

Running jobs: 236092
Transfer rate: 11.41 GiB/sec

Grid computing and forward

Catherine Biscarat
biscarat@lpsc.in2p3.fr

**Laboratoire de Physique Subatomique
et de Cosmologie de Grenoble**

IN2P3/CNRS, Université Grenoble-Alpes, INPG (France)

ESIPAP, the European School in Instrumentation
for Particle and Astroparticle Physics
March 10th 2014, Archamps, France

ographer

/BKG
NGA, GEBCO

Running jobs: 236092
Transfer rate: 11.41 GiB/sec

Today, Together

- Why grid computing

A success story: the grid for the LHC

- Other Grids

- Behind the scenes

Technical details

- Going forward

Standards, simplicity, clouds

- Accessing the grid



ographer

/BKG

NGA, GEBCO

A few words on the speaker

- Before 2011: experimental particle physicist (colliders)



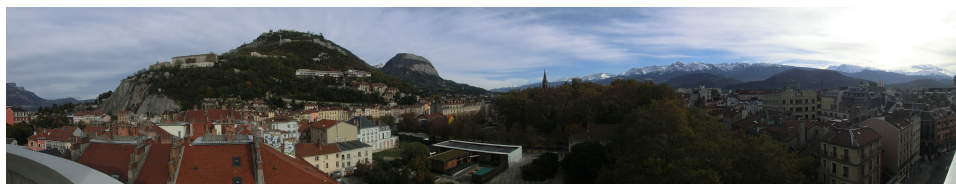
Detector in US, Chicago, ppbar collider Tevatron $\sqrt{s} = 2 \text{ TeV}$
- calorimetry, simulation, SUSY search



Detector at CERN, Geneva, pp collider LHC $\sqrt{s} = 14 \text{ TeV}$
- calorimetry, non standard Higgs boson search
- responsible for the ATLAS computing activities in a major centre (T1)

- Since 2011: Research engineer in computing at IN2P3 (National Institute for Nuclear and Particle Physics in France), part of CNRS (National Centre for Scientific Research)

- Grid computing
- Application porting



Running jobs: 236092
Transfer rate: 11.41 GiB/sec



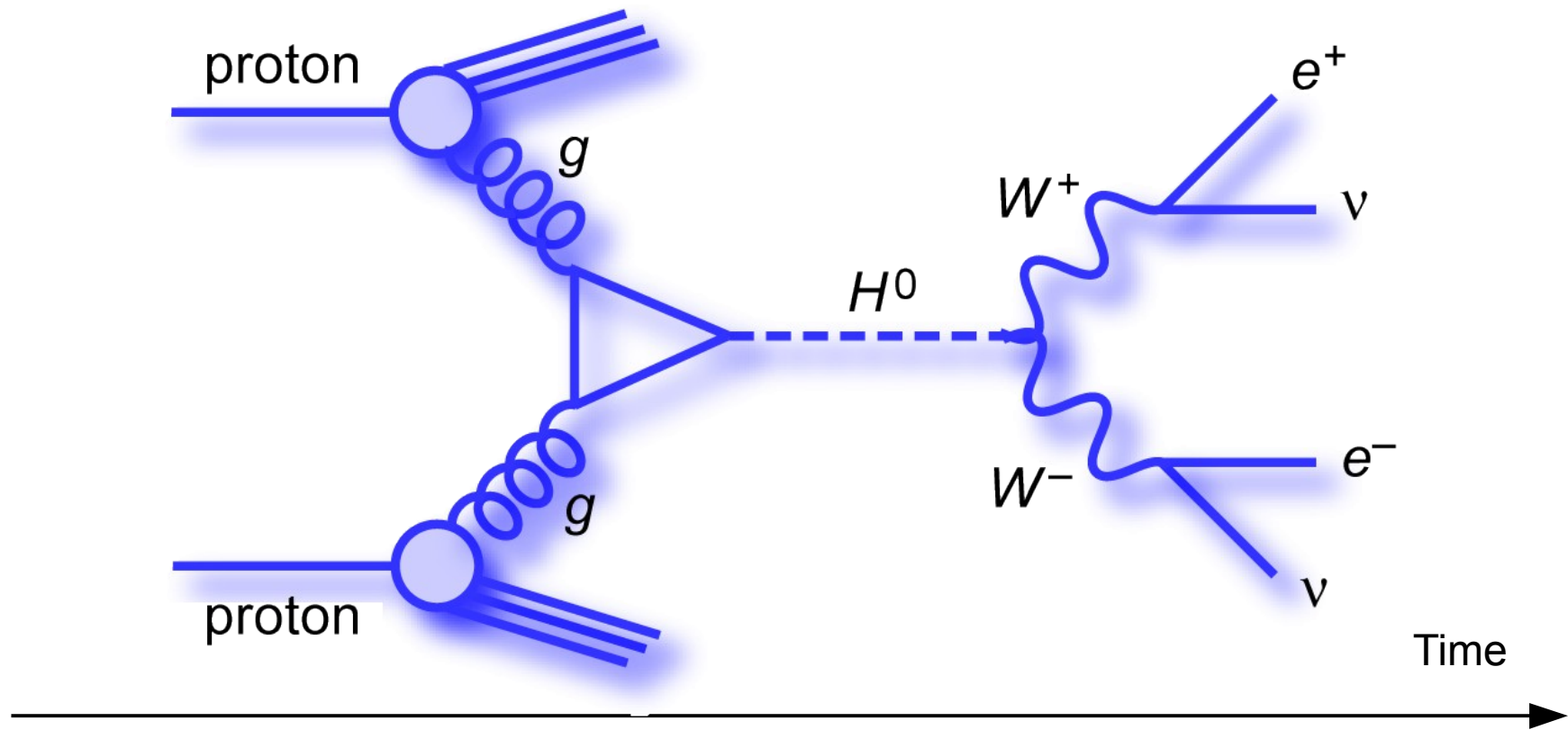
Why grid computing ?

ographer

/BKG
NGA, GEBCO

Searching for the Higgs

The Higgs boson could be produced in the collision of two protons



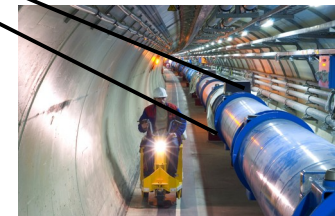
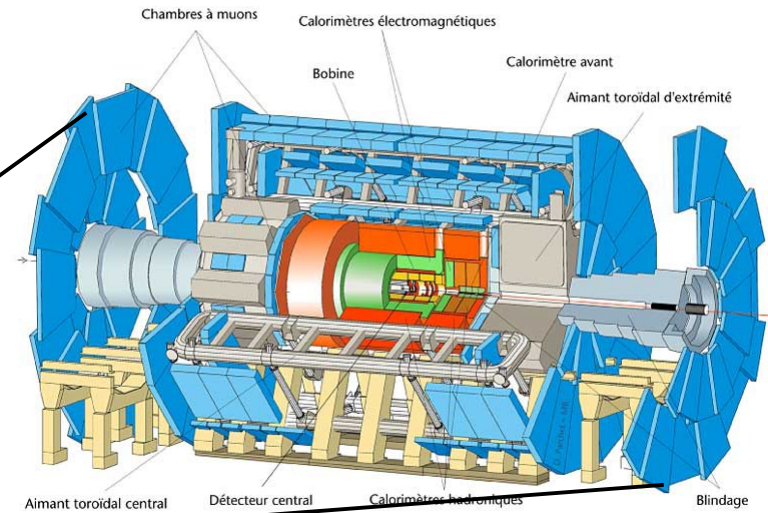
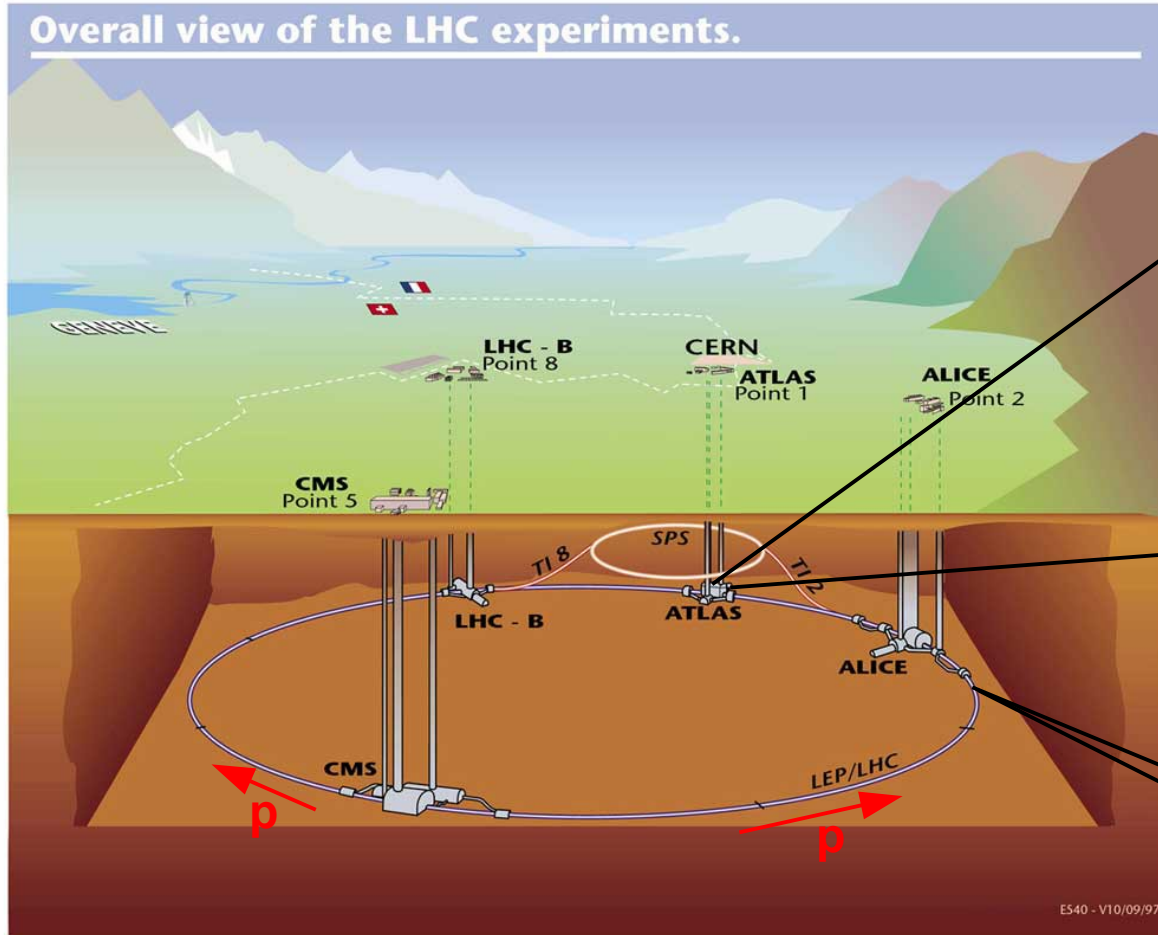
Two protons collide

Two gluons (constituents
Of the protons) do fusion
and a Higgs boson
gets created

The Higgs decays
instantaneously in a W
boson pair (which decay
themselves further)

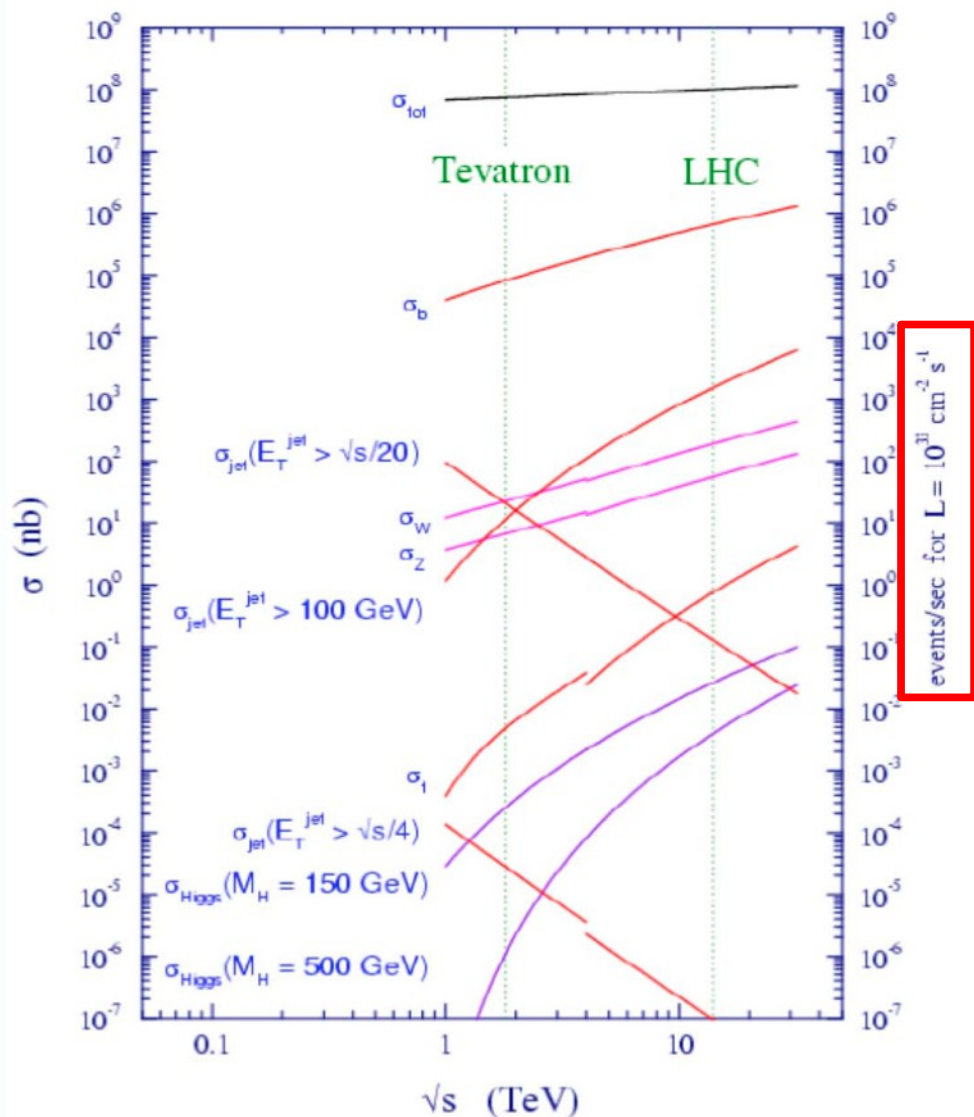
The stable particles
in the final state (here e , ν)
get to the detector

In practice



Event production rate at the LHC

proton - (anti)proton cross sections



- Bunch crossing: 40 MHz
 - 100 000 CDs written / s
 - distance(moon-earth) / 3 months
- Expensive to store
- Difficult to share
- Long to analyse
- But: all events do not carry the same interest for physics
- First selection at the experiment level before storage
 - Trigger (very fast electronic and algorithms)

Quantity of stored data

- Let's do a quick calculation
 - Each experiment stores ~100 events /s
 - Each event counts for ~1-2 MB
- several PB of data produced per year and per experiment
- equivalent to 600 000 DVDs movies produced per experiment and per year

1 MB
1 digital picture
1 dictionnaire
1 CD = 650 MB

1 GB = 1000 MB
5GB = 1 DVD movie

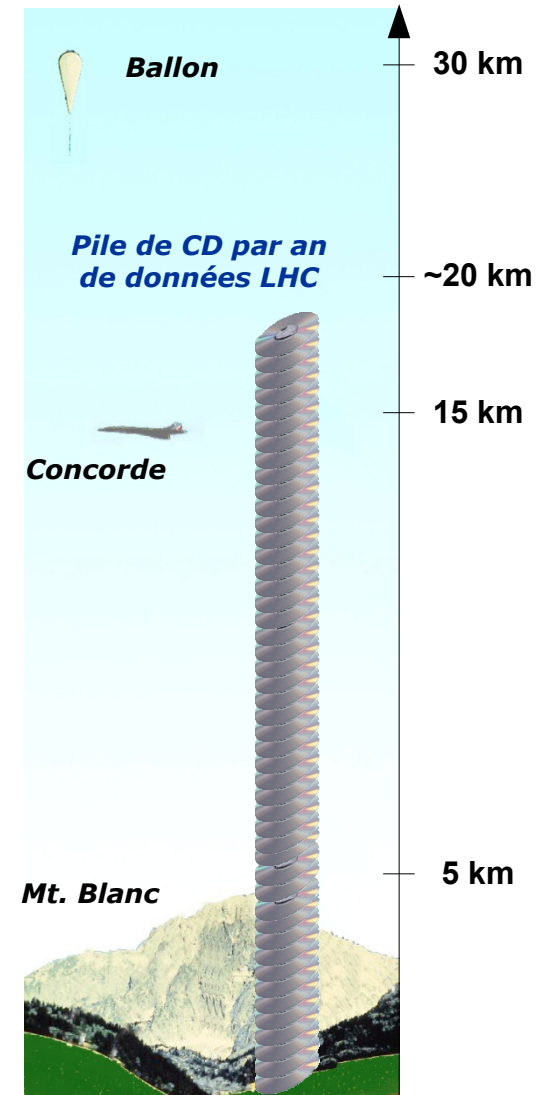
1TB = 1000 GB
Livres produits par an

1PB = 1000 TB
Production par an de
1 expérience LHC

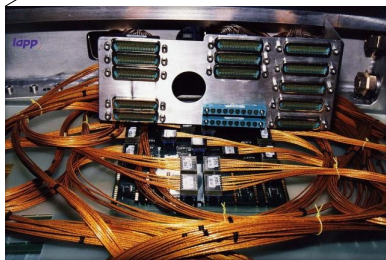
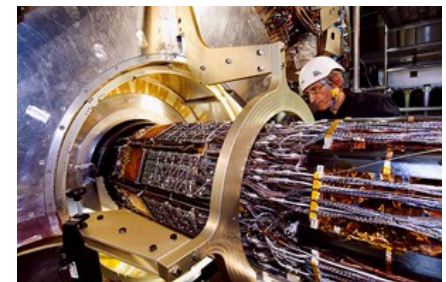
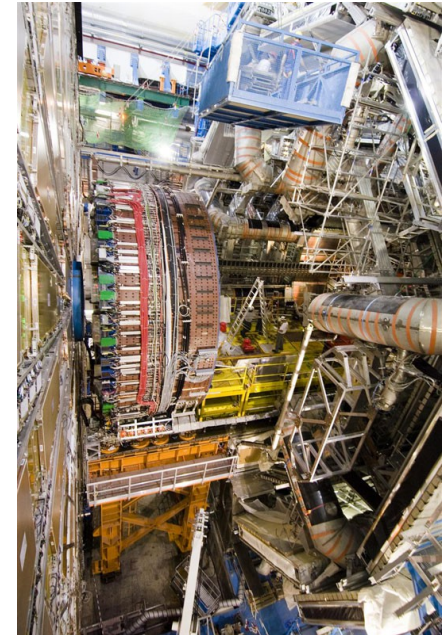
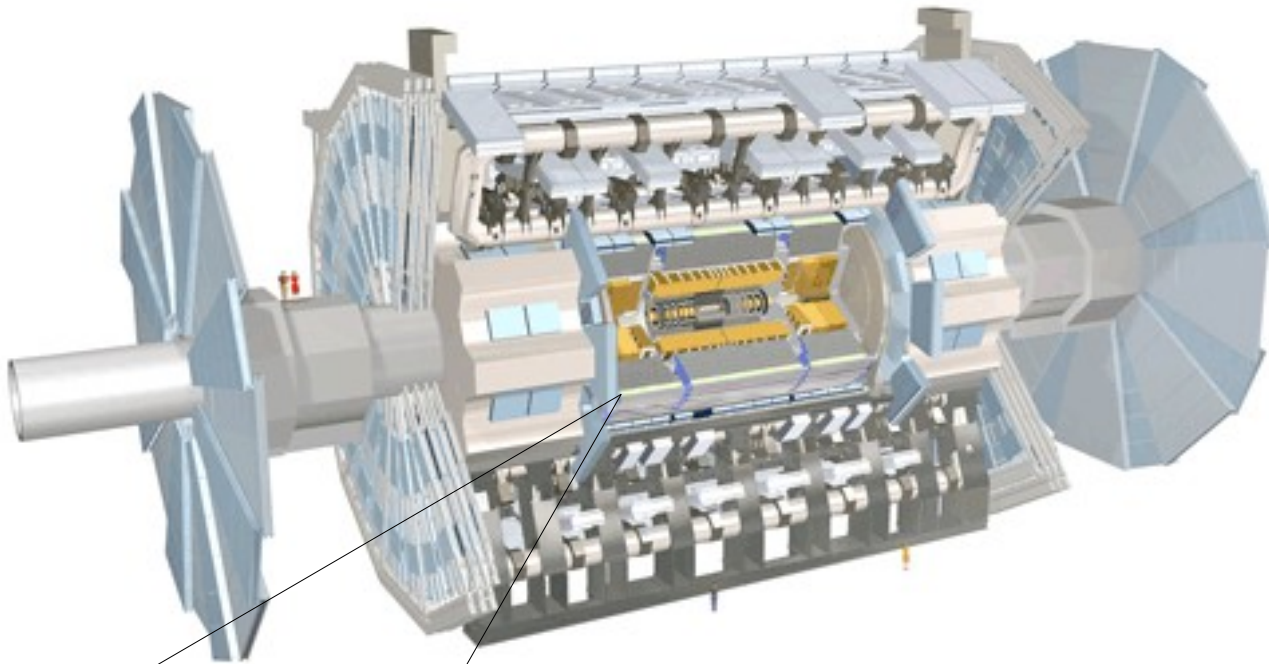
1 EB = 1000 PB
Production mondiale
d'informationen 1 an



Enormous amount of data
to analyse and to store;
No single computing center
could manage it alone.



Raw data



```
101100 101011 010001
110111 001011 001100
100001 111100 100110
110101 110011 100101
001010 101000 001010
111001 100101 000011
010111 001001 010100
100010 010100 101111
100100 101001 001010
000010 100101 111001
```

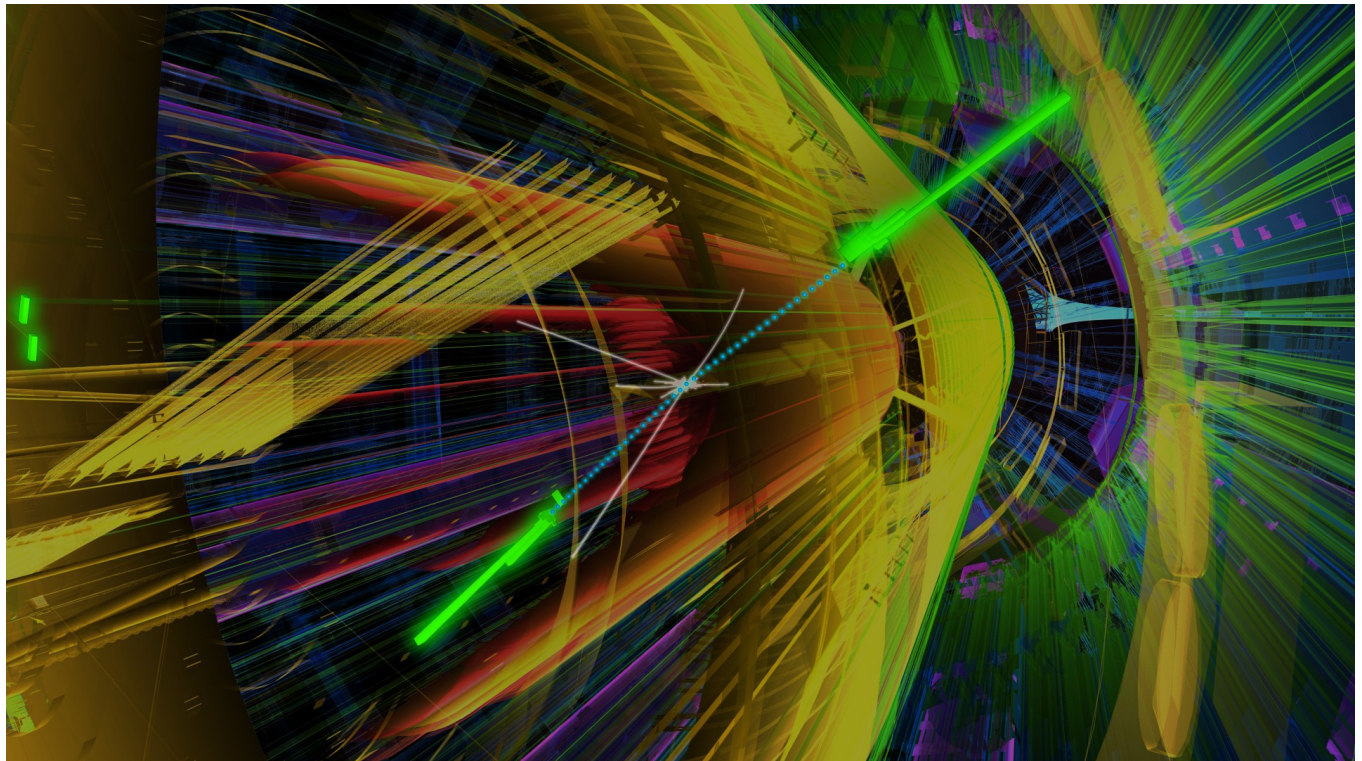
Pixels → yes/no
Calo → tensions en "Volt"

Stored on tape
at CERN



“Events”

- Data are organised in “events”
 - A “picture” of a collision
 - With millions of sensors
- Each event is independent from other events
 - And it is quite small
- Algorithms process events one by one

