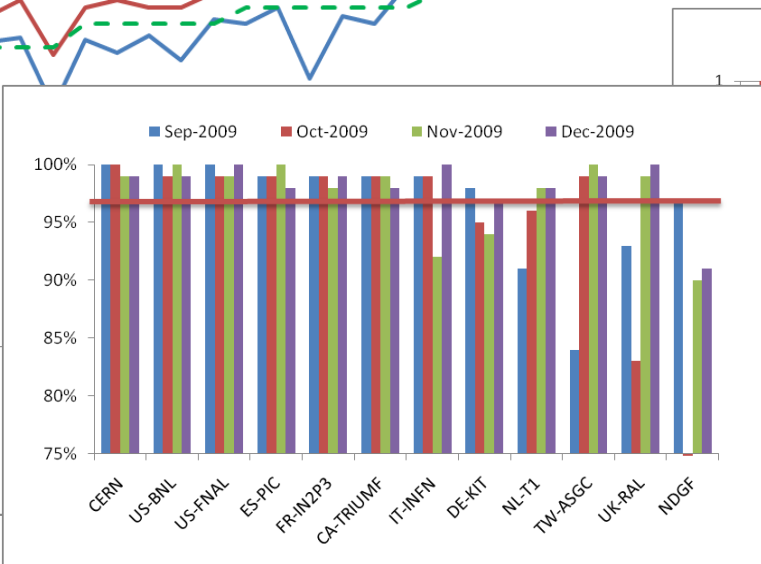
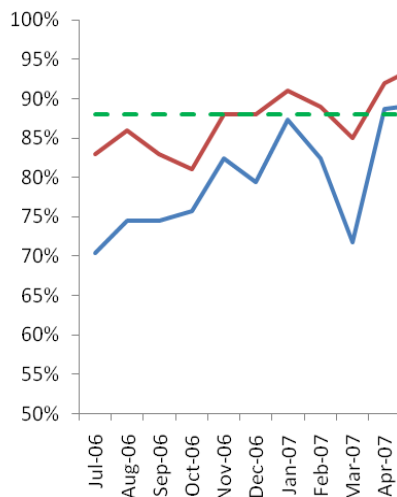