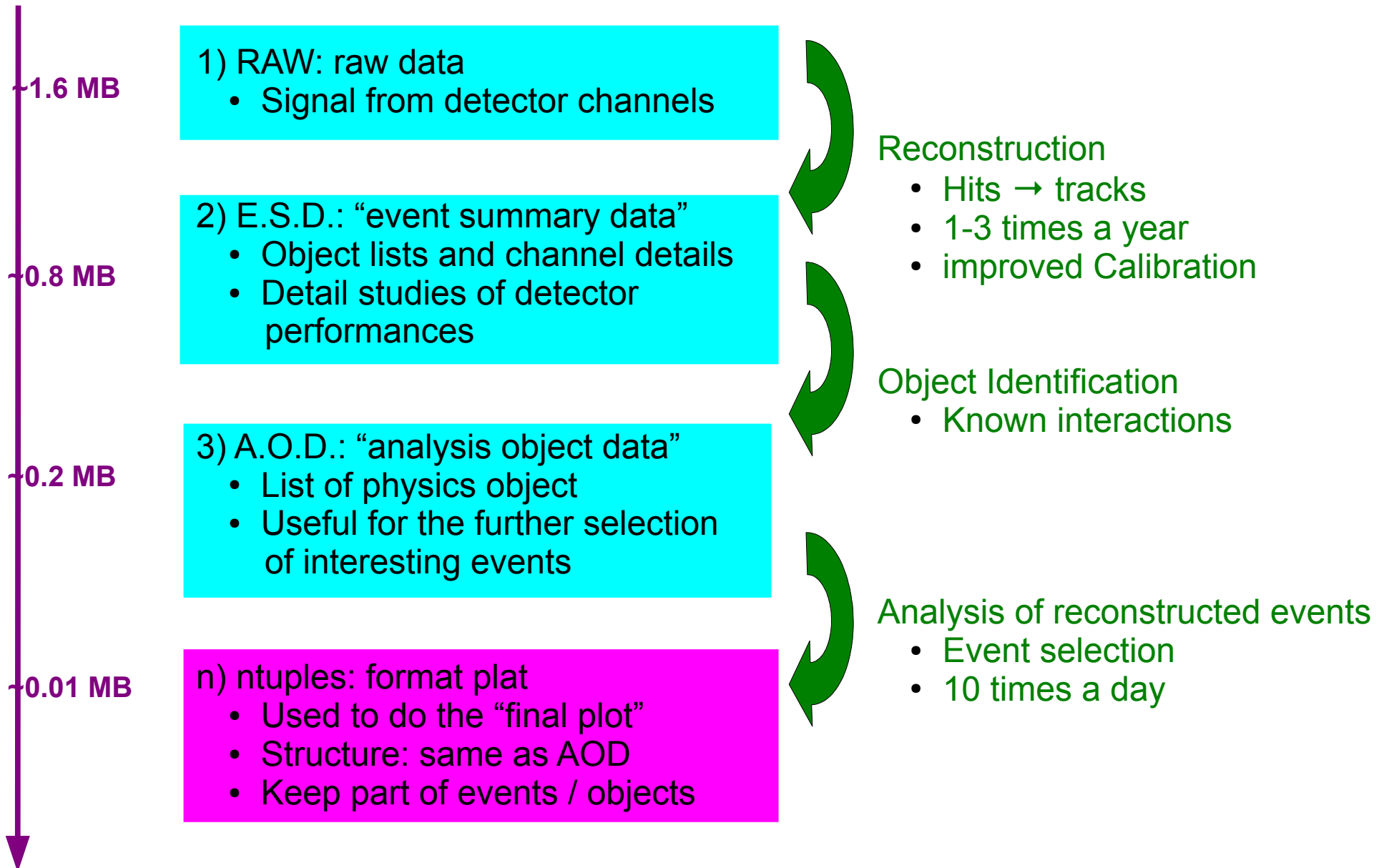


How do we look into data

Size/event

Data Format

Processing step



How do we look into data

Size/event

Data Format

Processing step

1.6 MB

- 1) RAW: raw data
- Signal from detector channels

Reconstruction

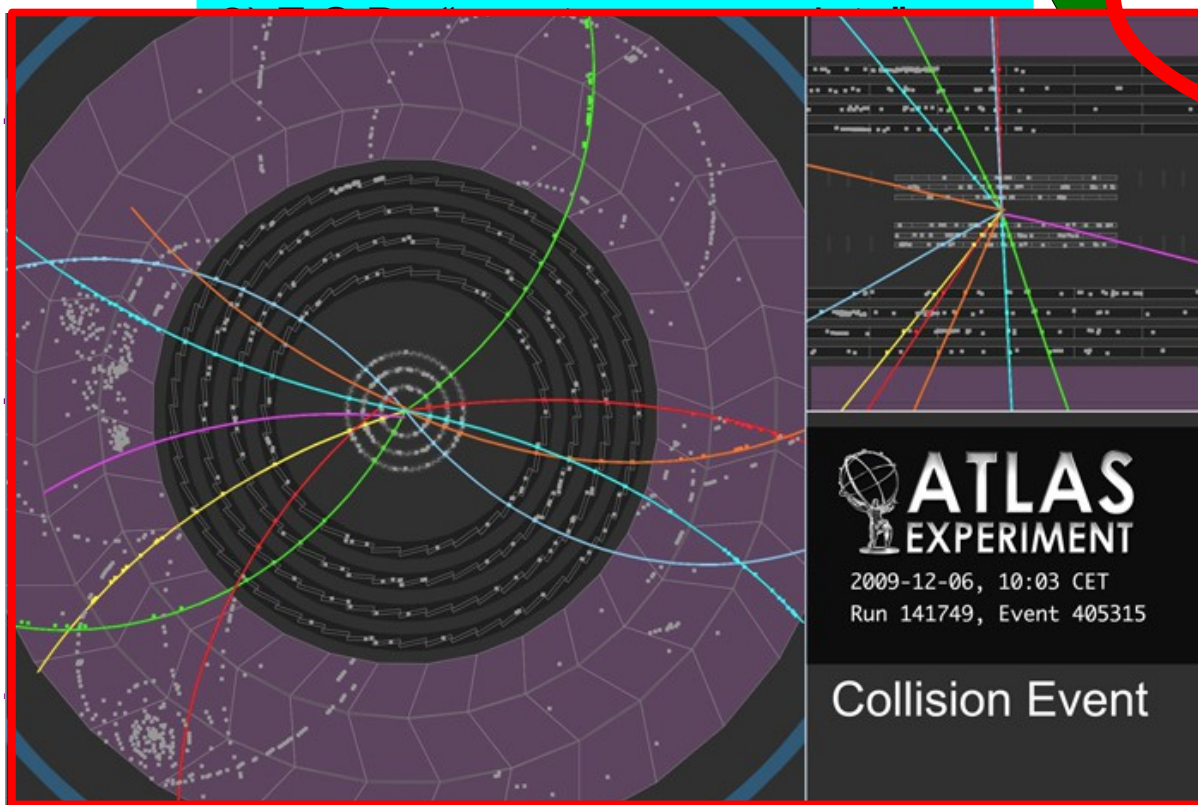
- Hits → tracks
- 1-3 times a year
- improved Calibration

Object Identification

- Known interactions

Analysis of reconstructed events

- Event selection
- 10 times a day



<http://atlas.web.cern.ch/Atlas/public/EVTDISPLAY/events.html>

Keep part of events / objects

How do we look into data

Size/event

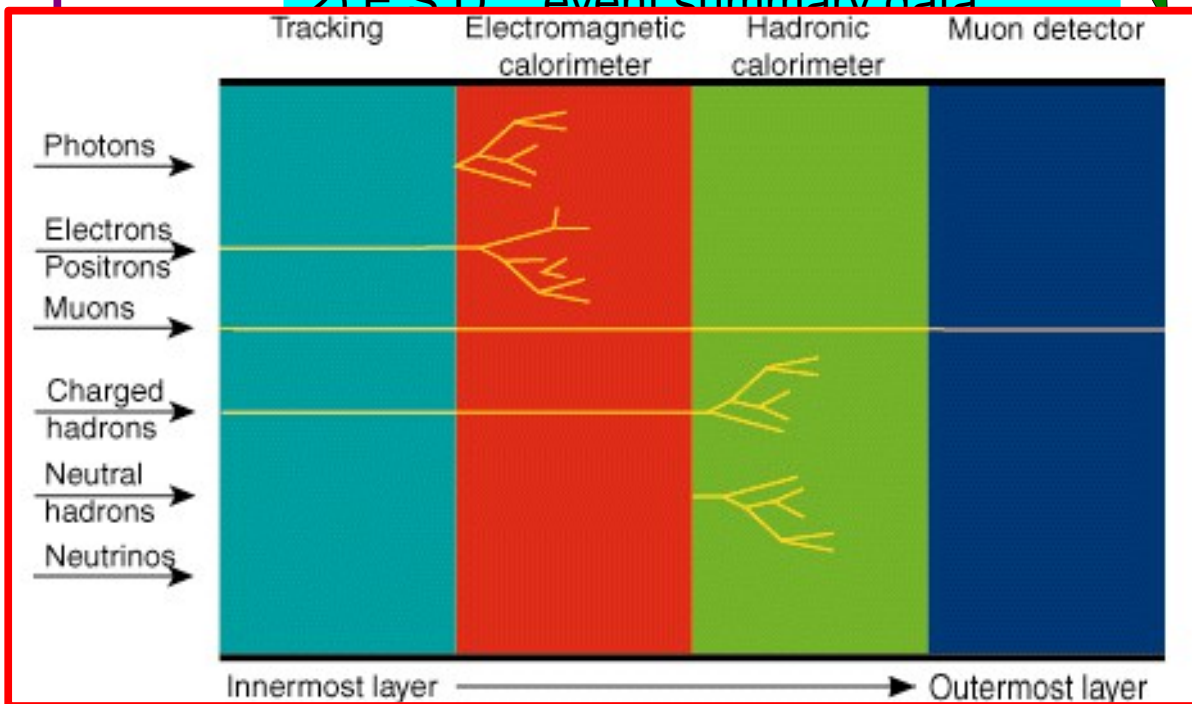
Data Format

Processing step

1.6 MB

- 1) RAW: raw data
- Signal from detector channels

2) E S D: "event summary data"



Reconstruction

- Hits → tracks
- 1-3 times a year
- improved Calibration

Object Identification

- Known interactions

Analysis of reconstructed events

- Event selection
- 10 times a day

- Structure: same as AOD
- Keep part of events / objects

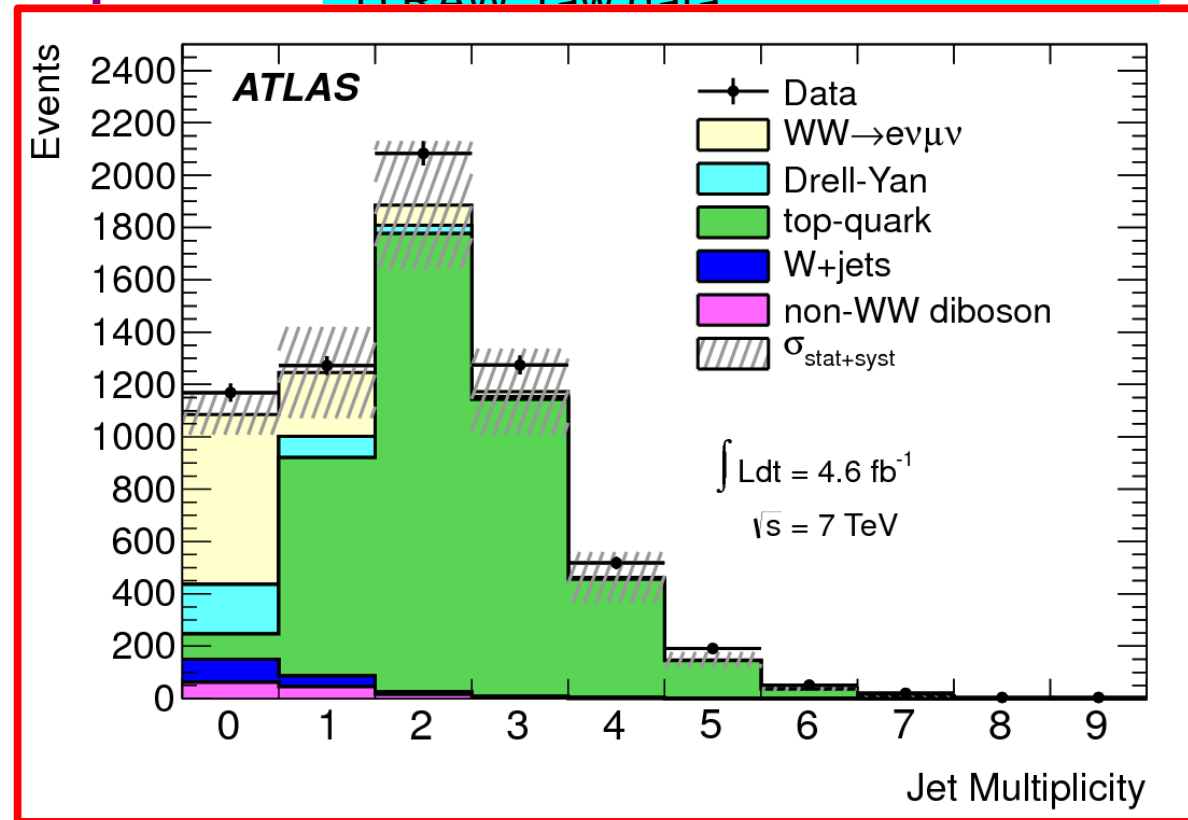
How do we look into data

Size/event

Data Format

Processing step

1) RAW: raw data



Reconstruction

- Hits \rightarrow tracks
- 1-3 times a year
- improved Calibration

Object Identification

- Known interactions

Analysis of reconstructed events

- Event selection
- 10 times a day

0.01 MB

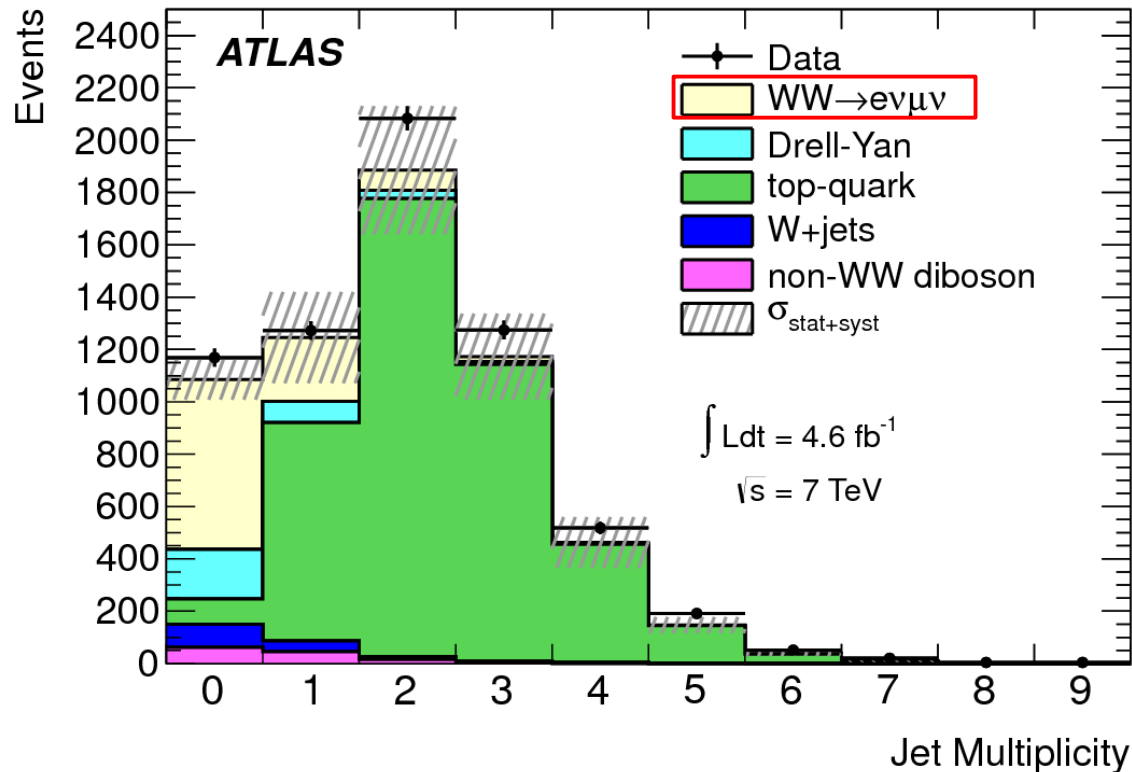
n) ntuples: format plat

- Used to do the “final plot”
- Structure: same as AOD
- Keep part of events / objects

Interpreting data

In other words:

**Compare the data
With a model (MS).**



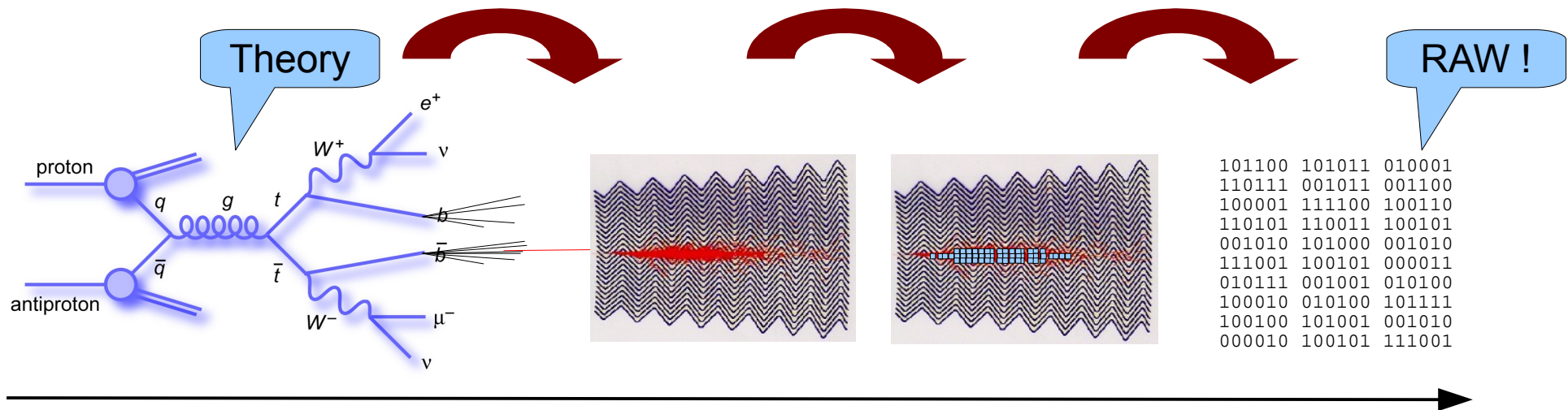
- **Agreement** data/simulation (understanding of the detector)
- Find **deviations** (discovery of new phenomena)

Getting results from an experiment: no way without simulation !

Simulation of RAW data

Three ingredients :

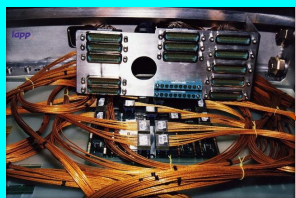
- 1) "Physics" modelisation (collisions, hard process, ...)
- 2) Particle/matter interactions in the detector layout
- 3) Signals transmitted by the detector



Summary

Centralised operations

Reconstruction

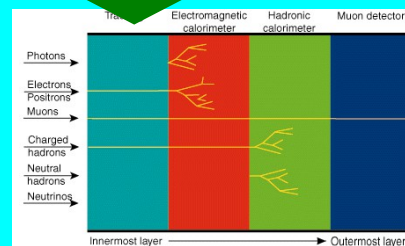


```

101100 010001
110111 001100
111100 100110
110101 110011
001010 001010
100101 000011
010111 010100
    
```



Identification



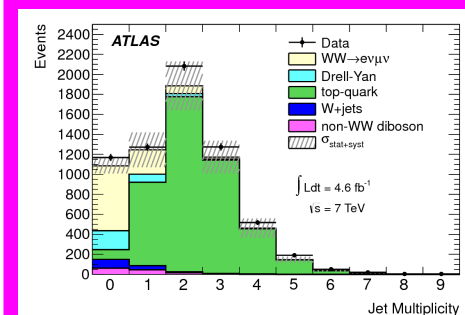
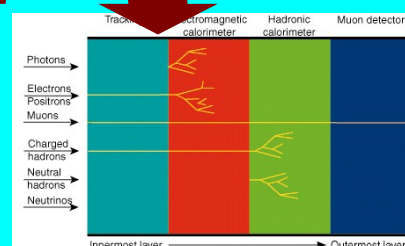
group/individual

Final analysis (n times / day)

Simulation

```

101100 010001
110111 001100
111100 100110
110101 110011
001010 001010
100101 000011
010111 010100
    
```



A new order of magnitude

High number of events

Complex detectors

Complex algorithms

Large distributed community

CMS Experiment at LHC, CERN
Data recorded: Mon Nov 8 11:30:53 2010 CEST
Run/Event: 150431 / 630470
Lumi section: 173

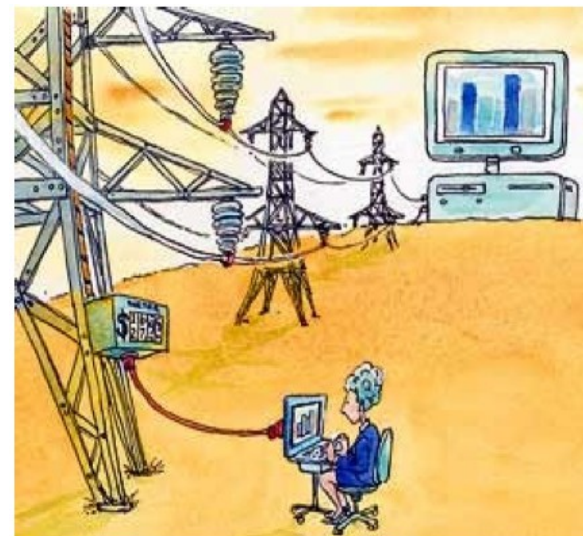
- Argentina
- Armenia
- Australia
- Austria
- Azerbaijan
- Belarus
- Brazil
- Canada
- Chile
- China
- Colombia
- Czech Republic
- Denmark
- France
- Georgia
- Germany
- Greece
- Israel
- Italy
- Japan
- Netherlands
- Norway
- Poland
- Portugal
- Romania
- Russia
- Serbia
- Slovakia
- Slovenia
- South Africa
- Spain
- Sweden
- Switzerland
- Taiwan
- Turkey
- UK
- USA
- CERN
- JINR

ATLAS
Collaboration
~3000 physicists

A new Challenge

- No new concepts, but new orders of magnitude
→ one single sit could not do all the computing alone
(e.g. cooling, power, money)

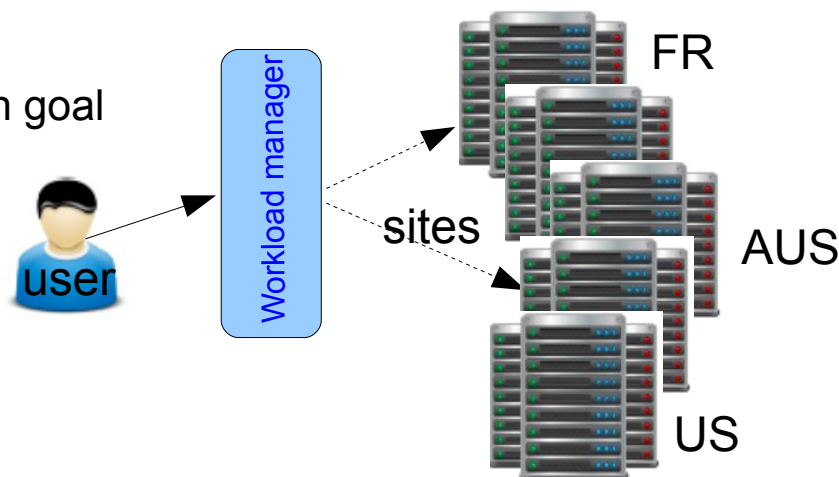
- Though,
 - Computing centers already existing worldwide
 - Often shares between several communities
 - Founding is provided locally
 - Concepts of grid computing fits to the HEP model
 - Authentication / trust / solutions for heterogeneity



Term taken from the
“electric power grid”

→ decision (technical & political) to build a grid infrastructure for the LHC computing

- Share of resources from several units for a common goal



Running jobs: 236092
Transfer rate: 11.41 GiB/sec

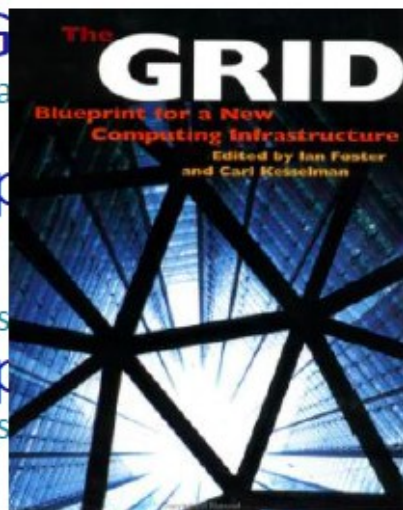
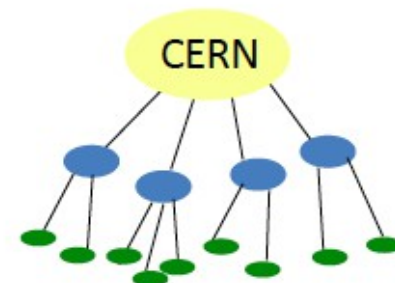


Building WLCG

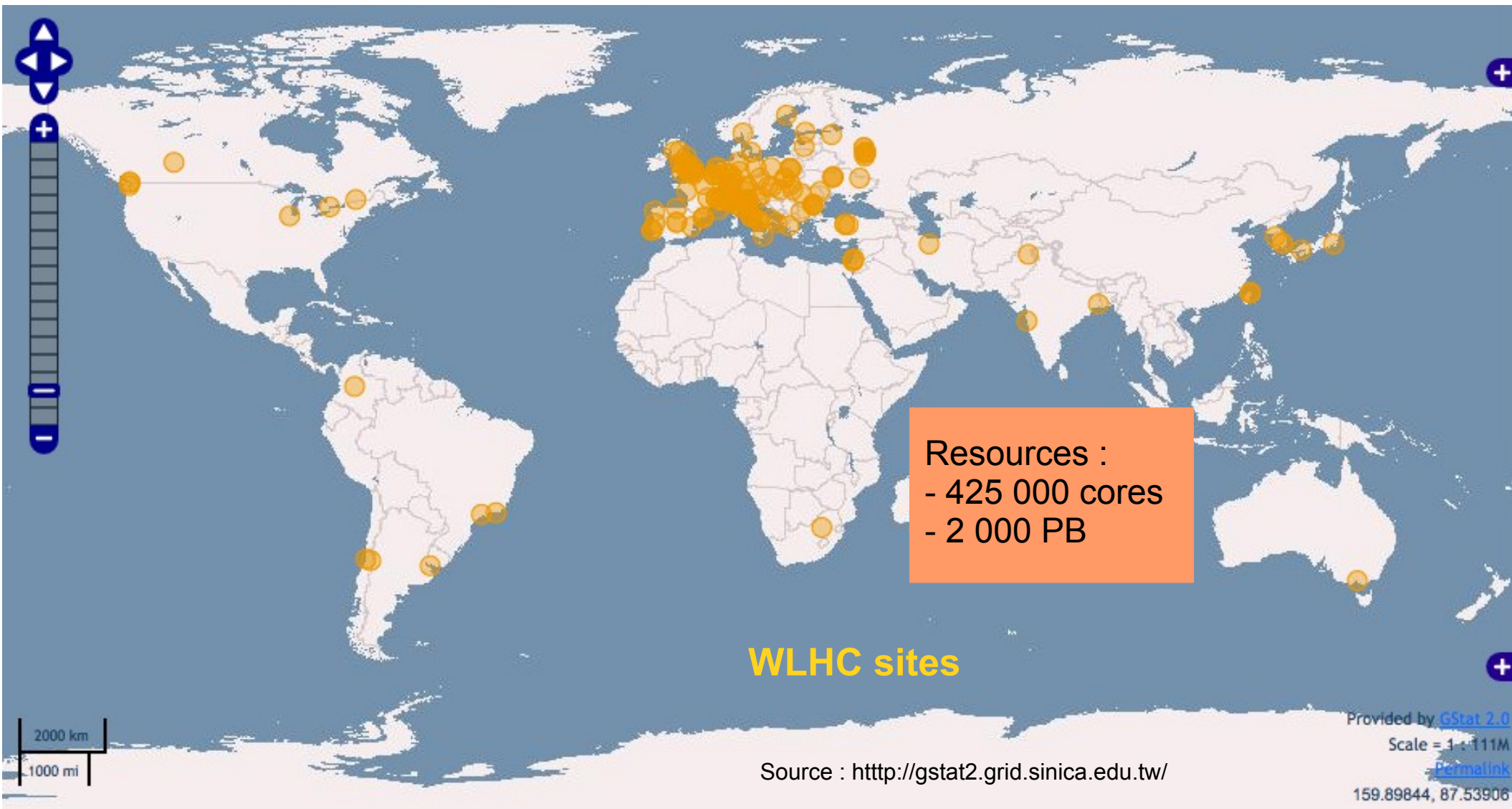
ographer
/BKG
NGA, GEBCO

A few words about history

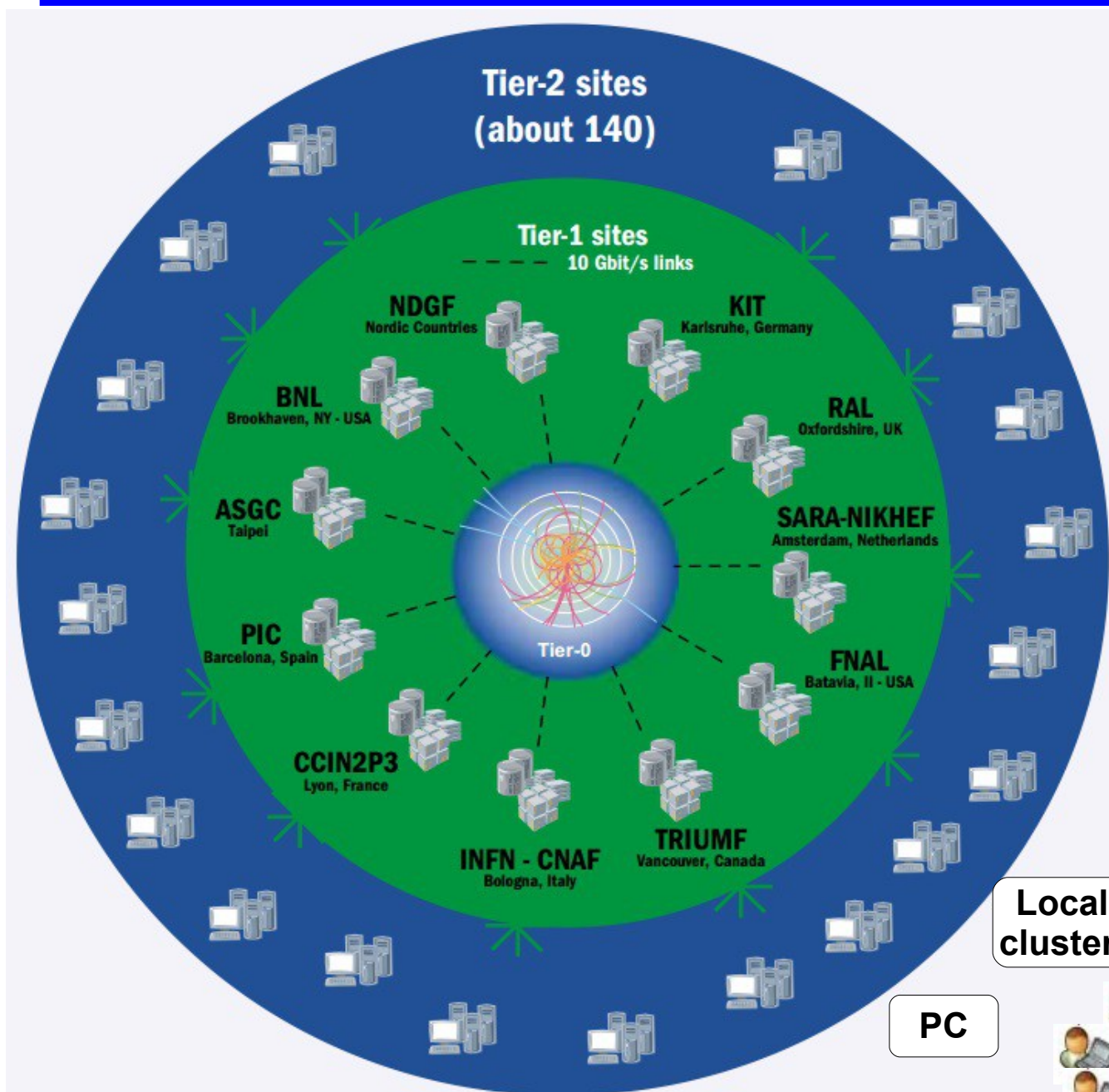
- 1999 - MONARC project
 - First LHC computing architecture – hierarchical distributed model, focus on network control
- 2000 – growing interest in grid technology
 - HEP community main driver in launching the DataGrid project
- 2001-2004 - EU DataGrid project
 - middleware & testbed for an operational grid
- 2002-2005 – LHC Computing Grid
 - deploying the results of DataGrid to provide a production facility for LHC experiments
- 2004-2006 – EU EGEE project phase 1
 - starts from the LCG grid
 - shared production infrastructure
 - expanding to other communities and sciences
- 2006-2008 – EU EGEE project phase 2
 - expanding to other communities and sciences
 - Scale and stability
 - Interoperations/Interoperability
- 2008-2010 – EU EGEE project phase 3
 - More communities
 - Efficient operations
 - Less central coordination
- 2010 – 201x EGI and EMI
 - Sustainable infrastructures based on National Grid Infrastructures
 - Decoupling of middleware development and infrastructure
 - Merging middleware stacks in europe



Sites around the world



Sites hierarchy



Tier0 (CERN):

- Raw data storage
- Initial reco pass
- Data distribution

Tier-1:

- Permanent storage
- Subsequent reco passes

Tier-2:

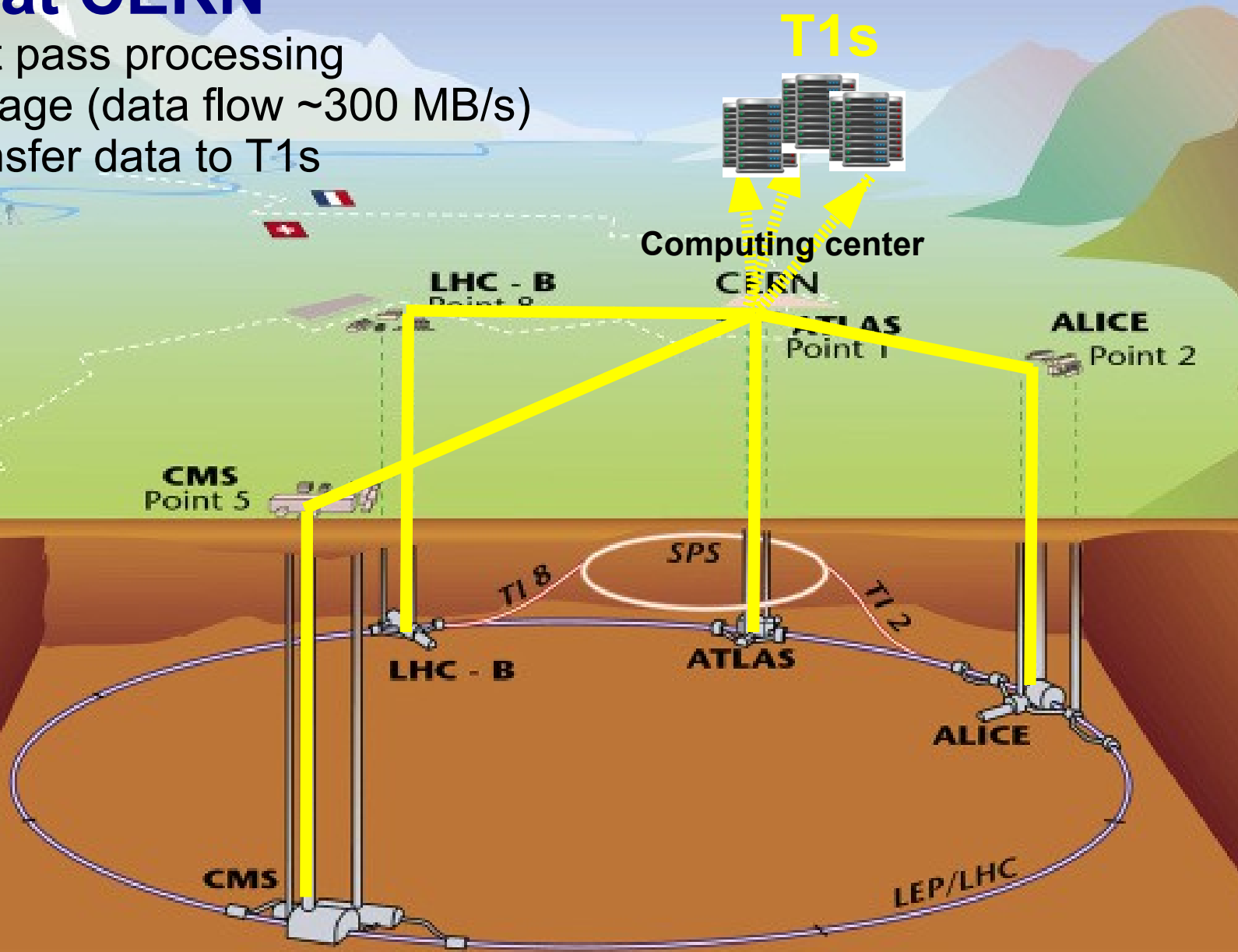
- Simulation
- End user analysis

In addition (end user analysis):

- Tier-3
- Local clusters

T0 at CERN

- First pass processing
- Storage (data flow ~ 300 MB/s)
- Transfer data to T1s

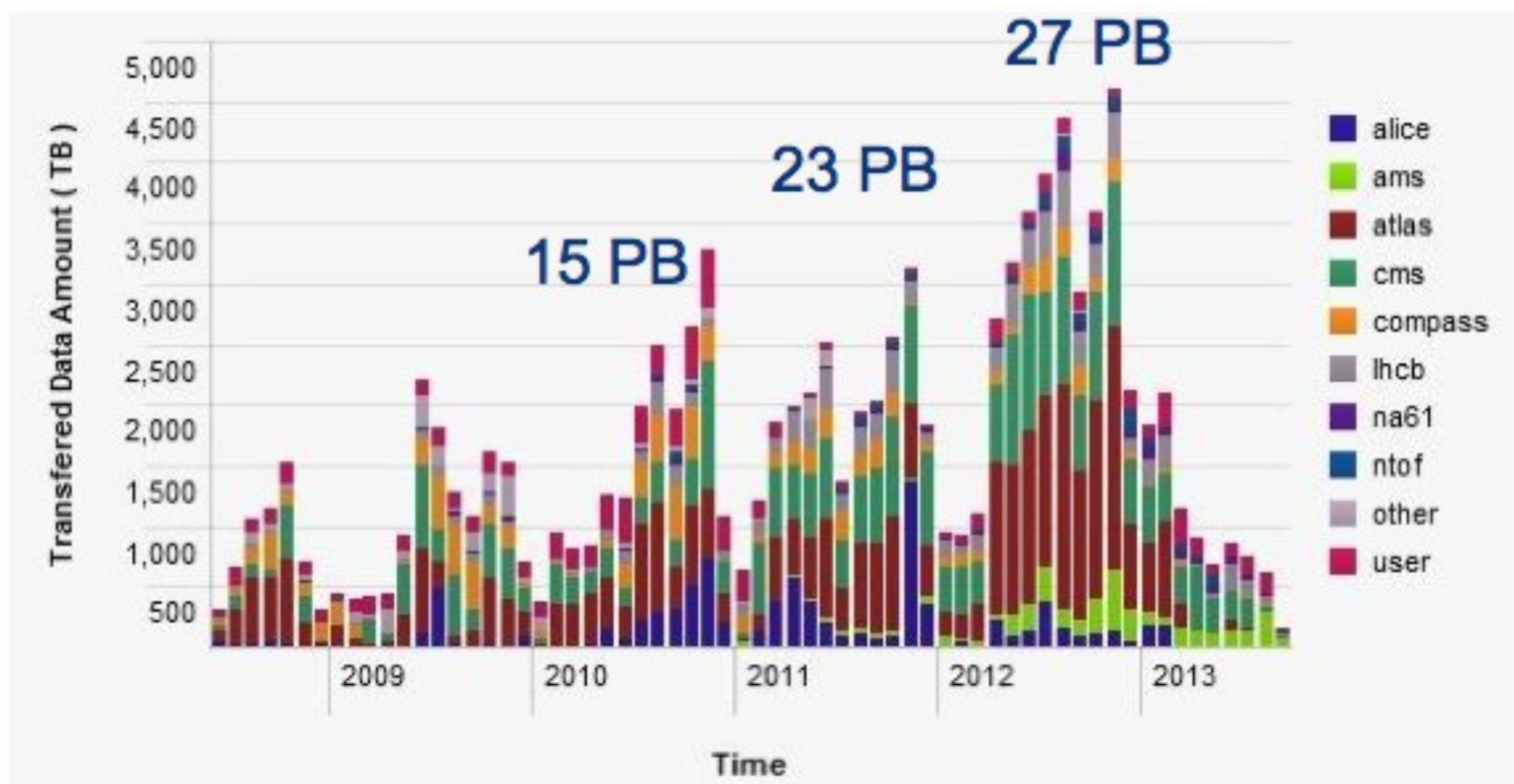


CERN tape writes

CERN tape writes

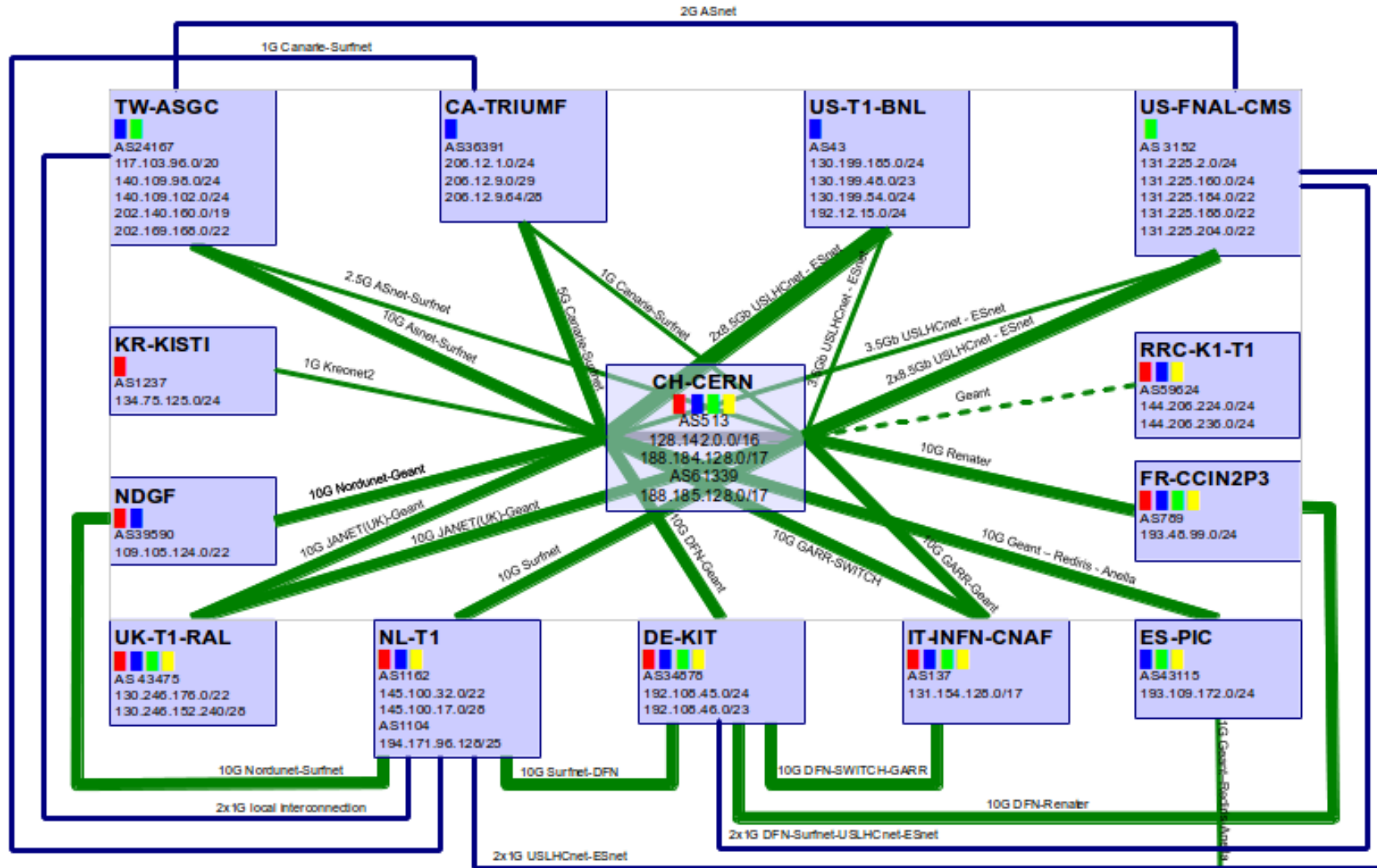
Usage:

- ATLAS 31%
- CMS 25%
- ALICE 9%
- LHCb 8%



LHC Optical Private Network (T0-T1)

Agreements to provide back-up links between T1s



	T0-T1 and T1-T1 traffic		= Alice		= Atlas
	T1-T1 traffic only		= CMS		= LHCb
	Not deployed yet				
	(thick) >=10Gbps				
	(thin) <10Gbps				

p2p prefix: 192.16.166.0/24
edoardo.marfell@cern.ch 20131113

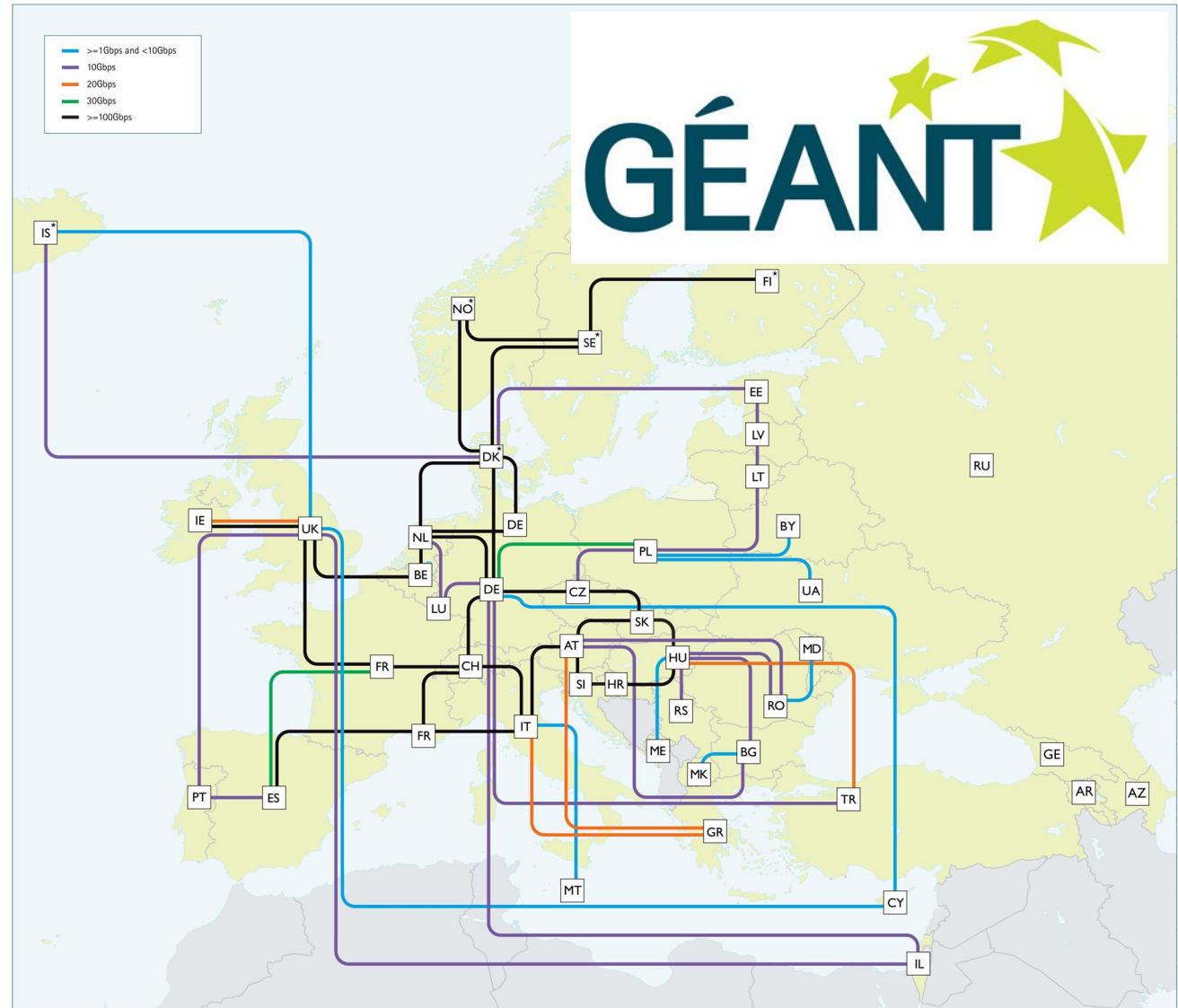
- Dedicated and redundant links
- T0->FR_T1: nominal ~225 MB/s



In Europe

The Pan-European Research and Education Network

GÉANT interconnects Europe's National Research and Education Networks (NRENs). Together we connect over 50 million users at 10,000 institutions across Europe.



- >=1Gbps and <10Gbps
- 10Gbps
- 20Gbps
- 30Gbps
- >=100Gbps














































DE Germany



FR France

GÉANT connectivity as at January 2014. GÉANT is operated by DANTE on behalf of Europe's NRENs.