ATLAS M4 Data Monitoring – August 31

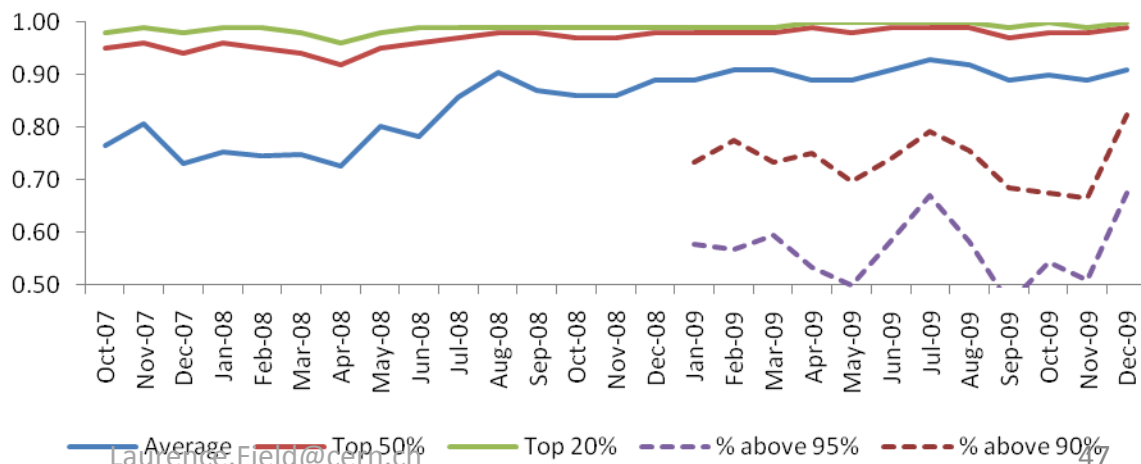


# Reliabilities

## Site Reliability: CERN + Tier 1s



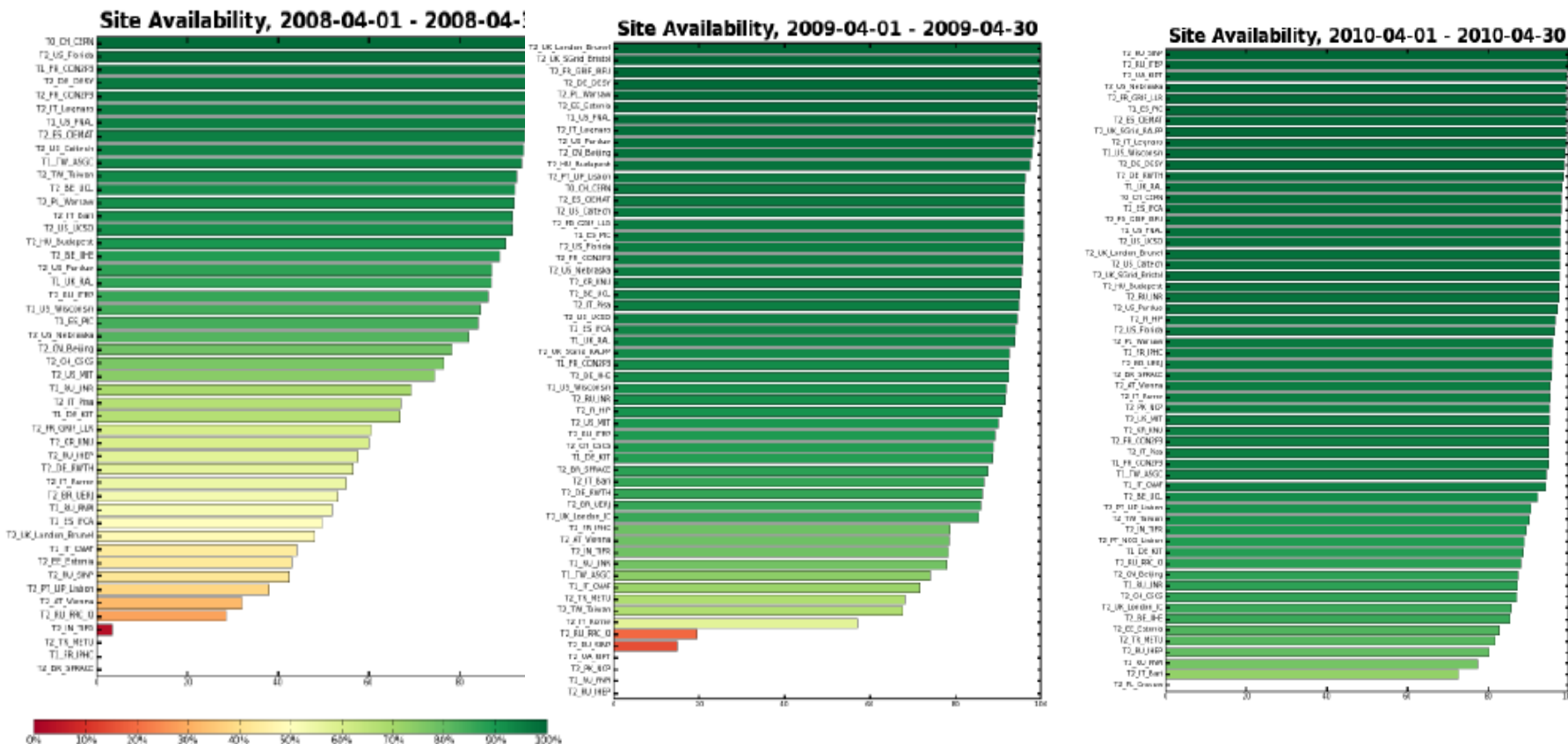
## Tier 2 Reliabilities



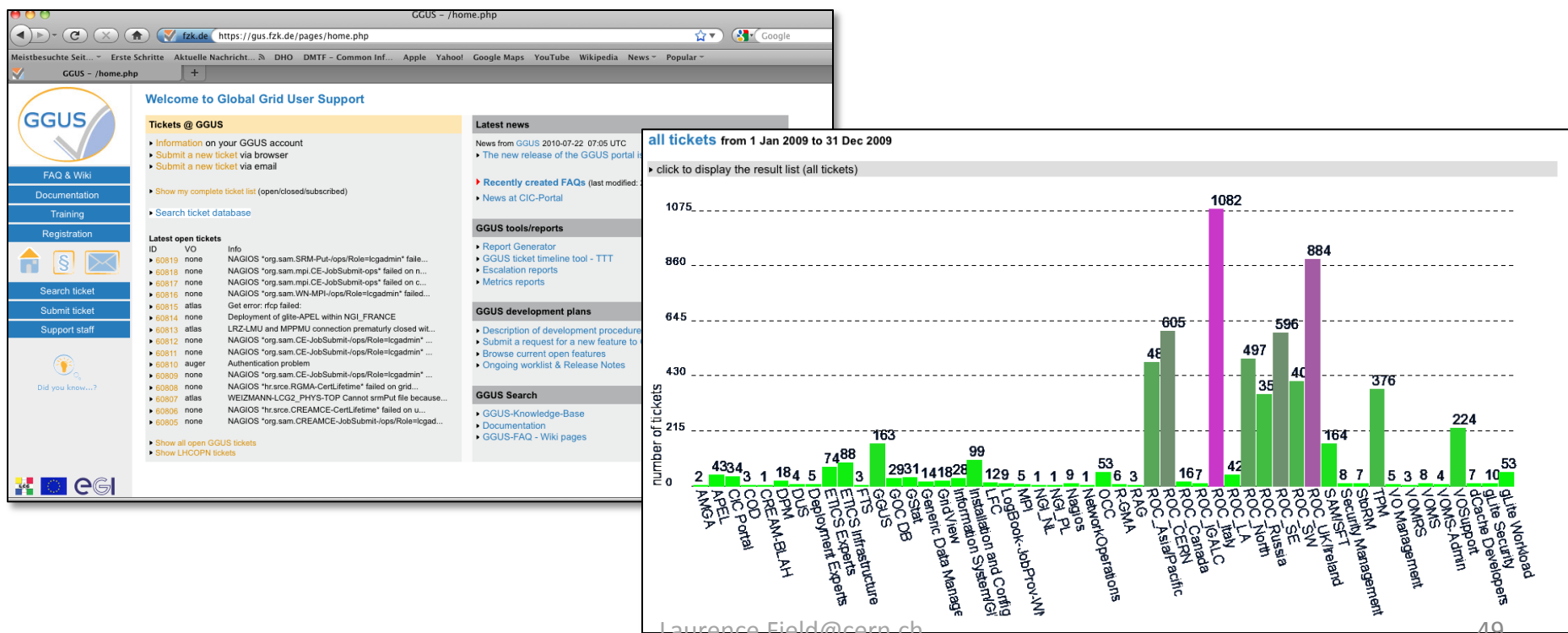
- This is not the full picture:
- Experiment-specific measures give complementary view
- Need to be used together with some understanding of underlying issues