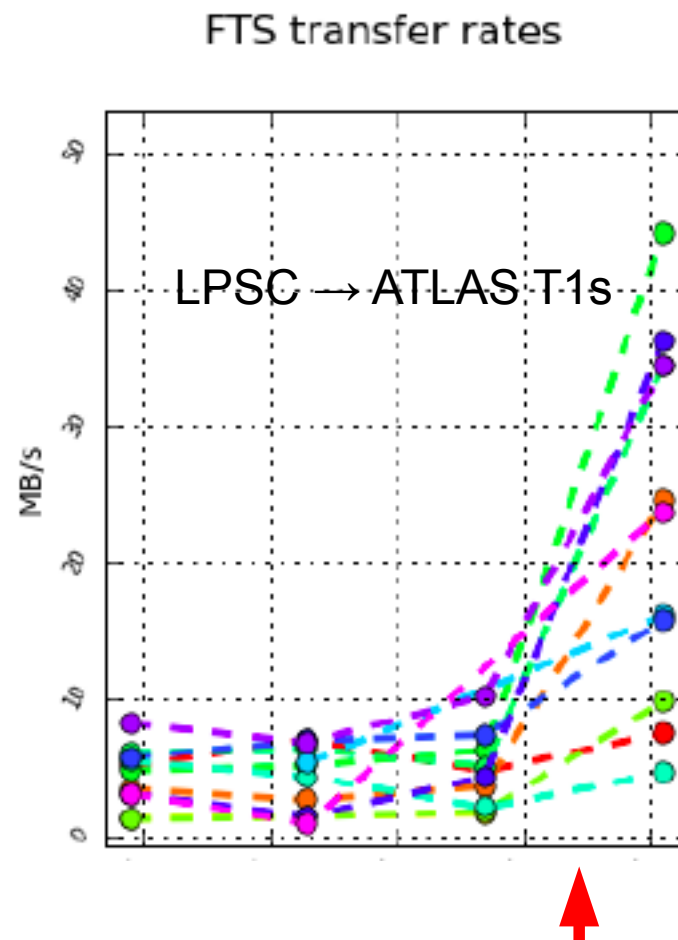
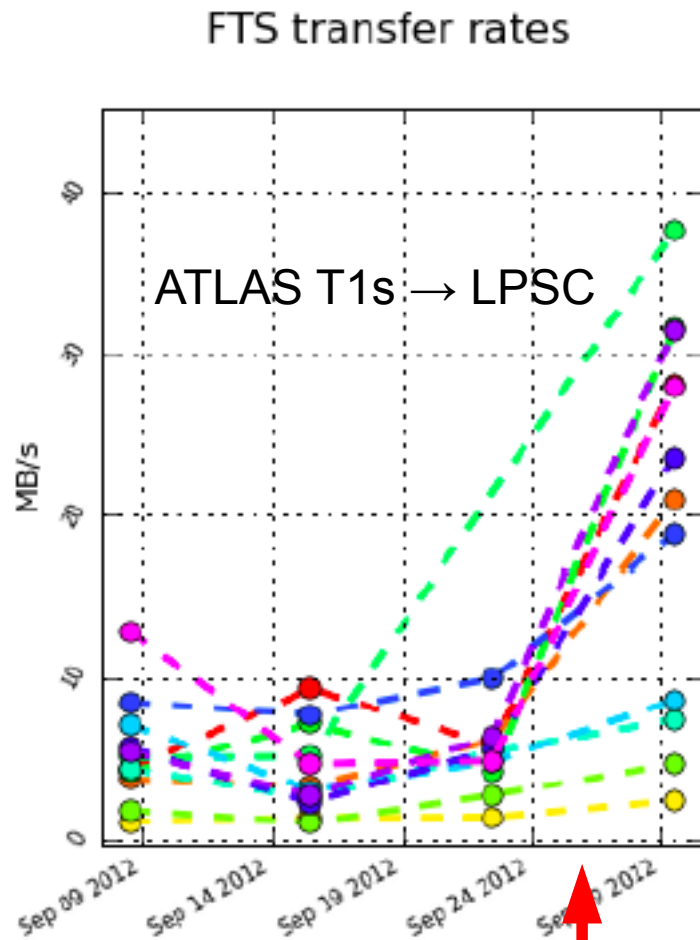
 Armenia	 Bulgaria	 Germany	 Finland	 Croatia	 Iceland	 Latvia	 Netherlands	 Romania	 Sweden	 Turkey
 Austria	 Switzerland	 Denmark	 France	 Hungary	 Italy	 Montenegro	 Norway	 Serbia	 Slovenia	 United Kingdom
 Azerbaijan	 Cyprus	 Estonia	 Georgia	 Ireland	 Lithuania	 F.Y.R. Macedonia	 Poland	 Belarus	 Moldova	 Ukraine
 Belgium	 Czech Republic	 Spain	 Greece	 Israel	 Luxembourg	 Malta	 Portugal	 Russia	 Slovakia	

Remote Tier-0

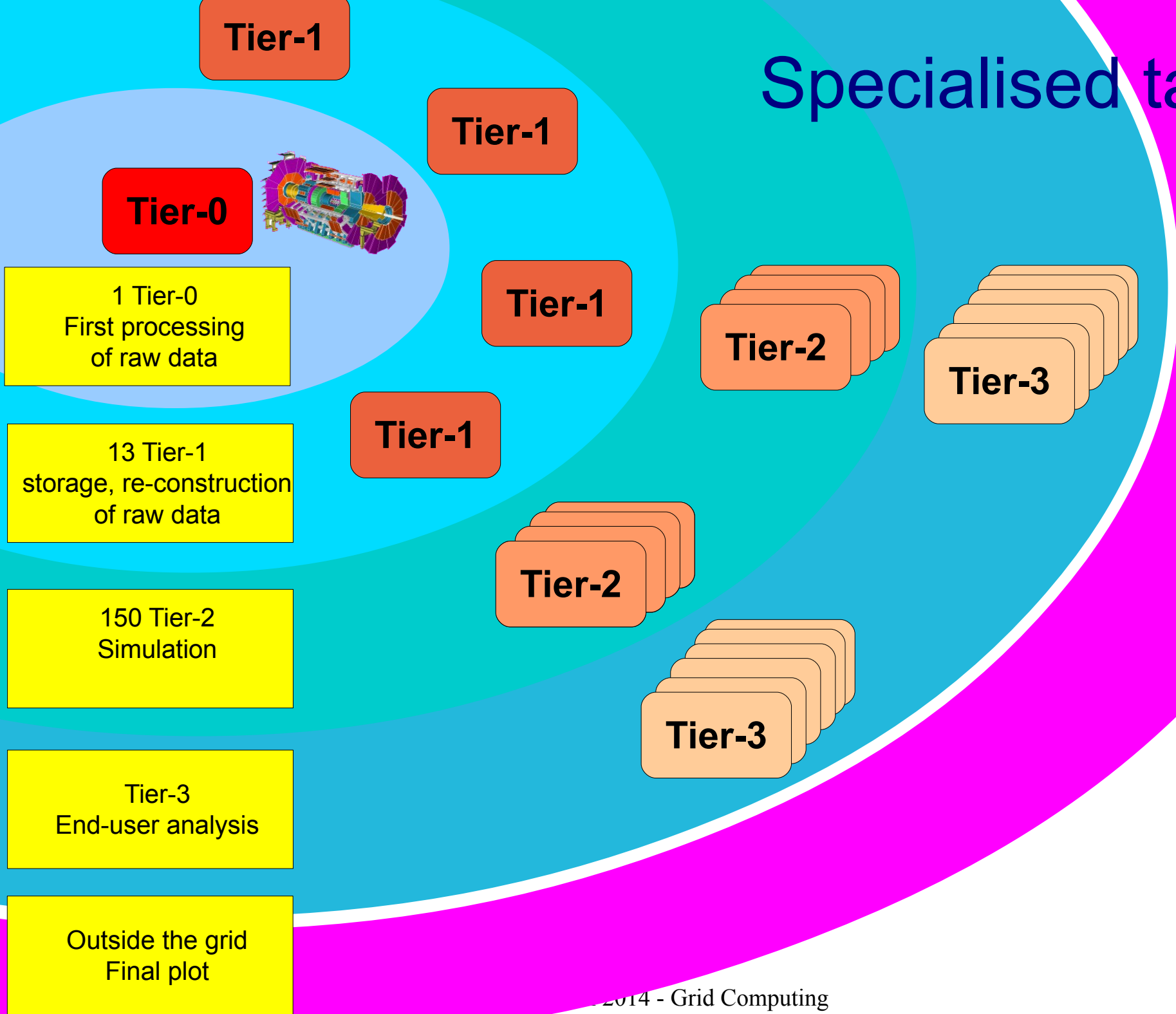


Upgrading the network

LPSC site, a WLCG Tier-2, serves ALICE and ATLAS
Network upgrade from 700Mb/s to 5Gb/s



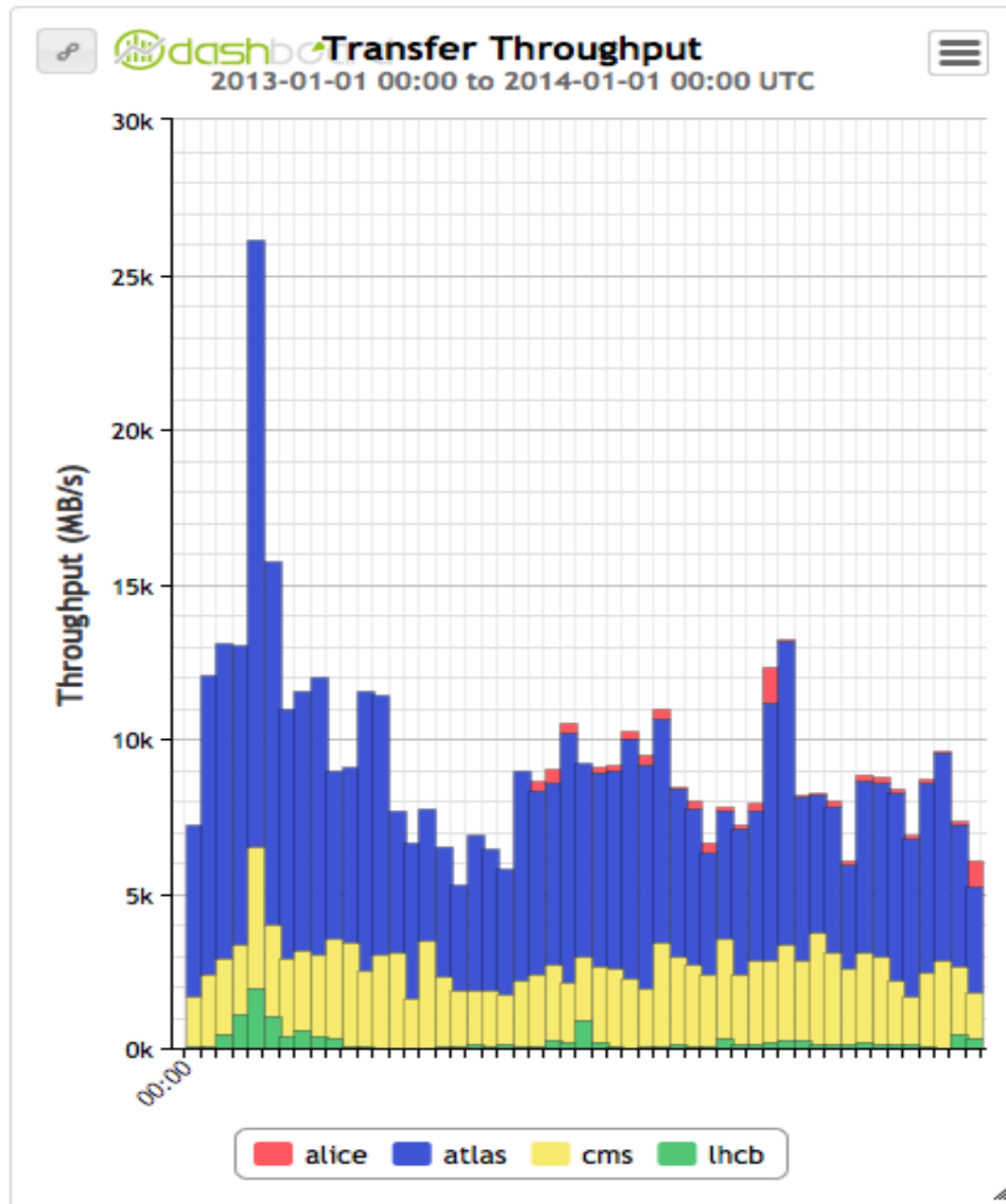
Specialised tasks



Global transfers

Year 2013
All VOs

Bin: per week



MoU requirements

T1 centres

Memorandum of Understanding between CERN and the institutions participating in WLCG

(CERN-C-RRB-2005-1/
Rev. April 2009)

Only an extract !

Service	Maximum delay in responding to operational problems			Average availability ⁵ measured on an annual basis	
	Service interruption	Degradation of the capacity of the service by more than 50%	Degradation of the capacity of the service by more than 20%	During accelerator operation	At all other times
Acceptance of data from the Tier-0 Centre during accelerator operation	12 hours	12 hours	24 hours	99%	n/a
Networking service to the Tier-0 Centre during accelerator operation	12 hours	24 hours	48 hours	98%	n/a
Data-intensive analysis services, including networking to Tier-0, Tier-1 Centres outwith accelerator operation	24 hours	48 hours	48 hours	n/a	98%
All other services ⁶ – prime service hours ⁹	2 hour	2 hour	4 hours	98%	98%
All other services ⁶ – outwith prime service hours ⁹	24 hours	48 hours	48 hours	97%	97%

The response times in the above table refer only to the maximum delay before action is taken to repair the problem. The mean time to repair is also a very important factor that is only covered in this table indirectly through the availability targets. All of these parameters will require an adequate level of staffing of the services, including on-call coverage outside of prime shift.

MoU requirements

T2 centres

- Less constraints
 - Receive no data from detectors
- Availability > 95%

Availability: time site is available
/ total time

Reliability: time site is available
/ (total time – **scheduled downtime**)

Test suit per experiment
with monthly report



Tier-2 Availability and Reliability Report

ATLAS

February 2014

Federation Details

Colour coding :

N/A <30% <60% <90% >=90%

Availability Algorithm: (OSG-CE + CE + CREAM-CE) * (SRMv2 + OSG-SRMv2)

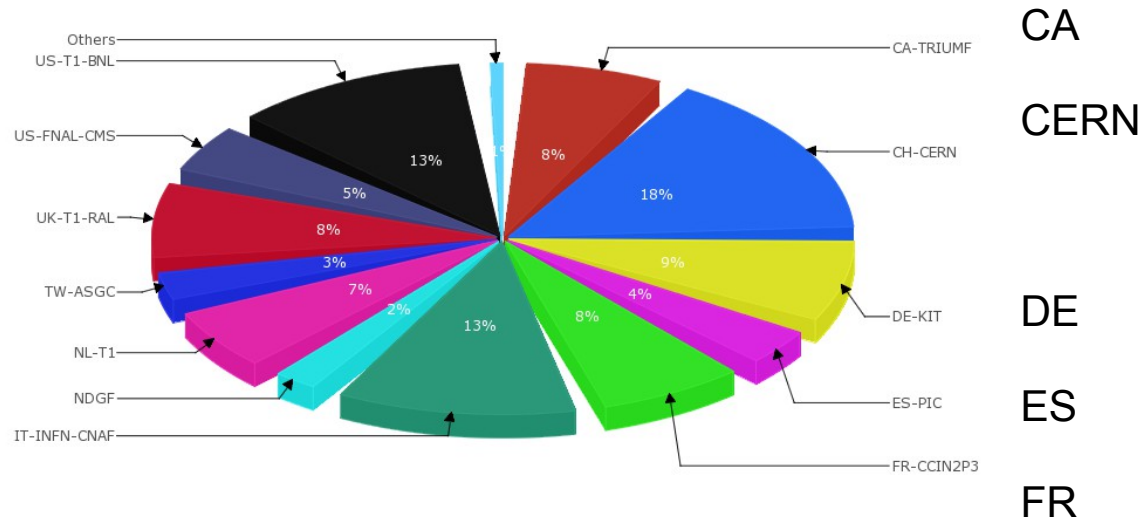
Federation	Site	Phy-CPU	Log-CPU	HS06	Availability	Reliability	Unknown	Availability History		
								Nov 2013	Dec 2013	Jan 2014
FR-IN2P3-LPSC	IN2P3-LPSC	136	660	5399	99 %	99 %	3 %	93 %	100 %	94 %

Tier-1

Developed by CESGA 'EGI View': / normcpu-HEPSPEC06 / 2013:4-2014:3 / VO-TIER1 / lhc (x) / ACCBAR-LIN / i

2014-03-05 01:53

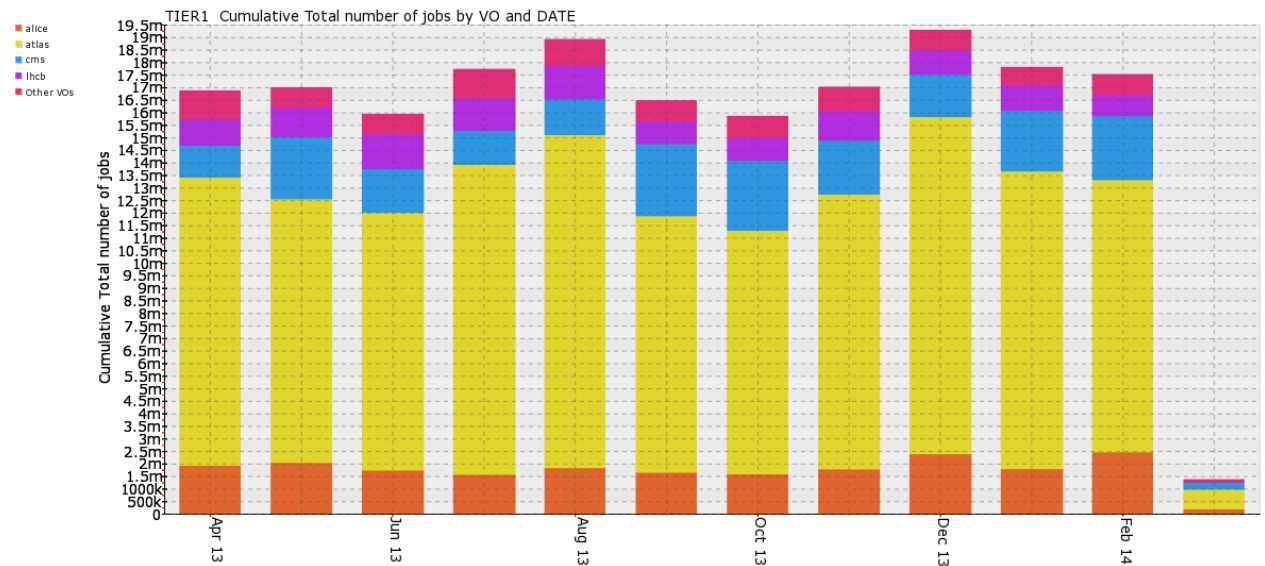
TIER1 Normalised CPU time (HEPSPEC06) per TIER1



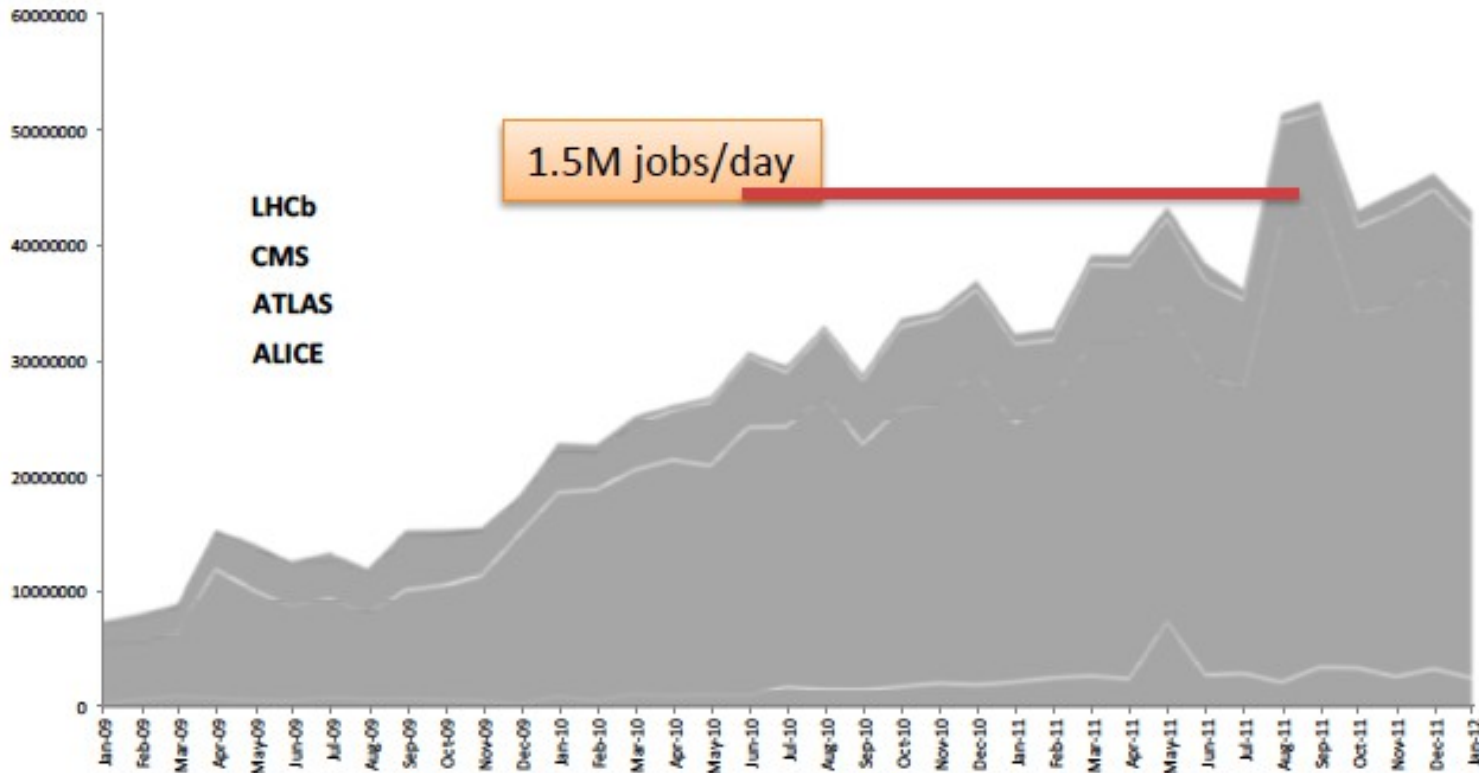
CA
CERN
DE
ES
FR

Developed by CESGA 'EGI View': / njobs / 2013:4-2014:3 / VO-DATE / lhc (i) / ACCBAR-LIN / i

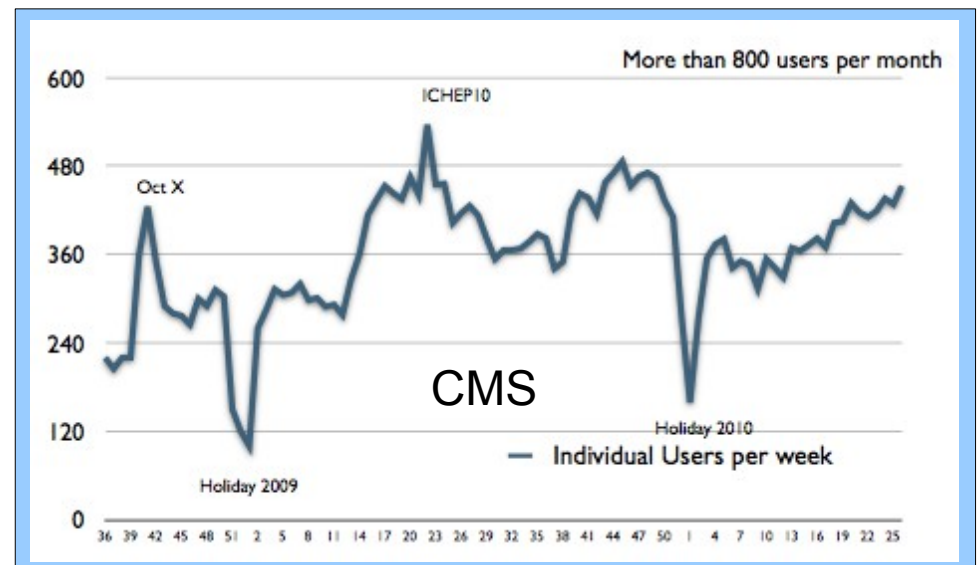
2014-03-05 01:53



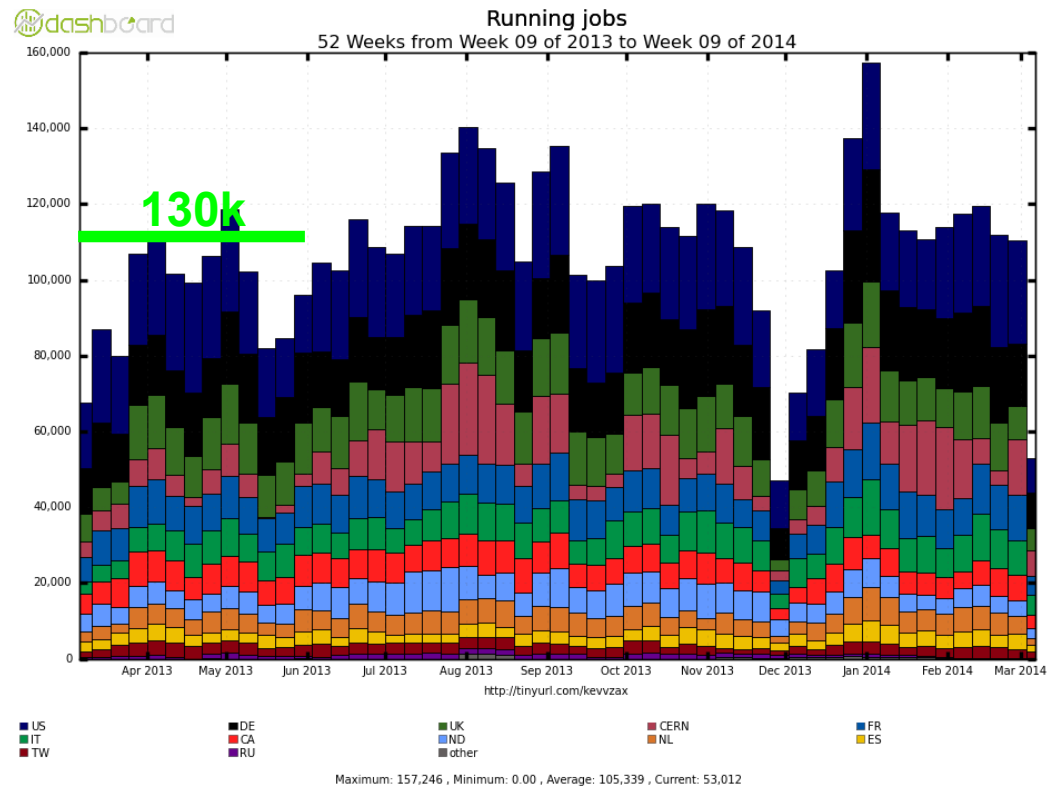
Jobs



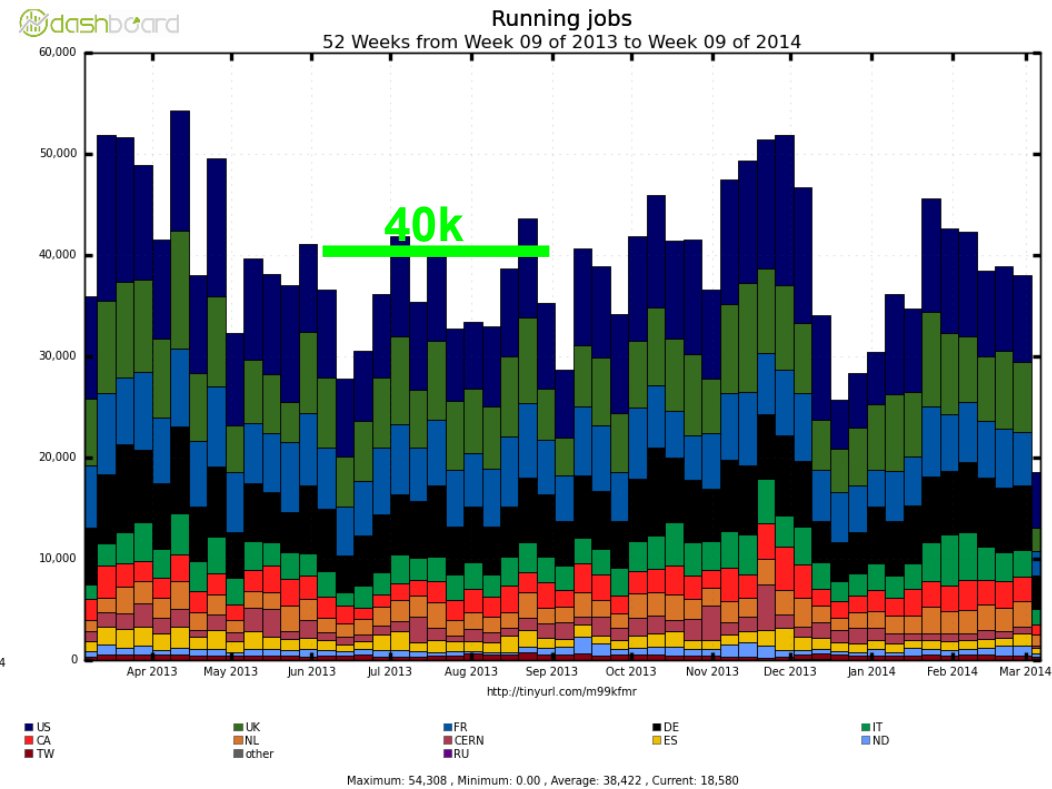
Currently:
 jobs completed/day
 - ALICE ~ 250K
 - ATLAS ~ 850k



ATLAS alone



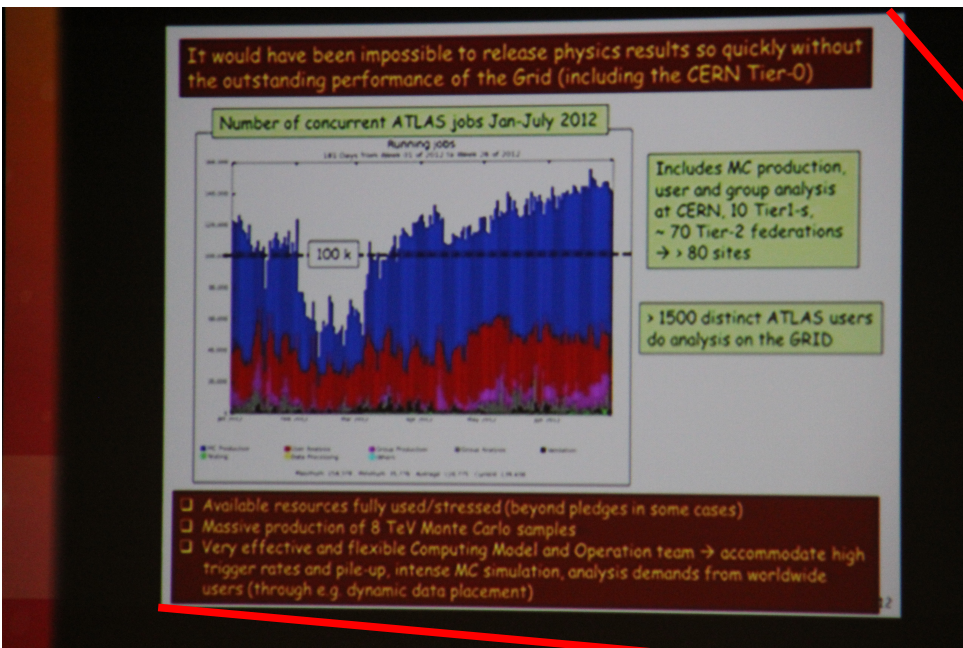
Production jobs



Analysis jobs

“Computing enables physics”

Photography: C. Biscarat



CERN seminar,
July 4th 2012,
retransmitted at
ICHEP (Melbourne)

Running jobs: 236092
Transfer rate: 11.41 GiB/sec



Other grids

ographer
/BKG
NGA, GEBCO

European Grid Initiative

- The WLCG shares the infrastructures with EGI
 - Organised in National Grid Initiative
- EGI project (and ancestors)
 - Build a large scale production grid
 - Secured and robust
 - Promote international collaboration
 - For sciences and technologies
- ~ 300 different user communities
- Large spectra of scientific domains

EGI web site: <http://www.egi.eu/>

Many sciences

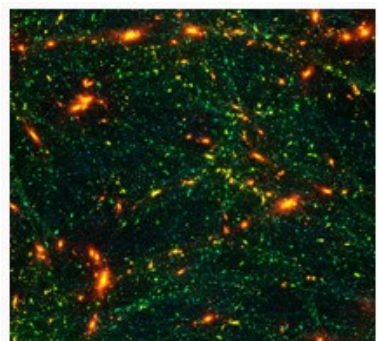


MRI scan with location of the hippocampus

Image: Amber Rieder, Jenna Traynor (wikicommons)

Medical and Health Sciences

Processing of thousands of MRI scans from patients with Alzheimer's disease



Seventy million elements of a cosmological N-body simulation using VisIVO and Splotch.