# Improving The Quality



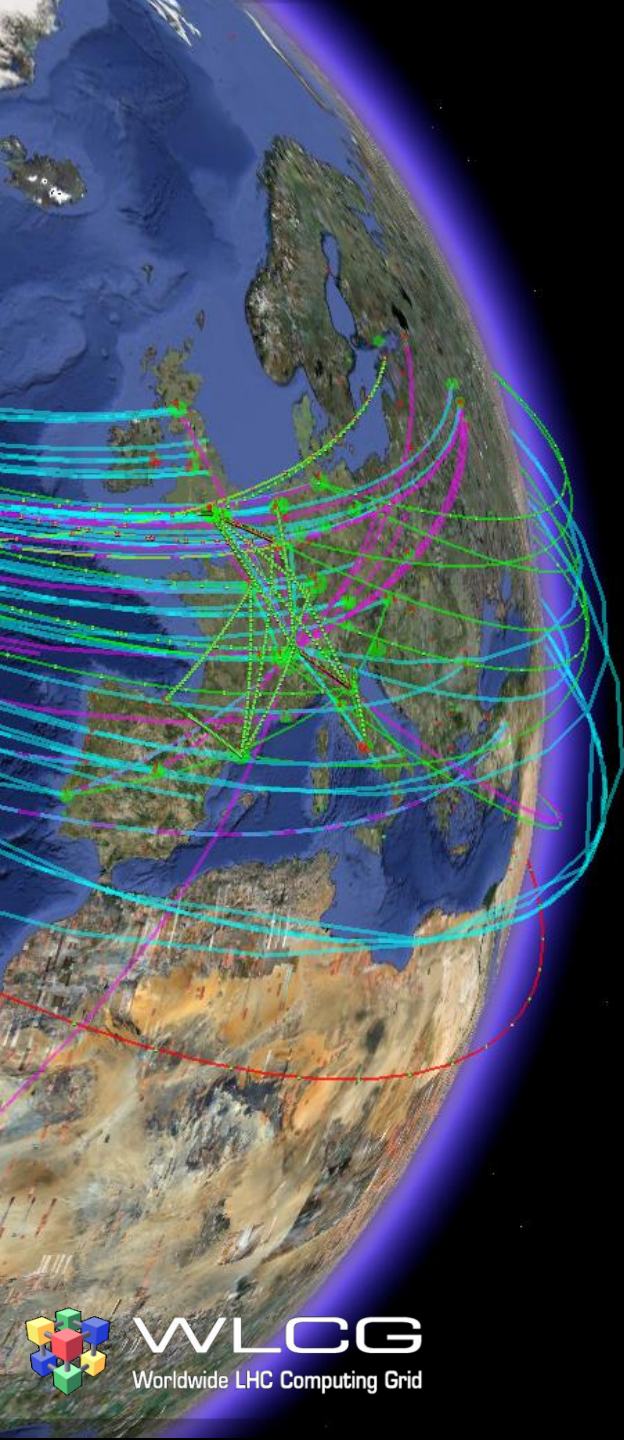
- GGUS: Web based portal
  - About 1000 tickets per months
  - Grid security aware
  - Interfaces to regional/national support structures





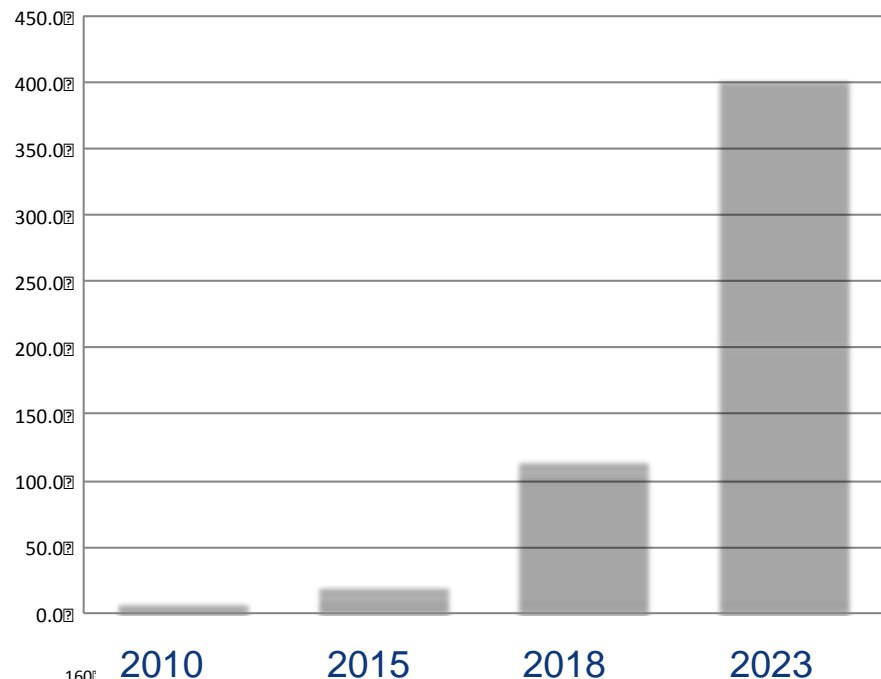
- Reduce operational overhead
  - Self-supporting WLCG Tiers
    - No need for external funds for operations
- Zero configuration
  - For both pledged and opportunistic resources
- Implications
  - Must simplify the grid model (middleware)
    - As thin a layer as possible
  - Make service management lightweight
  - Centralize key services at a few large centres