Physical Sciences

Creation of 3D visualisations of astronomical data on unprecedented scales.

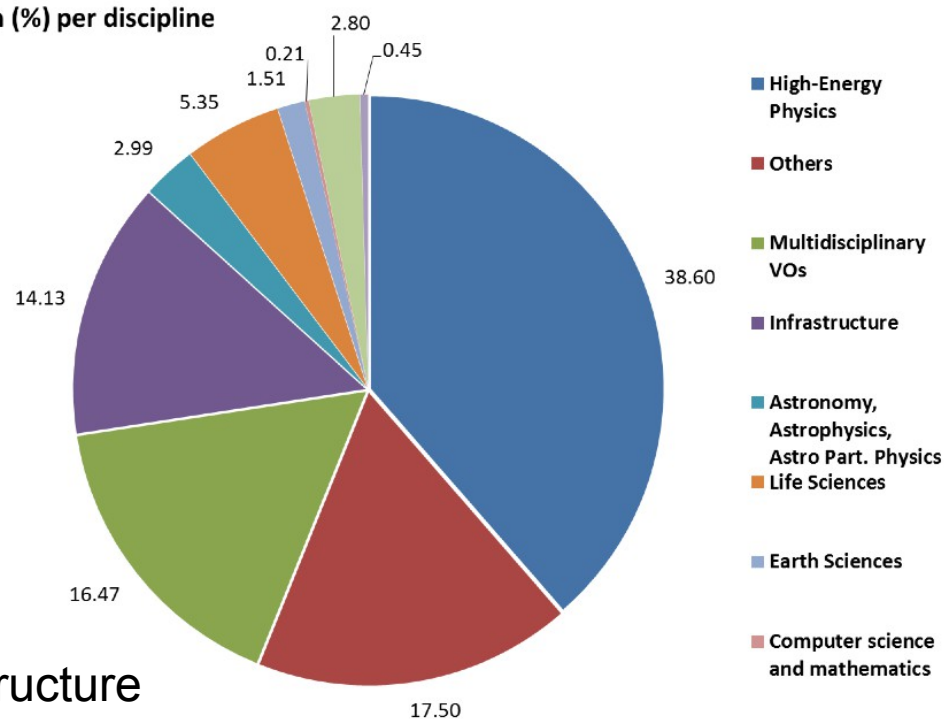


Oil exploration in the North Sea depends on a good knowledge of the underground rock structure. Grid computing helps scientists to make sense of all the seismic data.

Natural sciences

Correlate data from millions of calculations to unveil the rock structure of an oil field under the North Sea.

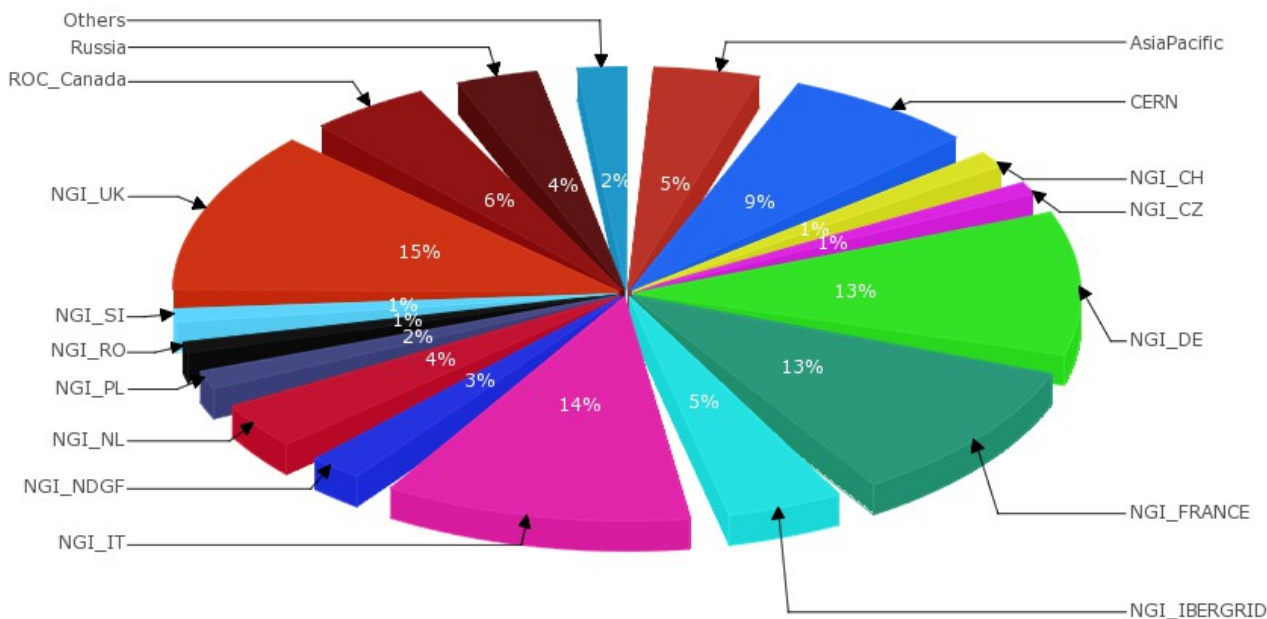
User distribution (%) per discipline



From ISGC 2013

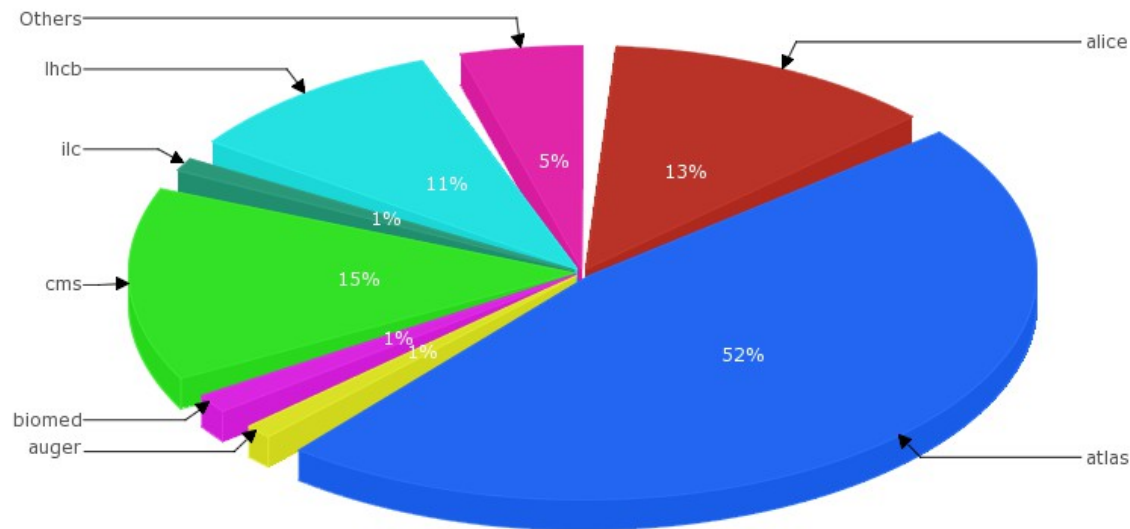
Normalised CPU time (kSI2K) per REGION

Resource consumption



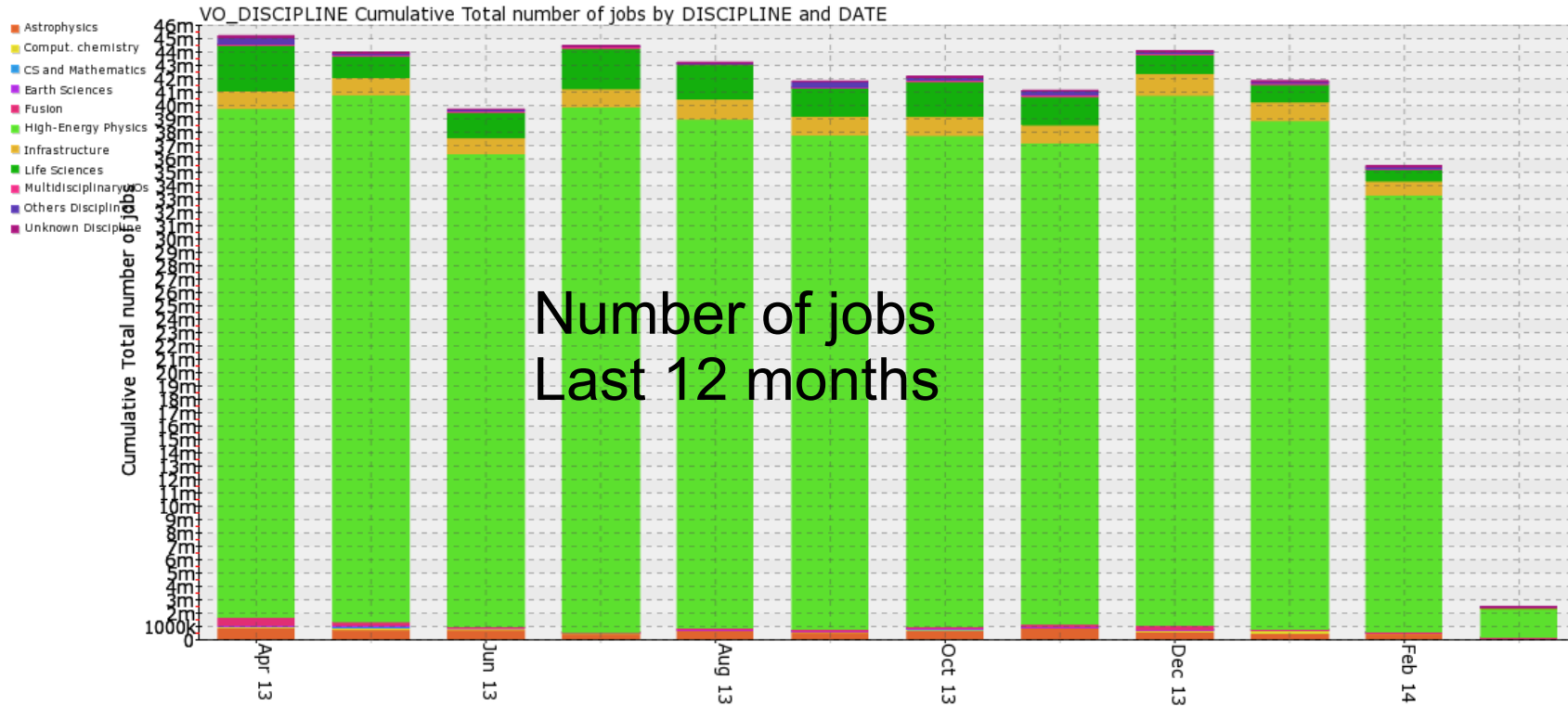
Normalised CPU time
 Last 12 months
 - Per region
 - Per VO

;) per VO



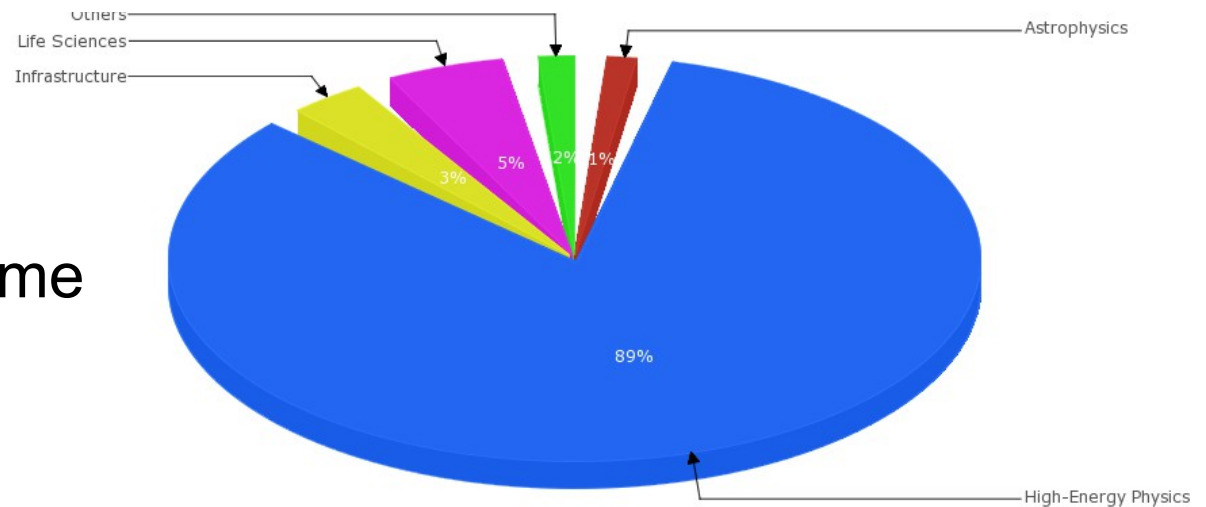
<https://accounting.egi.eu/>

Per discipline



<https://accounting.egi.eu/>

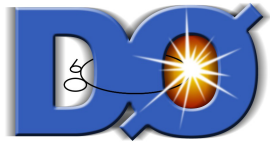
Normalised CPU time Last 12 months



A few user example

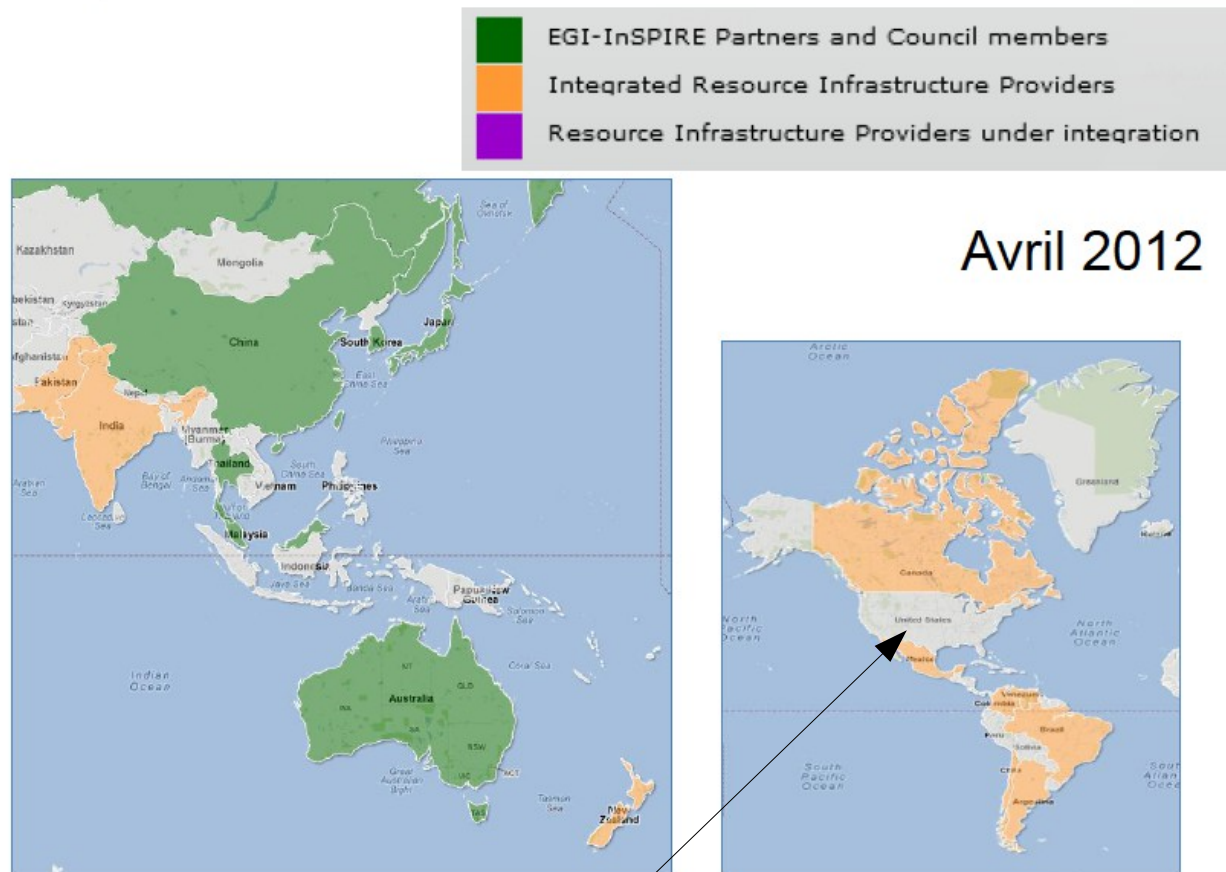


An observatory for ground-based
gamma-ray astronomy



EGI partners and resource providers

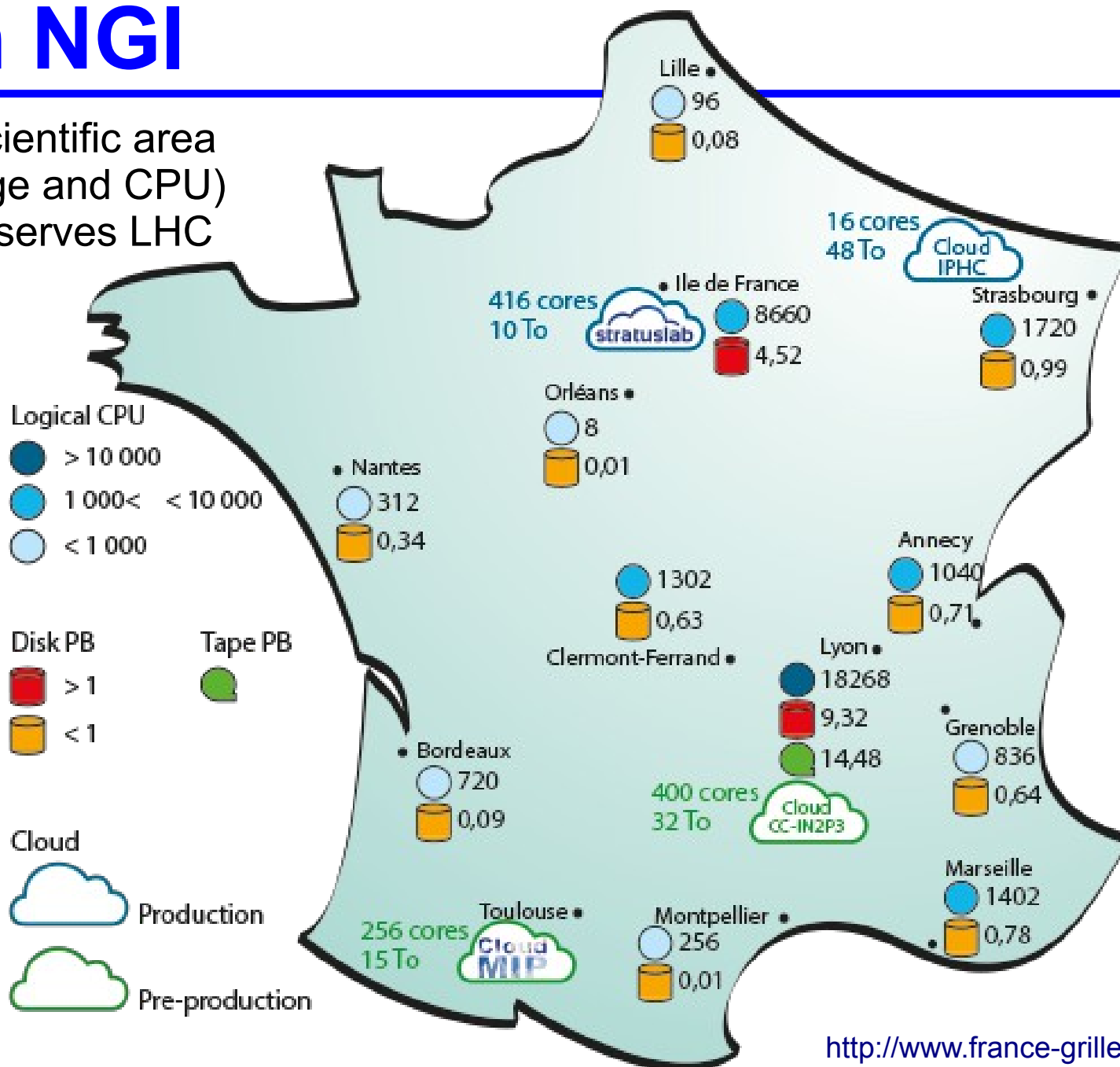
- The most extended grid infrastructure, 35 countries.



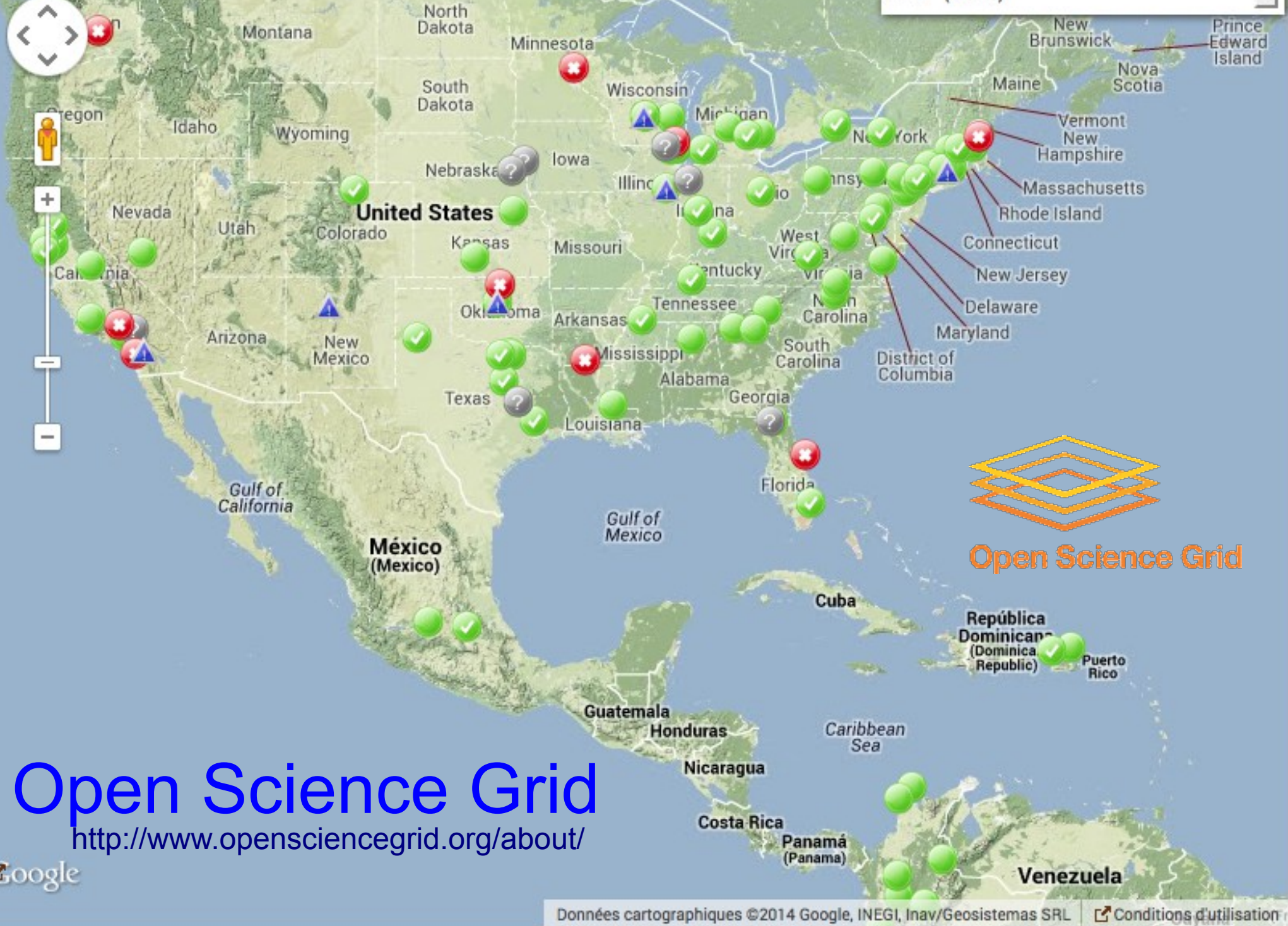
Open Science Grid (OSG)
Own operations tools and procedures,
compatible policies

French NGI

- Open to any scientific area
- 18 sites (storage and CPU)
- Half of them serves LHC



<http://www.france-grilles.fr/>



Many kinds of grids

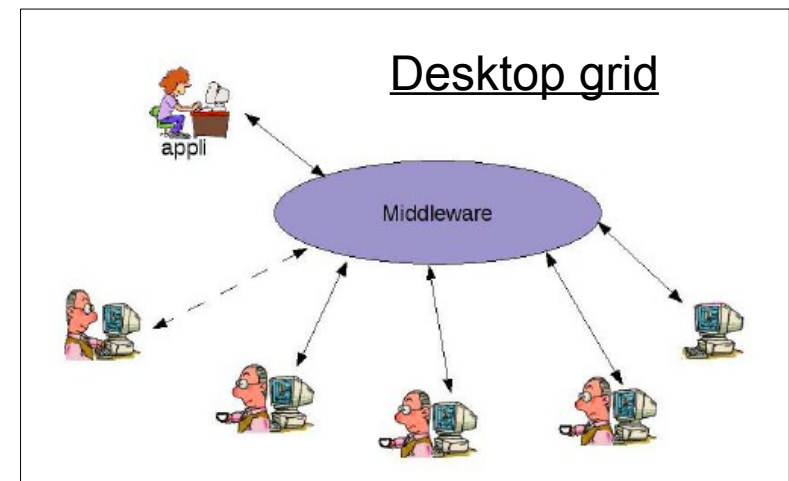
They might have different goals:

- **Production** grid
 - Shared resources for a common goal, known users
 - Shared “interfacing” software (middleware)
- **Development** grid (Grid5000 in France)
 - Shared resources for a common goal, known users
 - Testing grid software
- **Volunteer** grid (BOINC)
 - Resources given to a project with unknown users

They may differ in heterogeneity:

- Desktop grid (BOINC)
- Service grid (EGI, OSG,...)
- HPC grid (DEISA, TeraGrid,...)

They may be interoperable (EGI+OSG)



Many kinds of grids

Geographically:

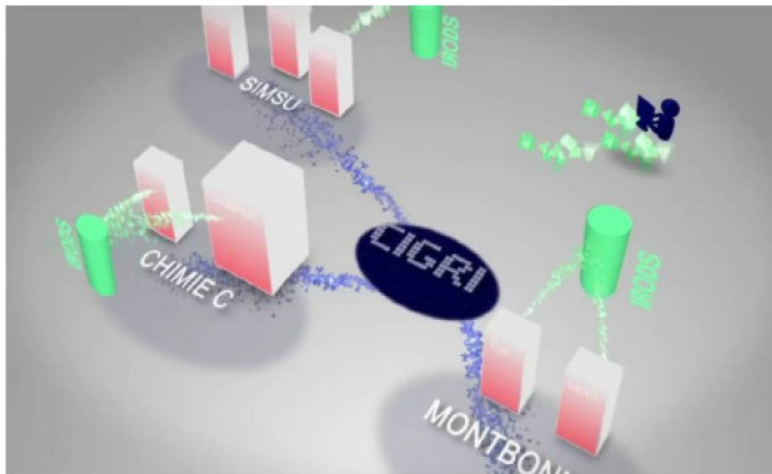
- International grid
- Regional grid
-

In term of science field:

- Thematic grid
- Multidisciplinary grid



GRISBI: bioinformatic (France)



CIGRI: grid on top of T2 HPC clusters (Grenoble Universities, INPG, and National research labs) – opportunistic usage

Running jobs: 236092
Transfer rate: 11.41 GiB/sec

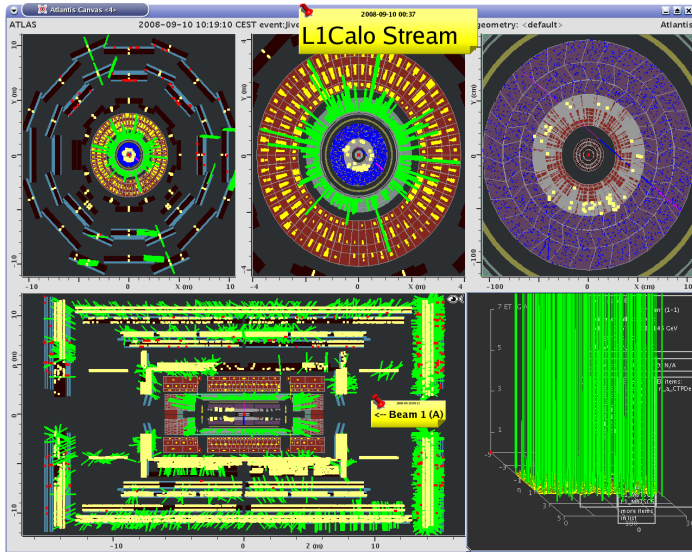


Behind the scenes

ographer
/BKG
NGA, GEBCO

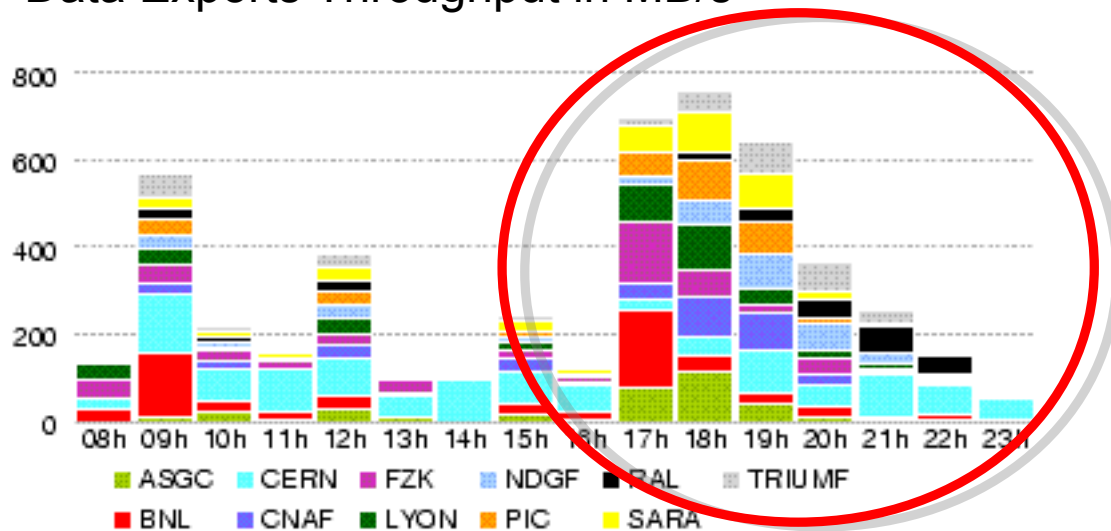
D-day « LHC first beam »

- Sept. 10th 2008



D-day « LHC first beam »

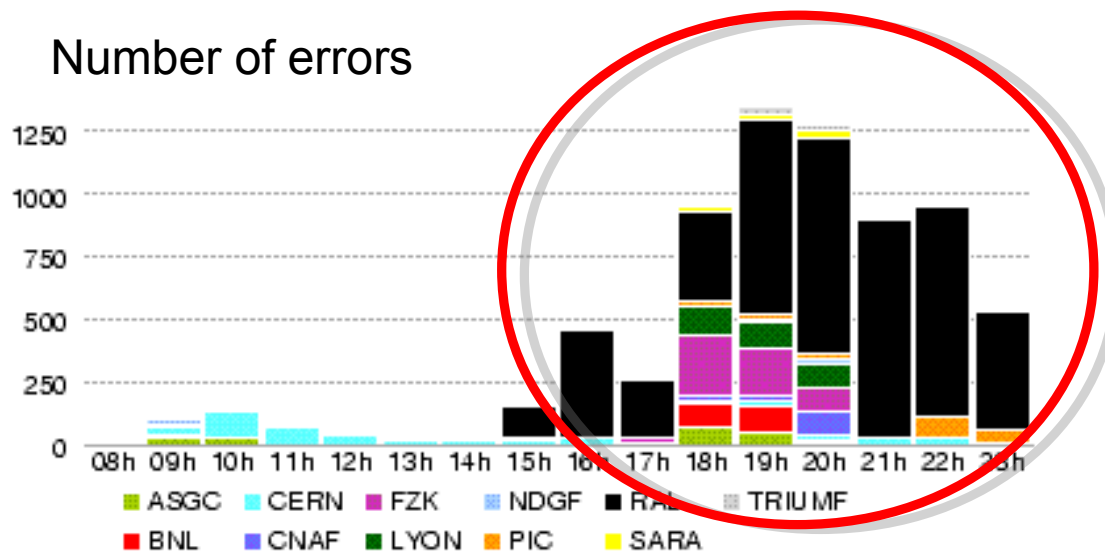
Data Exports Throughput in MB/s



In the Computing control room (2nd floor)

Concurrent access to data on central servers by the end-users

Number of errors

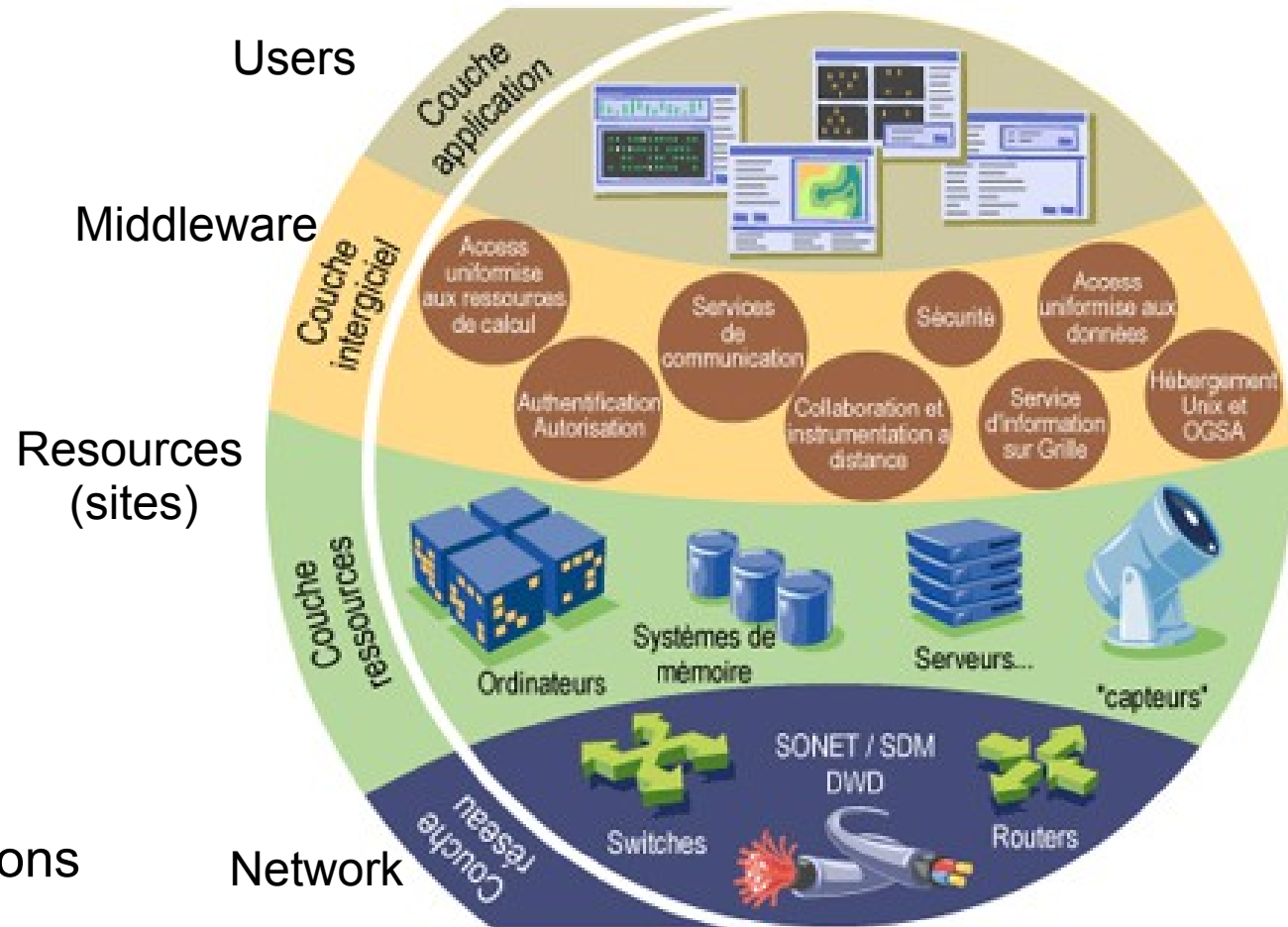


Middleware

- Heterogeneity of the sites
 - Batch systems
 - Storage
 - Operating systems
 - Worker nodes capacity
 - Authentication

Middleware

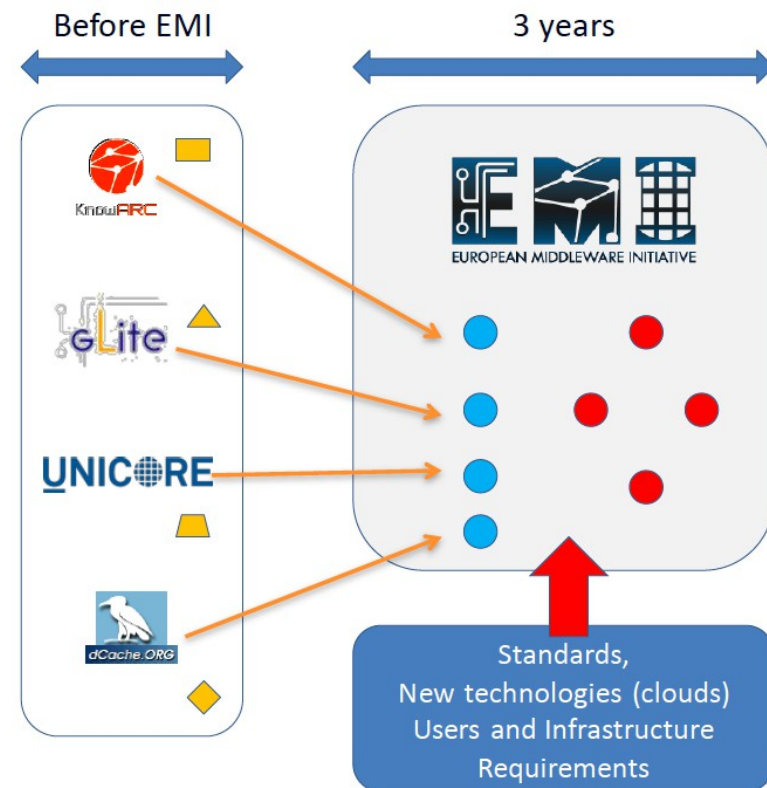
- Software layer
- Hiding and simplifying
 - No local user registration
 - Abstraction of resources
 - Interface to user applications



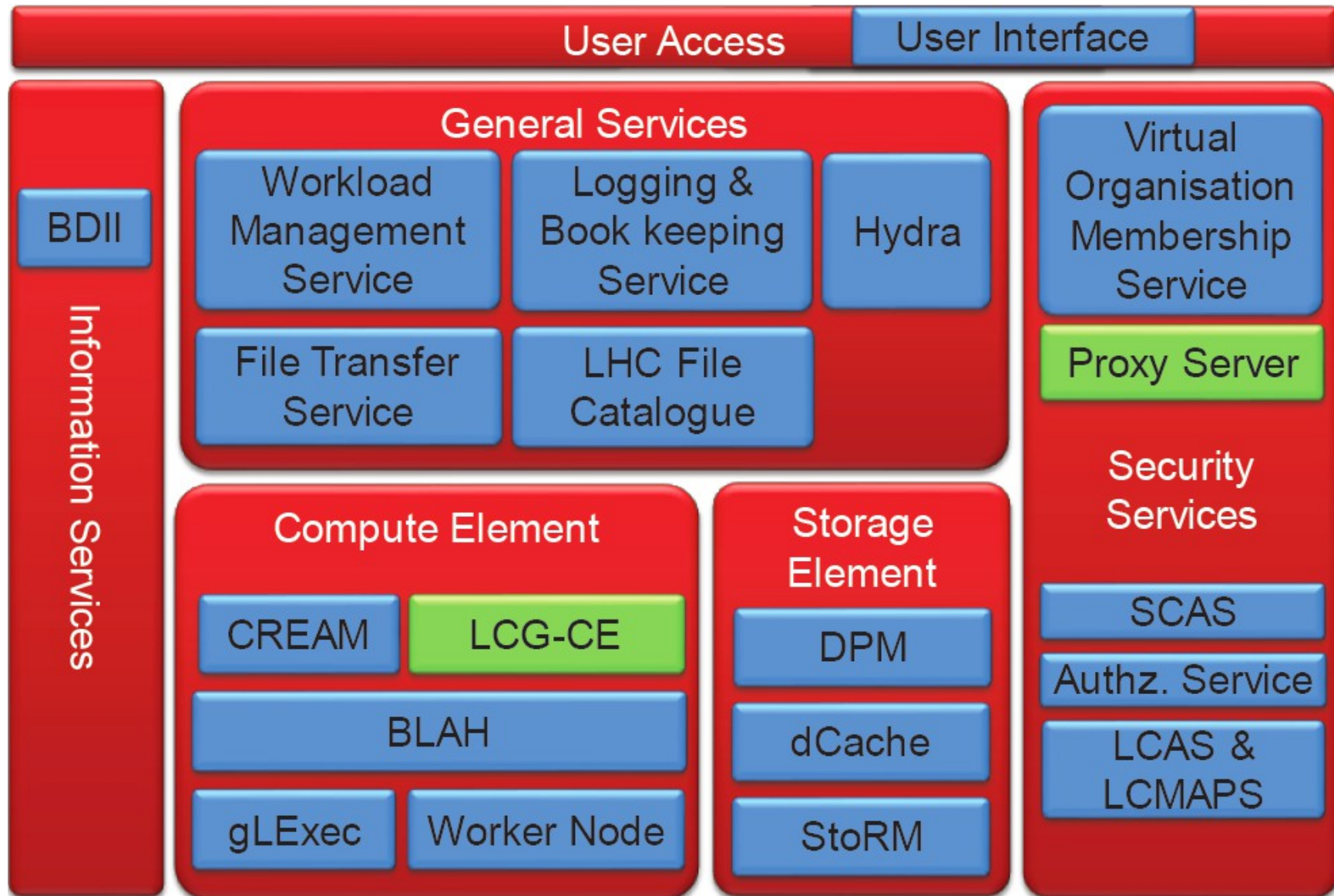
Ref : gridCafe

Towards standardisation

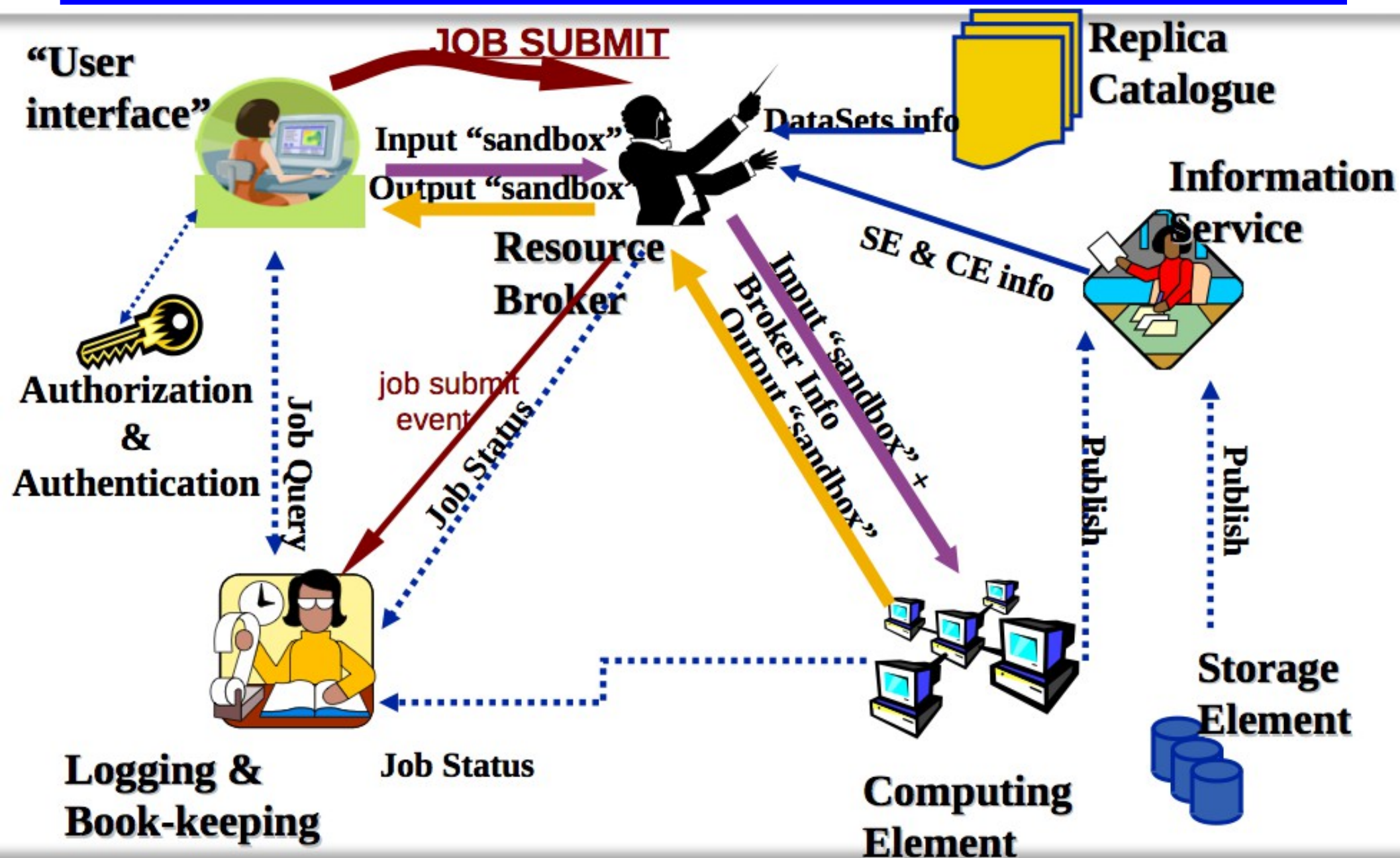
- Origin
 - Idea became popular in 1998 with the book “The Grid” (I. Foster & K. Kesselman)
 - Globus: the first middleware (1998) – evolution to globustoolkit-5
- Now : hundreds of middleware, tentative for standardisation
- In EGI: develop and use of EMI (European Middleware Initiative)