# The Future





# Scale of challenge



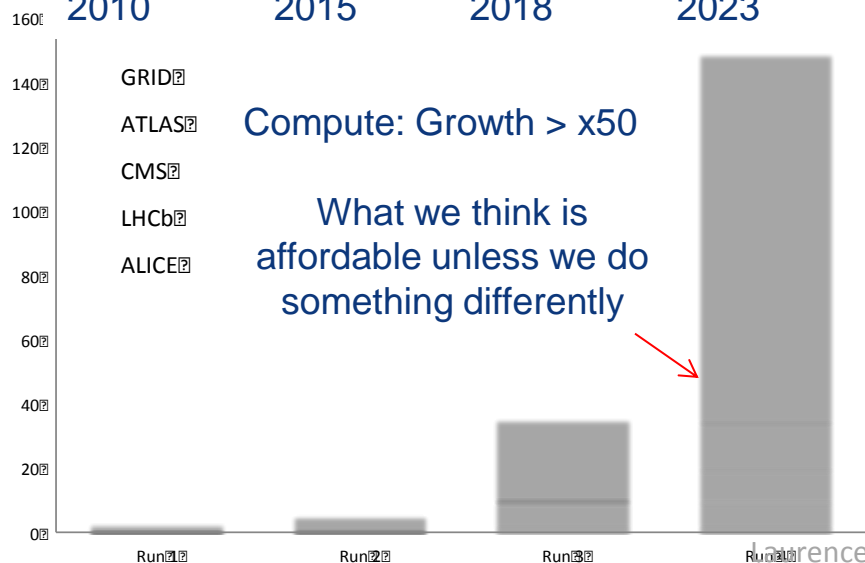
CMS  
ATLAS  
ALICE  
LHCb

- Computing challenge
  - Will “double” next run
  - Then explode thereafter

- Experiment upgrades
- High luminosity

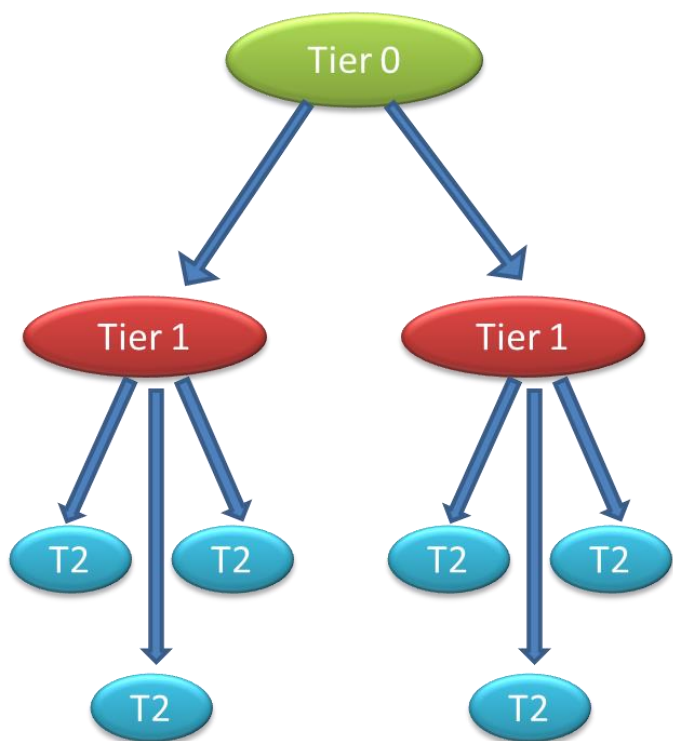
- Two solutions

- More efficient usage
  - Better algorithms
  - Better data management
- More resources
  - Opportunistic
  - Volunteer
- Move with technology
  - Clouds
  - Processor architectures