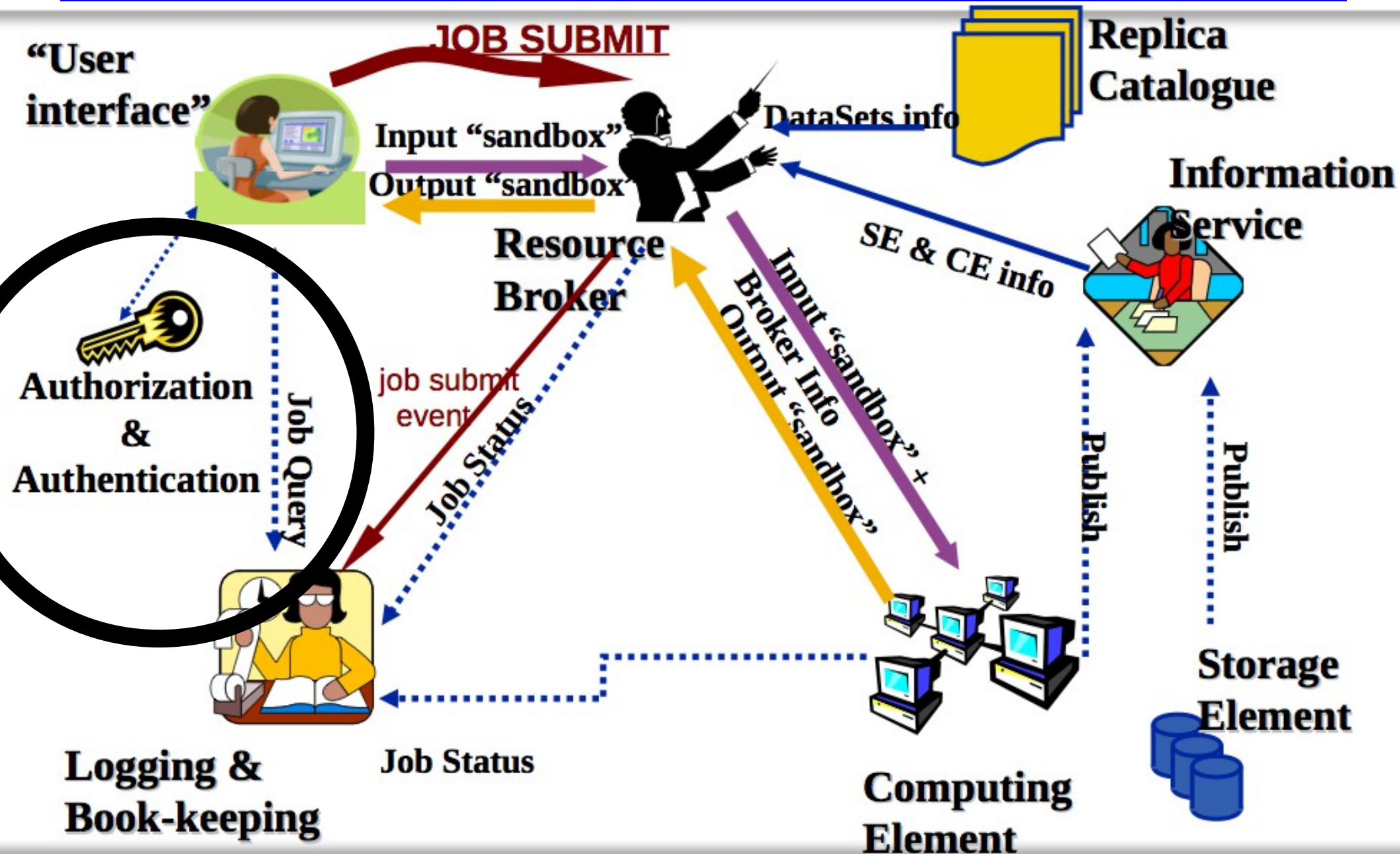
Major EMI services



Putting it all together

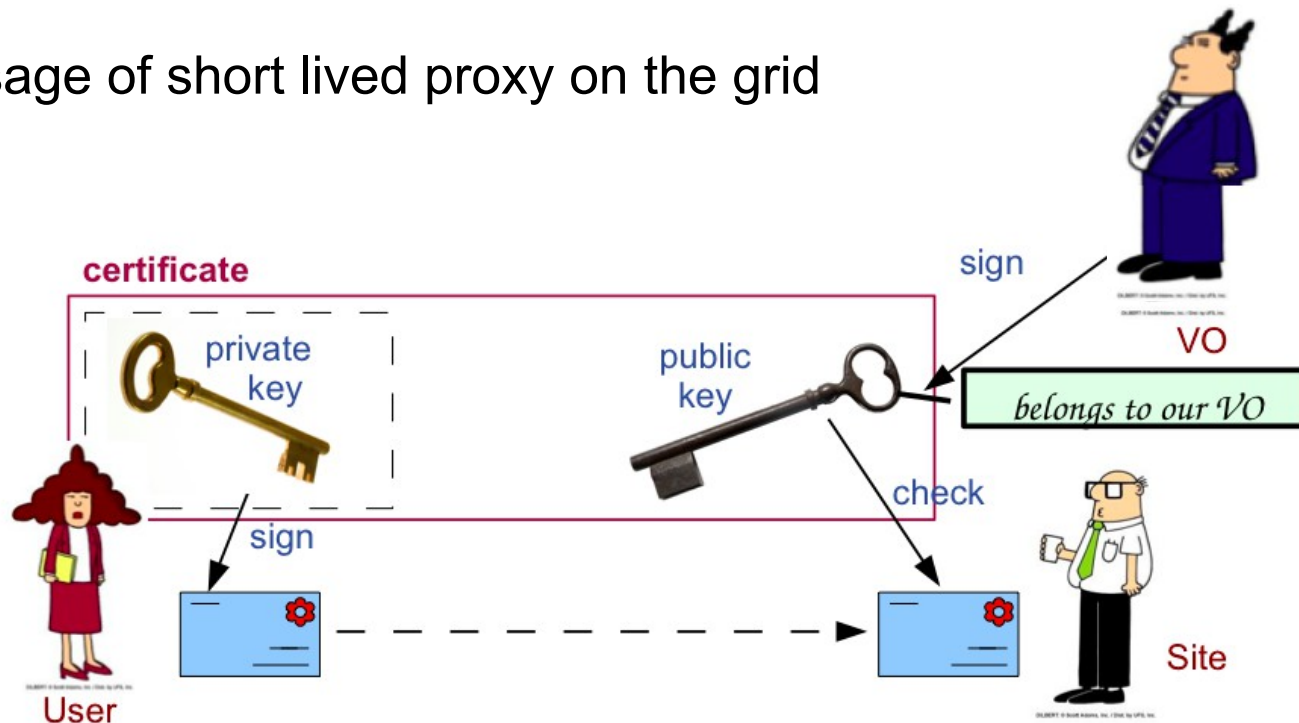


A word on



Authentication

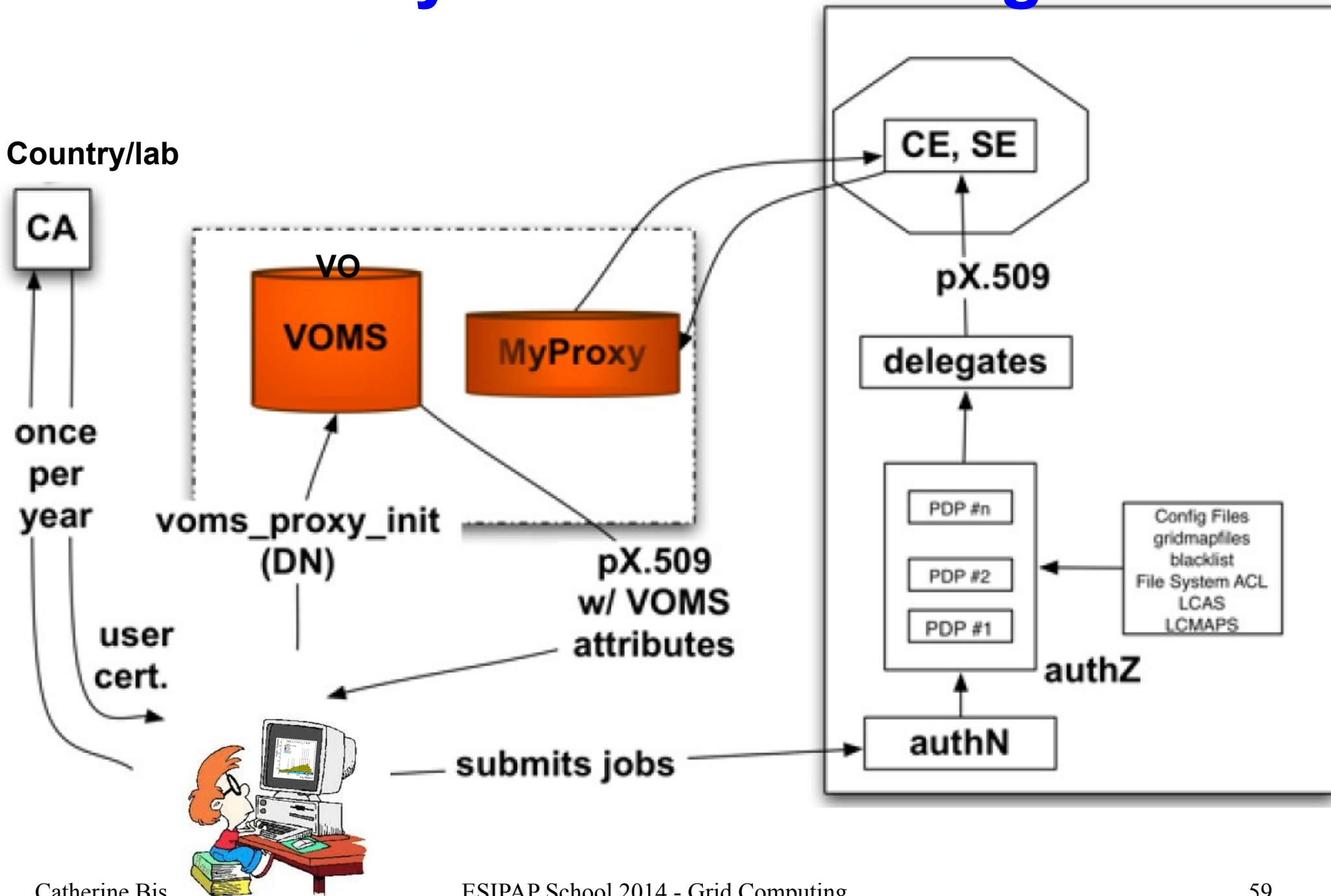
- The grid is large and (almost) nobody knows you
- **Authentication : who you are**
 - **Certificate authority** delivers certificate (long lived)
 - What you may already use in your web browser
 - Based on a public/private pair of keys
 - X509
 - For security, usage of short lived proxy on the grid



Authorisation

- **Authorisation: what can you do**
 - Users are organised in Virtual Organisation (VOs)
 - They generally belong to different organisations
 - They have a common goal / accept to share resources
 - Examples: one VO per LHC experiment, CTA, biomed, france-grilles, ...
 - VOMS (virtual Organisation MemberShip)
 - Delivered by your VO
 - Allow groups and roles (capabilities)
- additional attributes to your personal certificate
- Authorisation system must be agreed with all partners
 - Key ingredient : **trust**

Security: the root of the grid



EGI virtual organisations

Data to graph: Active VO Information about Active VOs

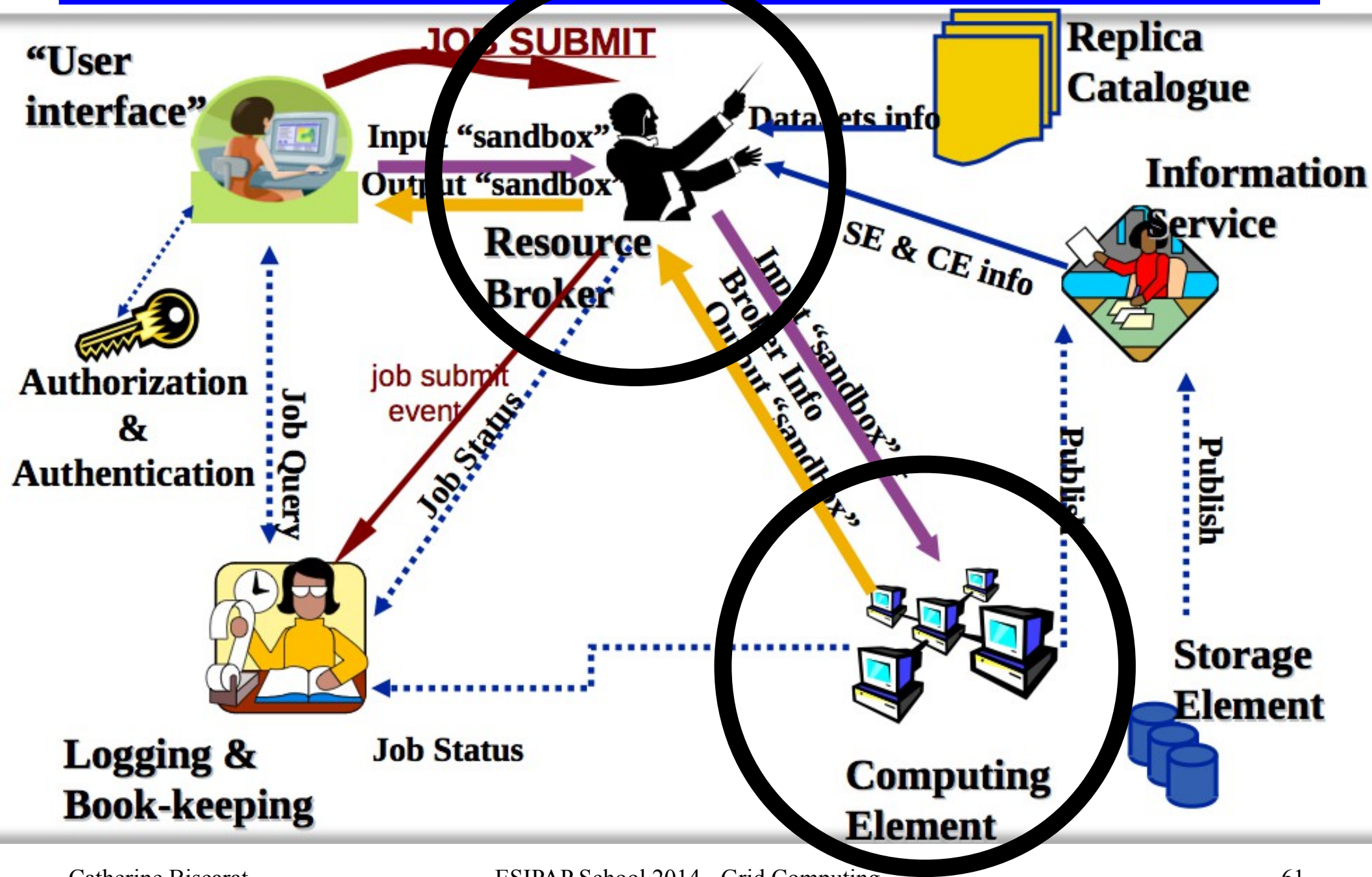
Period: Start year: 2013 Start month: 4
End year: 2014 End month: 3

Official VOs:

<input type="checkbox"/> aegis	<input type="checkbox"/> alice	<input type="checkbox"/> armgrid.grid.am	<input type="checkbox"/> astro.vo.eu-egee.org
<input type="checkbox"/> atlas	<input checked="" type="checkbox"/> auger	<input type="checkbox"/> balticgrid	<input type="checkbox"/> bbmri.nl
<input type="checkbox"/> belle	<input type="checkbox"/> biomed	<input type="checkbox"/> calice	<input type="checkbox"/> cdf
<input type="checkbox"/> cernatschool.org	<input type="checkbox"/> cesga	<input type="checkbox"/> chem.vo.ibergrid.eu	<input type="checkbox"/> cms
<input type="checkbox"/> compchem	<input type="checkbox"/> d4science.research-infrastructures.eu	<input type="checkbox"/> dech	<input type="checkbox"/> desy
<input type="checkbox"/> drihm.eu	<input type="checkbox"/> dteam	<input type="checkbox"/> dzero	<input type="checkbox"/> earth.vo.ibergrid.eu
<input type="checkbox"/> eng.vo.ibergrid.eu	<input type="checkbox"/> enmr.eu	<input type="checkbox"/> env.see-grid-sci.eu	<input type="checkbox"/> envirogrids.vo.eu-egee.org
<input type="checkbox"/> epic.vo.gridpp.ac.uk	<input type="checkbox"/> esr	<input type="checkbox"/> eumed	<input type="checkbox"/> fusion
<input type="checkbox"/> gaussian	<input type="checkbox"/> geant4	<input type="checkbox"/> ghep	<input type="checkbox"/> gilda
<input type="checkbox"/> glast.org	<input type="checkbox"/> gridmosi.ici.ro	<input type="checkbox"/> gridpp	<input type="checkbox"/> hermes
<input type="checkbox"/> hone	<input type="checkbox"/> hungrid	<input type="checkbox"/> iber.vo.ibergrid.eu	<input type="checkbox"/> icecube
<input type="checkbox"/> ict.vo.ibergrid.eu	<input type="checkbox"/> ific	<input type="checkbox"/> ilc	<input type="checkbox"/> infngrid
<input type="checkbox"/> lhcb	<input type="checkbox"/> life.vo.ibergrid.eu	<input type="checkbox"/> lofar	<input type="checkbox"/> lsgrid
<input type="checkbox"/> magic	<input type="checkbox"/> meteo.see-grid-sci.eu	<input type="checkbox"/> mice	<input type="checkbox"/> na62.vo.gridpp.ac.uk
<input type="checkbox"/> ncf	<input type="checkbox"/> nw_ru	<input type="checkbox"/> ops	<input type="checkbox"/> pheno
<input type="checkbox"/> phys.vo.ibergrid.eu	<input type="checkbox"/> planck	<input type="checkbox"/> prod.vo.eu-eela.eu	<input type="checkbox"/> projects.nl
<input type="checkbox"/> pvier	<input type="checkbox"/> see	<input type="checkbox"/> shiwa-workflow.eu	<input type="checkbox"/> snoplus.snolab.ca
<input type="checkbox"/> superbvo.org	<input type="checkbox"/> t2k.org	<input type="checkbox"/> theophys	<input type="checkbox"/> trgridb
<input type="checkbox"/> tut.vo.ibergrid.eu	<input type="checkbox"/> twgrid	<input type="checkbox"/> uniandes.edu.co	<input type="checkbox"/> verce.eu
<input type="checkbox"/> virgo	<input type="checkbox"/> vmed	<input type="checkbox"/> vo.agata.org	<input type="checkbox"/> vo.aginfra.eu
<input type="checkbox"/> vo.apc.univ-paris7.fr	<input type="checkbox"/> vo.complex-systems.eu	<input type="checkbox"/> vo.cta.in2p3.fr	<input type="checkbox"/> vo.formation.idgrilles.fr
<input type="checkbox"/> vo.france-asia.org	<input type="checkbox"/> vo.france-grilles.fr	<input type="checkbox"/> vo.general.csic.es	<input type="checkbox"/> vo.grand-est.fr
<input type="checkbox"/> vo.grif.fr	<input type="checkbox"/> vo.hess-experiment.eu	<input type="checkbox"/> vo.ifisc.csic.es	<input type="checkbox"/> vo.ipnl.in2p3.fr
<input type="checkbox"/> vo.ipno.in2p3.fr	<input type="checkbox"/> vo.irfu.cea.fr	<input type="checkbox"/> vo.landslides.mosaic.org	<input type="checkbox"/> vo.llr.in2p3.fr
<input type="checkbox"/> vo.londongrid.ac.uk	<input type="checkbox"/> vo.lpnhe.in2p3.fr	<input type="checkbox"/> vo.mcia.fr	<input type="checkbox"/> vo.msfg.fr
<input type="checkbox"/> vo.mure.in2p3.fr	<input type="checkbox"/> vo.neugrid.eu	<input type="checkbox"/> vo.northgrid.ac.uk	<input type="checkbox"/> vo.panda.gsi.de
<input type="checkbox"/> vo.paus.pic.es	<input type="checkbox"/> vo.plgrid.pl	<input type="checkbox"/> vo.rhone-alpes.idgrilles.fr	<input type="checkbox"/> vo.sbg.in2p3.fr
<input type="checkbox"/> vo.scotgrid.ac.uk	<input type="checkbox"/> vo.southgrid.ac.uk	<input type="checkbox"/> vo.u-psud.fr	<input type="checkbox"/> vo.up.pt
<input type="checkbox"/> voce	<input type="checkbox"/> xenon.biggrid.nl	<input type="checkbox"/> zeus	

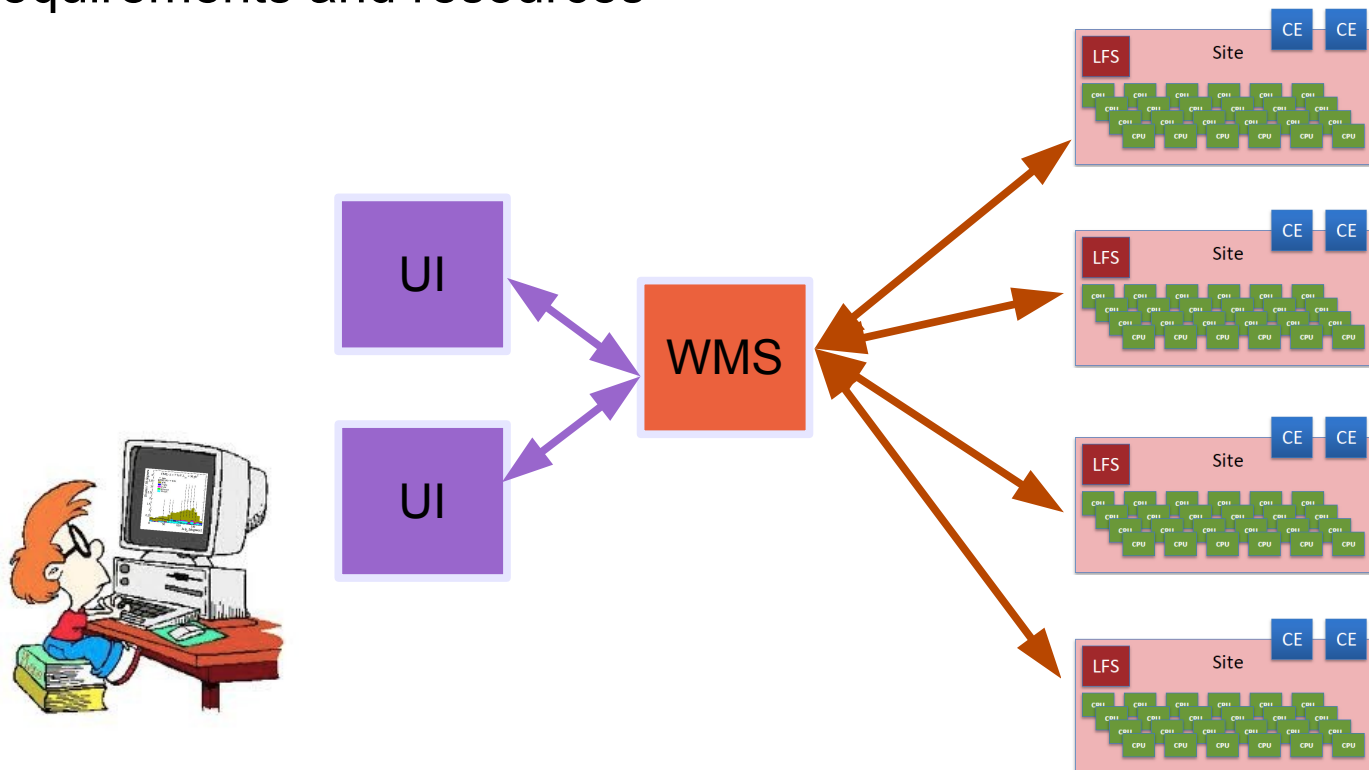
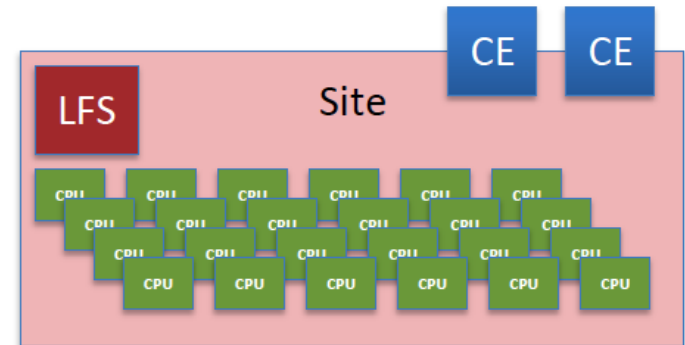
<https://accounting.egi.eu/egi.php>

A word on

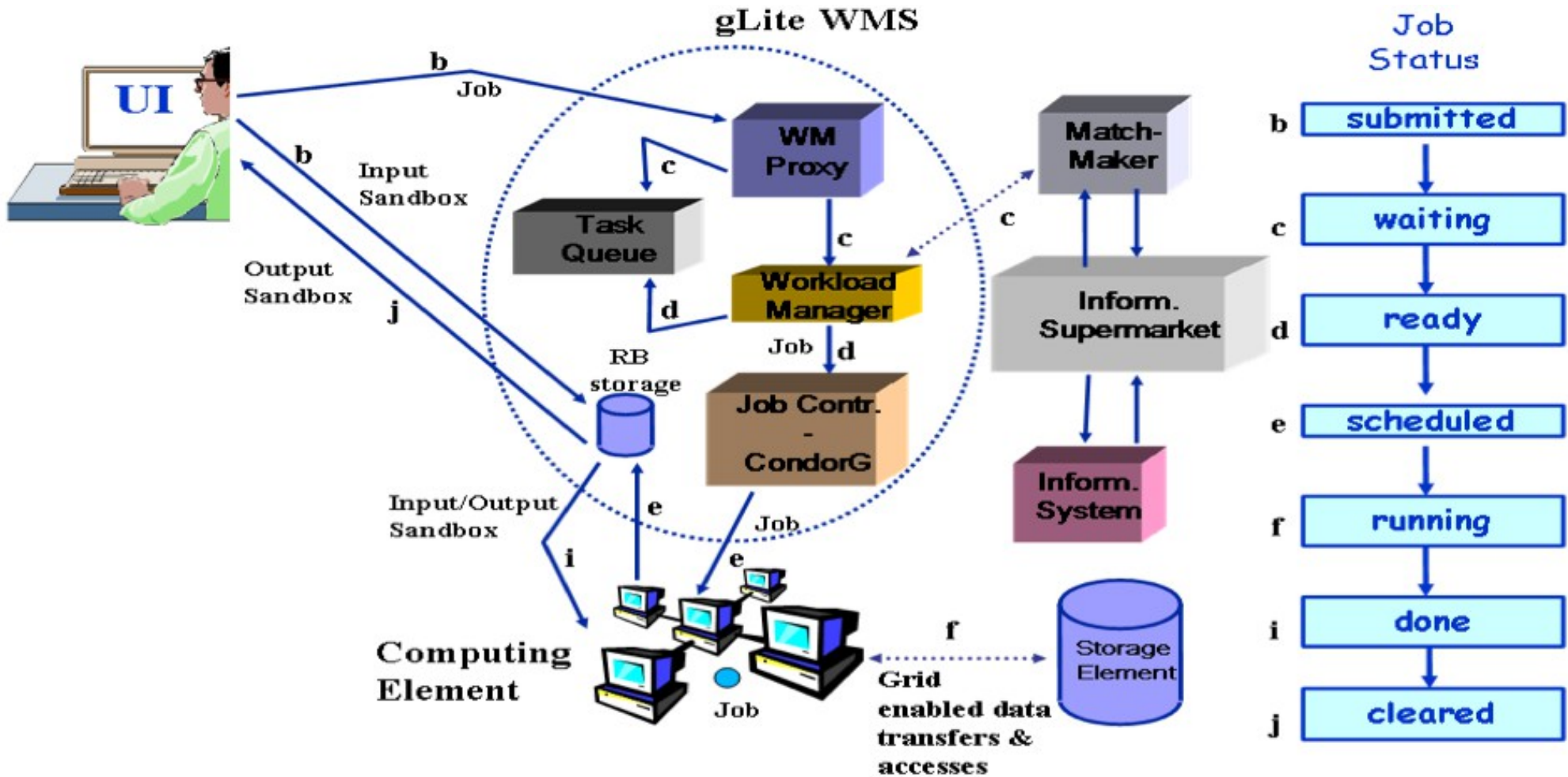


Workload management

- Computing elements
 - Gateways to the local computing resources
- Workload management system
 - akas Resource broker
 - Matches requirements and resources