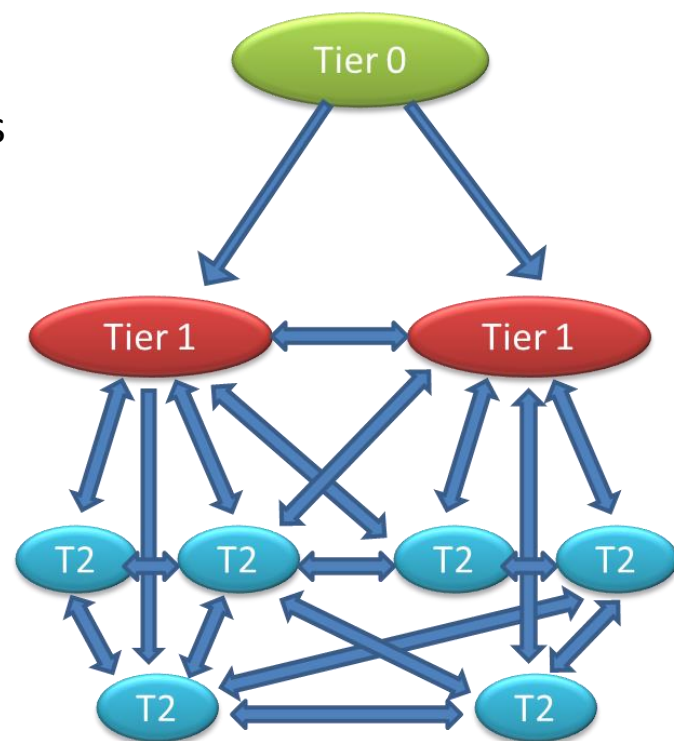


# Computing Model Evolution



Hierarchy

Evolution of  
computing models

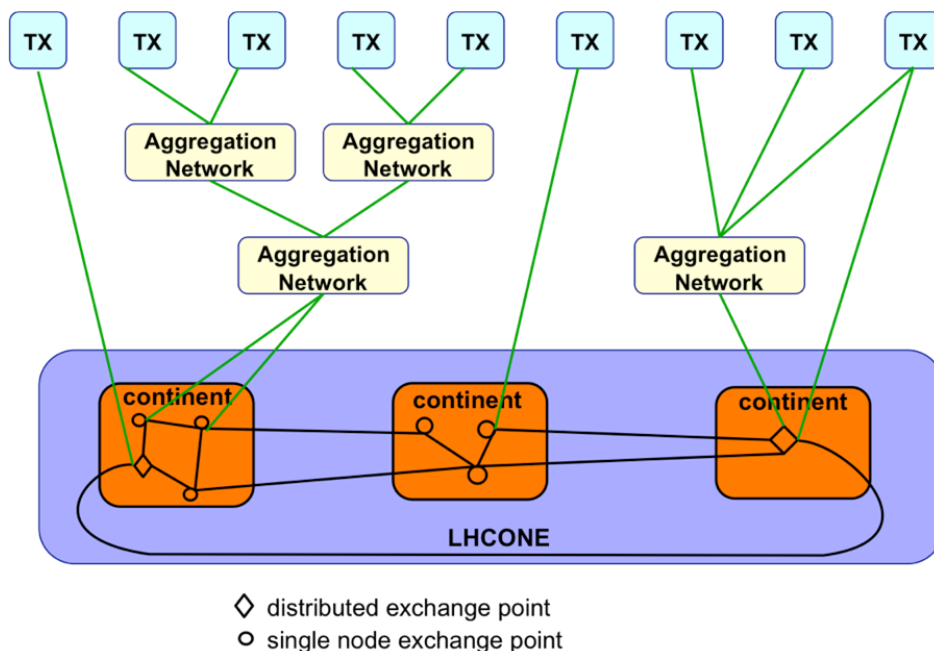
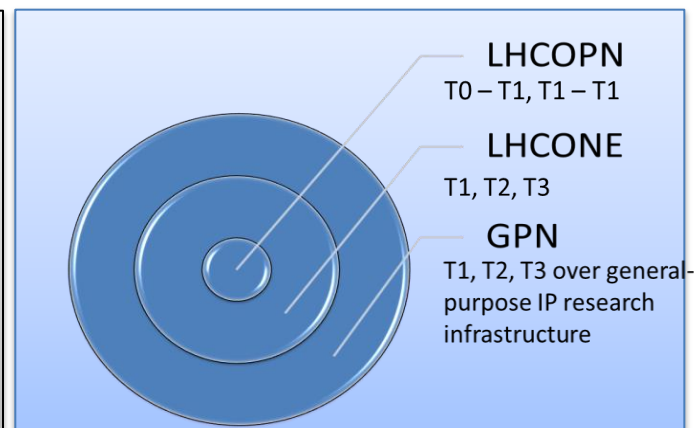


Mesh



Evolution of computing models also require evolution of network infrastructure

- Enable any Tier 2, 3 to easily connect to any Tier 1 or 2

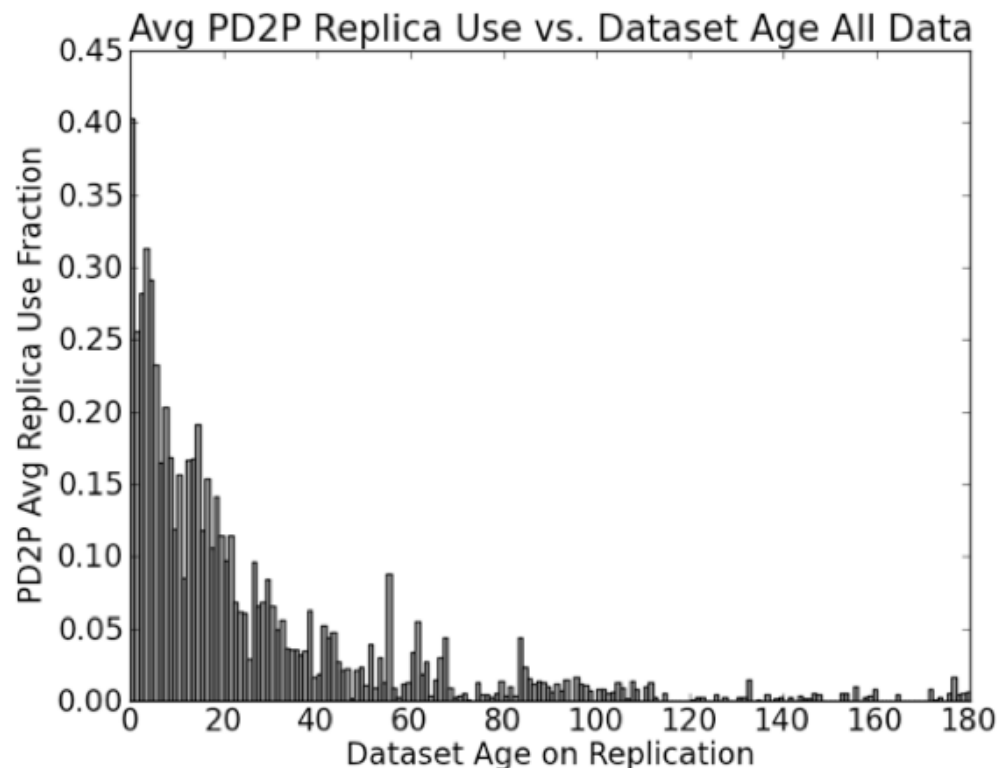
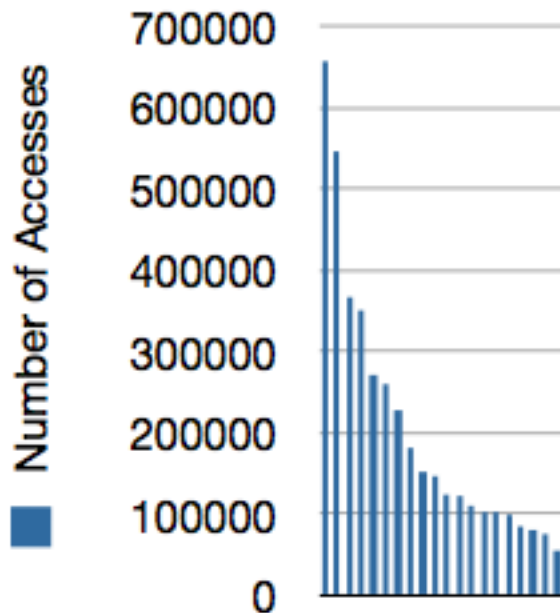


- Use of Open Exchange Points
- Do not overload the general R&E IP infrastructure with LHC data
- Connectivity to T1s, T2s, and T3s, and to aggregation networks: NRENs, GÉANT, etc.



# Data Popularity

- Usage of data is highly skewed
- Dynamic data placement can improve efficiency
- Data replicated to T2s at submission time (on demand)



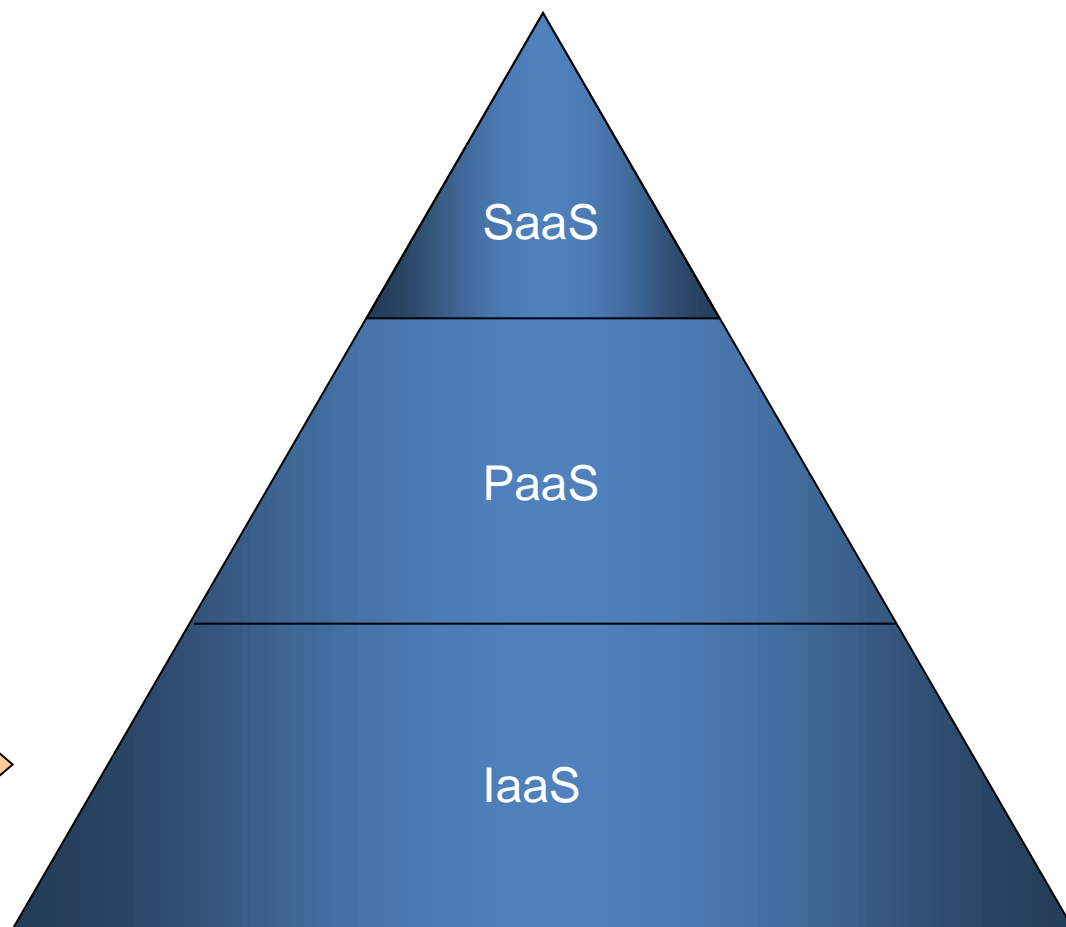
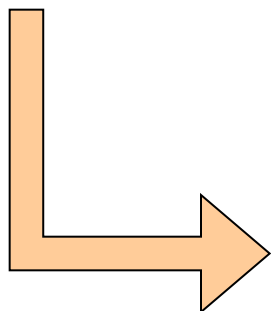


# Storage Federations

- Transparent access to distributed resources
  - through a unique namespace.
- Advantages
  - Resilience
    - Jobs will not fail due to unavailable data as another replica will be found
  - Overflow
    - Send jobs to a data-less site with free CPU
  - Storage efficiency
    - Fewer replicas of data need
  - Transparency
    - All data available through a single namespace
- Experiments expect 10% of the access may be this way



VMs on demand

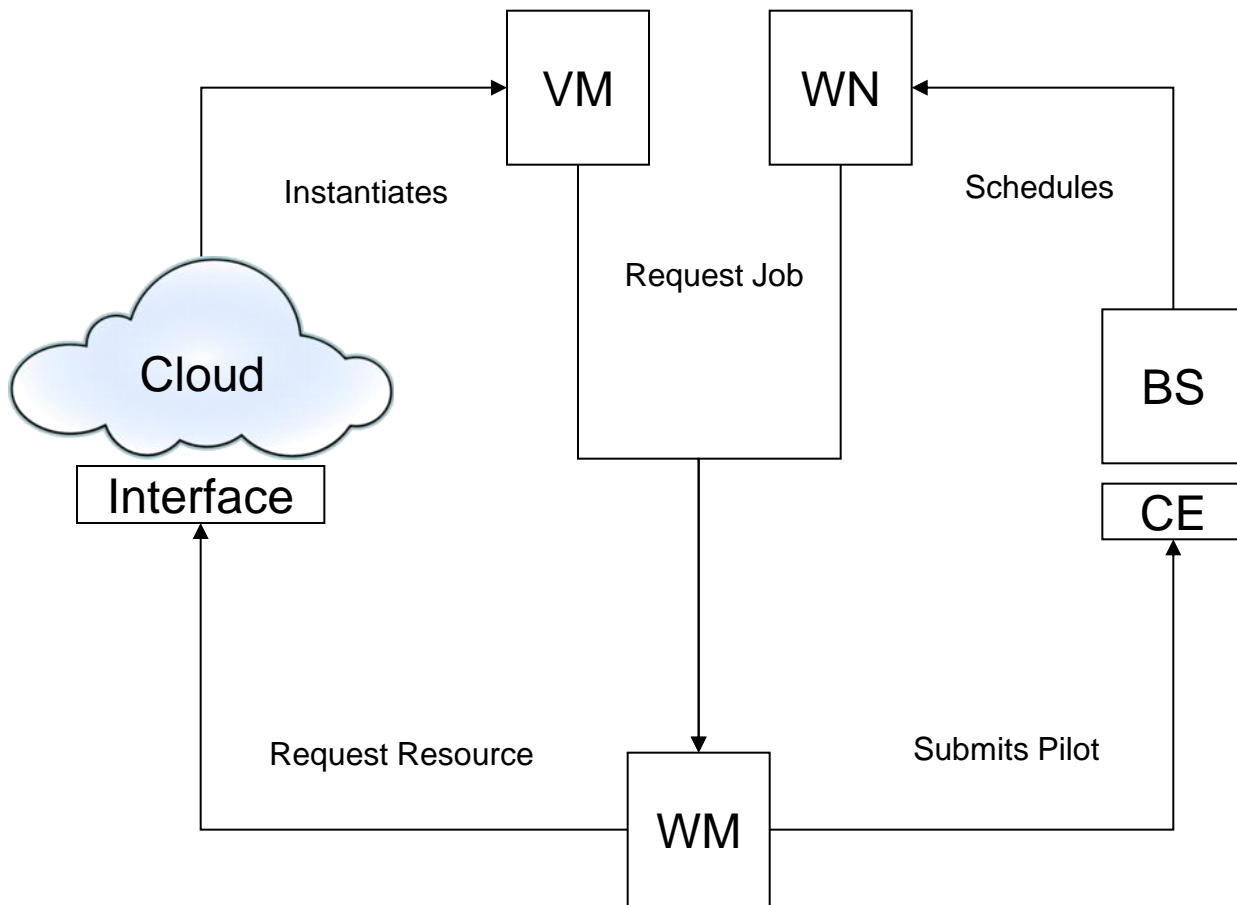




- General solution
  - Originated and supported outside of HEP
- Delivered as a metered service
  - Commercial providers
    - Sustainability
  - Mature SLAs
  - Opportunistic use
    - Simplified and broad approach
- Many sites are deploying cloud stacks internally
  - OpenStack, OpenNebula, ...
- Experiments have used many cloud instances
  - WLCG sites
  - HLT farms
  - Helix Nebula
  - Commercial providers
- Utility Computing?



# High-level View





- Image Management
- Capacity Management
- Monitoring
- Accounting
- Pilot Job Framework
- Supporting Services



## About ATLAS@Home

ATLAS@Home is a research project that uses volunteer computing to run simulations of the ATLAS experiment at CERN. You can participate by downloading and running a free program on your computer.

ATLAS is a particle physics experiment taking place at the Large Hadron Collider at CERN, that searches for new particles and processes using head-on collisions of protons of extraordinary high energy. Petabytes of data were recorded, processed and analyzed during the first three years of data taking, leading to up to 300 publications covering all the aspects of the Standard Model of particle physics, including the discovery of the Higgs boson in 2012.

Large scale simulation campaigns are a key ingredient for physicists, who permanently compare their data with both "known" physics and "new" phenomena predicted by alternative models of the universe, particles and interactions. This simulation runs on the WLCG Computing Grid and at any one point there are around 150,000 tasks running. You can help us run even more simulation by using your computer's idle time to run these same tasks.

No knowledge of particle physics is required, but for those interested in more details, at the moment we simulate the creation and decay of supersymmetric bosons and fermions, new types of particles that we would love to discover next year, as they would help us to shed light on the dark matter mystery!

The program you will download runs simulation software inside a virtual machine hosted by your computer. The virtual machine image is ~500MB but is only downloaded once. Each workunit downloads a small set of input data and runs for approx 1 to 2 hours depending on the computer's processor speed.

THE ATLAS@HOME PROJECT IS STILL UNDER DEVELOPMENT, and it cannot be guaranteed that jobs will be free from errors or that there will always be work available, but do not hesitate to contact us at [atlas.home@cern.ch](mailto:atlas.home@cern.ch)

## Join ATLAS@Home

- **Read our rules and policies**
- This project uses BOINC. If you're already running BOINC, select Add Project. If not, download BOINC.
- This project also requires VirtualBox to be installed.
  - Windows: VirtualBox is included in Windows distributions of Boinc and so does not have to be installed separately.
  - Mac and Linux: VirtualBox must be installed separately
- In case of problems using the latest version of VirtualBox try using version 4.2.10.
- After choosing Add Project in Boinc Manager, choose ATLAS@Home from the list of projects
- Then either enter details for to set up a new account, or if you have previously registered enter the existing account details.
- **IMPORTANT:**
  - A reasonably powerful modern 64-bit computer with at least 4GB of memory is required to run ATLAS@Home. Enabling 64-bit virtualisation may require some changes in BIOS settings.
  - In Boinc Manager set the % of processors used to be at most 50%, otherwise your computer may run out of memory.
  - The work units may use a lot of network bandwidth, so a slow internet connection may reduce work unit efficiency
- If you're running a command-line version of BOINC, create an account first.
- If you have any problems, [get help here](#).

LHC@home Portal SixTrack vLHC@home (Test4Theory)



LHC@home is a platform for volunteers to help physicists develop and exploit particle accelerators like CERN's Large Hadron Collider, and to compare theory with experiment in the search for new fundamental particles.

By contributing spare processing capacity on their home and laptop computers, volunteers may run simulations of beam dynamics and particle collisions in the LHC's giant detectors.



## Homepage

### The Sixtrack project

Help us to study the LHC machine and its upgrade to understand the fundamental laws of the universe.

[View details >](#)

### The vLHC@home project

Help us to do research about the elusive Higgs particle with our virtual atom smasher. (formerly known as Test4Theory)

[View details >](#)



## Project Partners



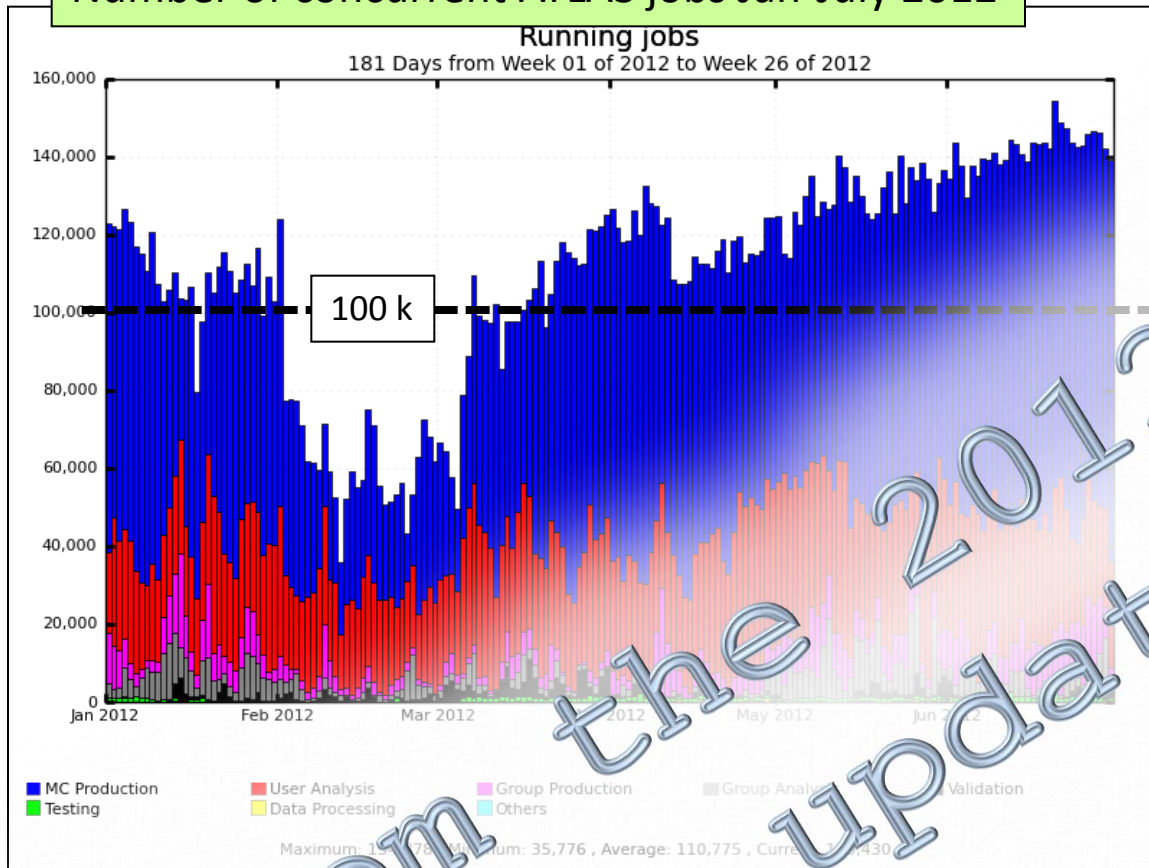
Copyright CERN, 2013 -- European Organization for Nuclear Research  
With the support of the Citizen Science Centre and the LHC Physics Centre at CERN

[Login Contact us](#)

Powered by Drupal, an open source content management system

It would have been impossible to release physics results so quickly without the outstanding performance of the Grid (including the CERN Tier-0)

### Number of concurrent ATLAS jobs Jan-July 2012



Includes MC production,  
user and group analysis  
at CERN, 10 Tier1-s,  
~ 70 Tier-2 federations  
→ > 80 sites

> 1500 distinct ATLAS users  
do analysis on the GRID

- ❑ Available resources fully used/stressed (beyond pledges in some cases)
- ❑ Massive production of 8 TeV Monte Carlo samples
- ❑ Very effective and flexible Computing Model and Operation team → accommodate high trigger rates and pile-up, intense MC simulation, analysis demands from worldwide users (through e.g. dynamic data placement)