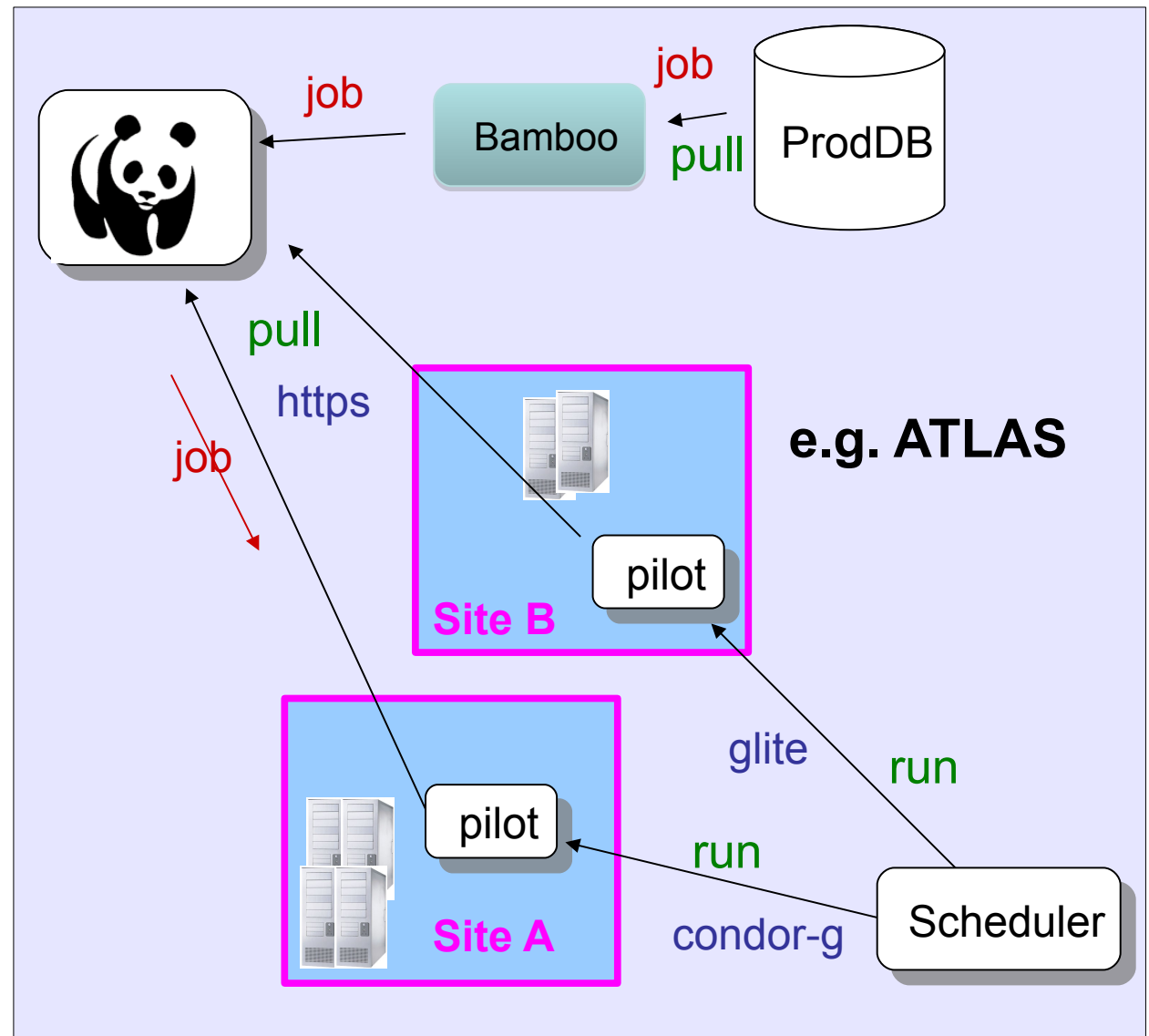


WMS – jobs route



Pilot job mechanism

- Place holder jobs get queued at the sites
- Real jobs get pulled according to the available resources and the environment
- Mechanism to change user ID (gLExec)
- Each LHC experiment uses pilot jobs



Job Description Language

attribut job

```
Executable = "gridTest";  
StdError = "stderr.log";  
StdOutput = "stdout.log";  
InputSandbox = {"/home/joda/test/gridTest"};  
OutputSandbox = {"stderr.log", "stdout.log"};
```

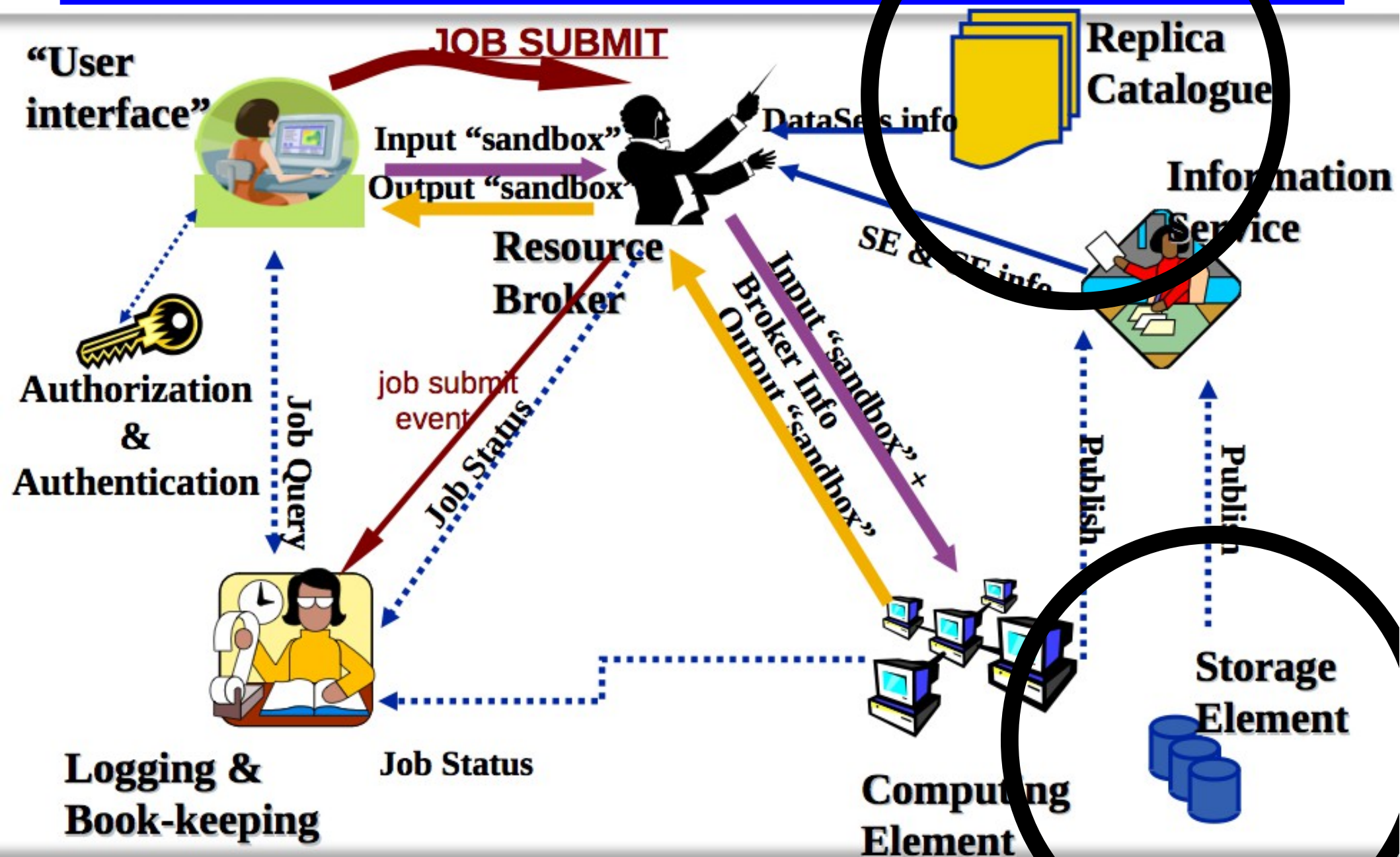
attribut données

```
InputData = "lfn:testbed0-00019";  
DataAccessProtocol = "gridftp";
```

attributs ressources

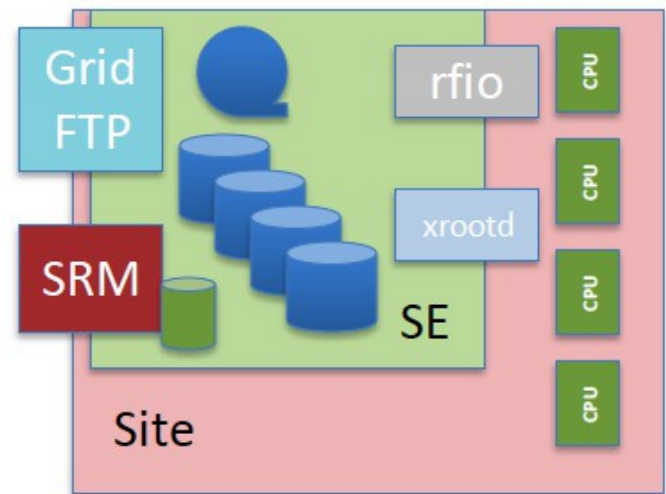
```
Requirements = other.Architecture=="INTEL" && \  
                other.OpSys=="LINUX" && other.FreeCpus\  
                >=4;  
Rank = other.GlueHostBenchmarkSF00;
```


A word on

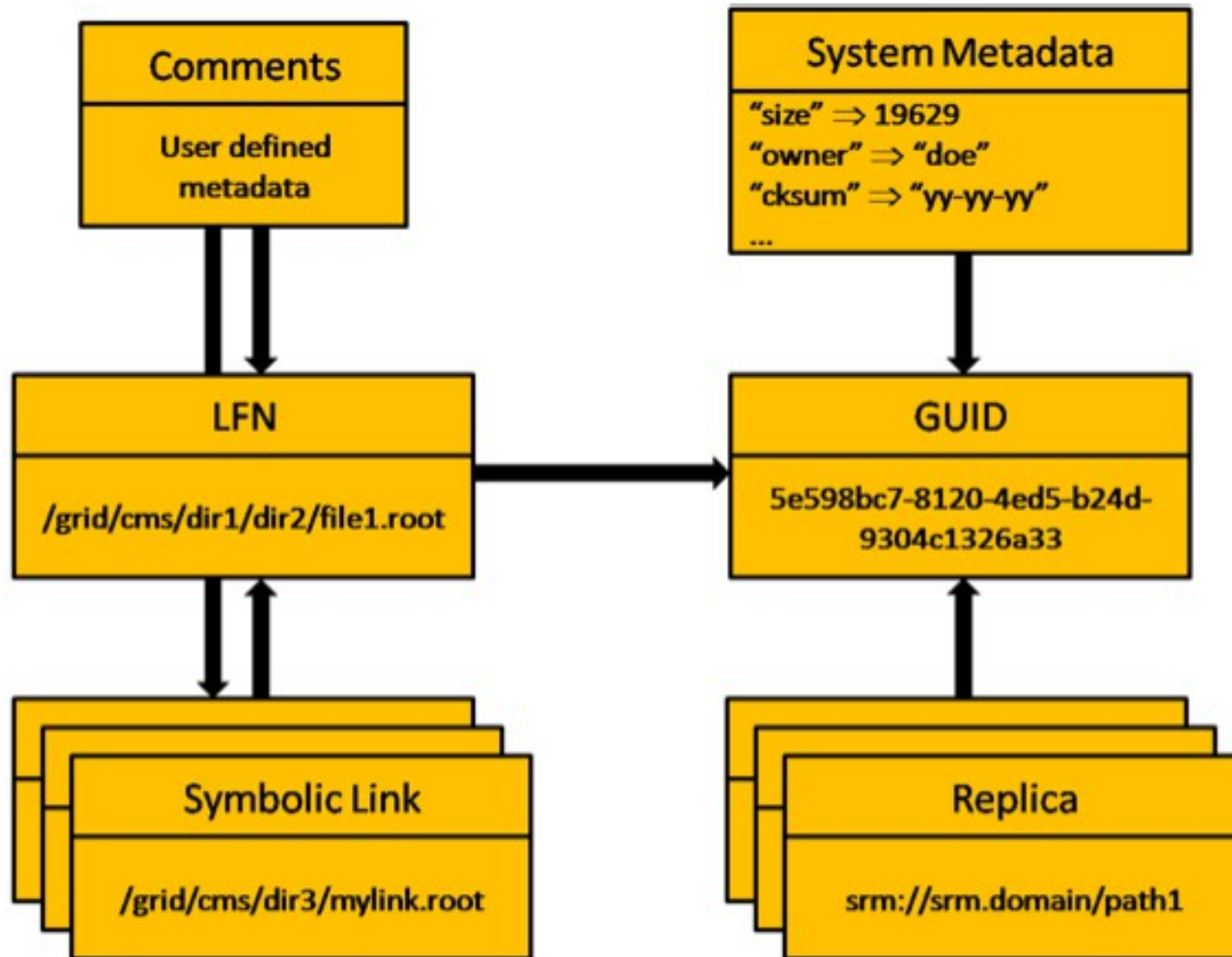


Data management

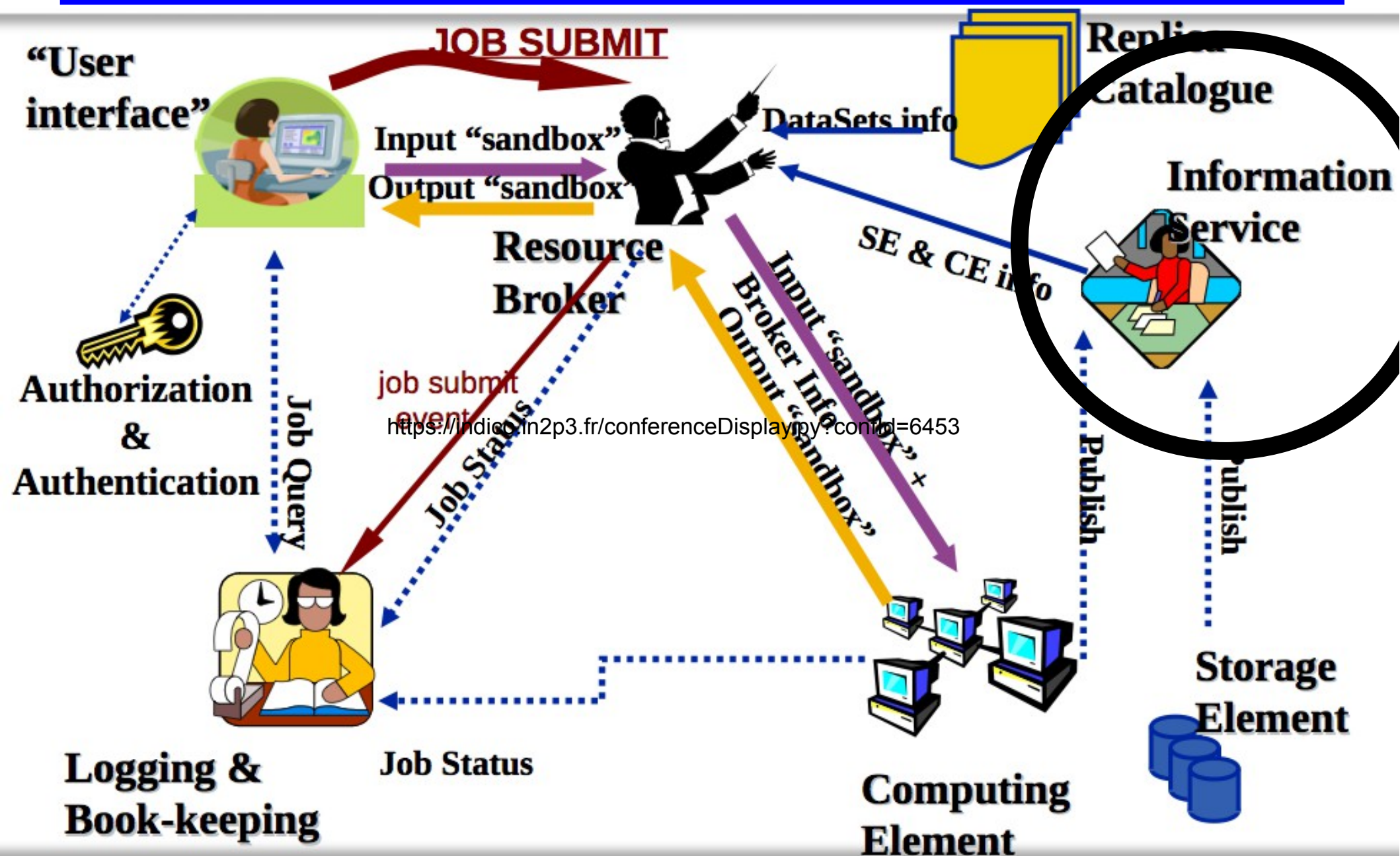
- **Storage elements (SE)**
 - External interface to physical storage at sites
 - Different protocols (rfio, srm,)
 - Different storage system (dCache, xrootd, ...)
- **Storage resource manager (SRM)**
 - Hide heterogeneous systems
 - Handle authorisation
- **Local File Catalog (LFC)**
 - Locate files on the grid
 - Keep track of the file (user's name ↔ file location)
- **File Transfer Service (FTS)**
 - Multi-VO service
 - Handles prioritisation



LFC architecture

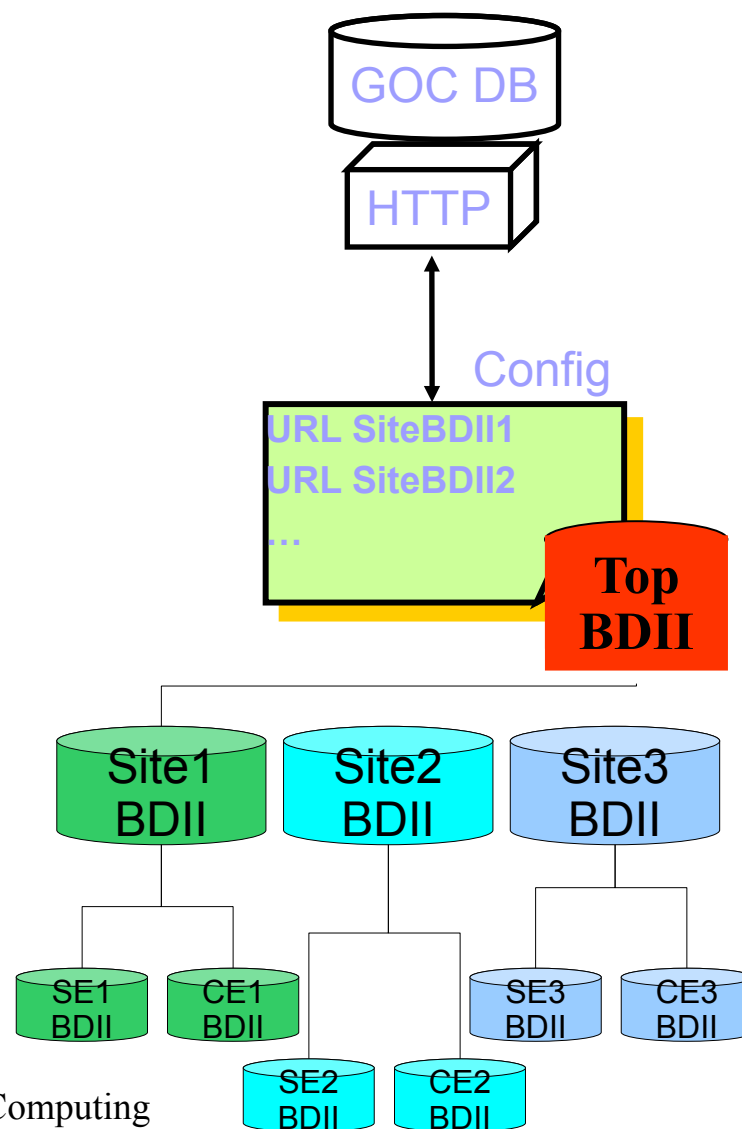


A word on



Information system

- In charge of collecting the information on the status of the site and services
 - Format understandable by both parties
- Sites publish:
 - Status of the services
 - Their resources
- VOs publish:
 - The software installed on the sites
- BDII (GlueSchema):
 - Sites BDII
 - Top BDII
 - Commands:
 - Ldapsearch
 - lcg-infosites
- GOC DB: reference for all the EGI sites
<http://goc.egi.eu/>





Site: IN2P3-LPSC

Laboratoire de Physique Subatomique et de Cosmologie de Grenoble



Edit

Browse

- My Sites
- Projects
- NGIs
- Sites
- Service Groups
- Services

Add

- Add Site
- Add Service Group
- Add Service
- Add Downtime

Downtimes

- Recent & Planned
- Active & Imminent

About GOCD 5.2

- Doc, Help & Support

Search

Submit

User Status

Registered as:
Catherine Biscarat

[View Details](#)
[Manage Roles](#)

Contact



E-Mail	grid.admin@lpsc.in2p3.fr
Telephone	(33) 4 76 28 41 58
Emergency Telephone	(33) 4 76 28 41 58
CSIRT Telephone	(33) 4 76 28 41 58
CSIRT E-Mail	grid.security@lpsc.in2p3.fr
Emergency E-Mail	
Helpdesk E-Mail	

Project Data



NGI/ROC	NGI_FRANCE
Infrastructure	Production
Certification Status	Certified Change
Scope(s)	EGI

Networking



Home URL	http://lpsc.in2p3.fr/
GIIS URL	ldap://lpsc-bdii.in2p3.fr:2170/mds-vo-name=IN2P3-LPSC,o=grid
IP Range	0.0.0.0/255.255.255.255
Domain	in2p3.fr

Location



Country	France
Latitude	0
Longitude	0
Time Zone	Europe/Paris
Location	

Site Extension Properties



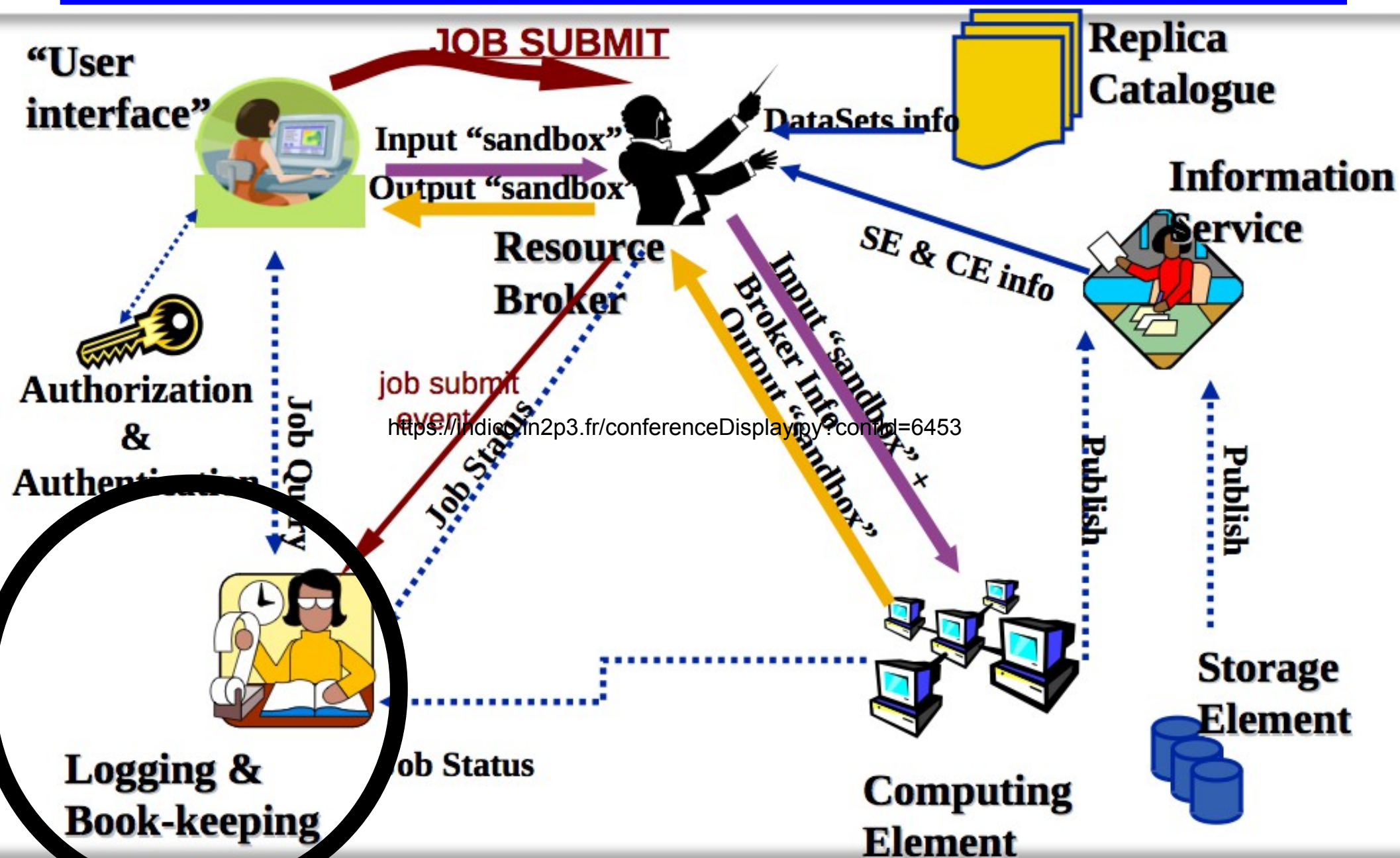
Name	Value	Edit	Remove
------	-------	------	--------

Services 

Hostname (service type)	URL	Production	Monitored	Scope(s)
lpsc-cream-ce.in2p3.fr (APEL)		✓	✓	EGI
lpsc-bdii.in2p3.fr (Site-BDII)		✓	✓	EGI
lpsc-vobox.in2p3.fr (VO-box)		✓	✓	EGI
lpsc-ce.in2p3.fr (CREAM-CE)		✓	✓	EGI
lpsc-cream-ce.in2p3.fr (CREAM-CE)		✓	✓	EGI
lpsc-se-dpm-server.in2p3.fr (SRM)		✓	✓	EGI
lpsc-ce.in2p3.fr (gLExec)		✓	✓	EGI
lpsc-cream-ce.in2p3.fr (gLExec)		✓	✓	EGI
lpsc-ce.in2p3.fr (eu.egi.MPI)		✓	✓	EGI
lpsc-cream-ce.in2p3.fr (eu.egi.MPI)		✓	✓	EGI
lpsc-perfsonar.in2p3.fr (net.perfSONAR.Bandwidth)		✓	✓	EGI
lpsc-perfsonar2.in2p3.fr (net.perfSONAR.Latency)		✓	✓	EGI
lpsc-ce2.in2p3.fr (CREAM-CE)		✓	✓	EGI
lpsc-ce2.in2p3.fr (eu.egi.MPI)		✓	✓	EGI
lpsc-ce2.in2p3.fr (gLExec)		✓	✓	EGI

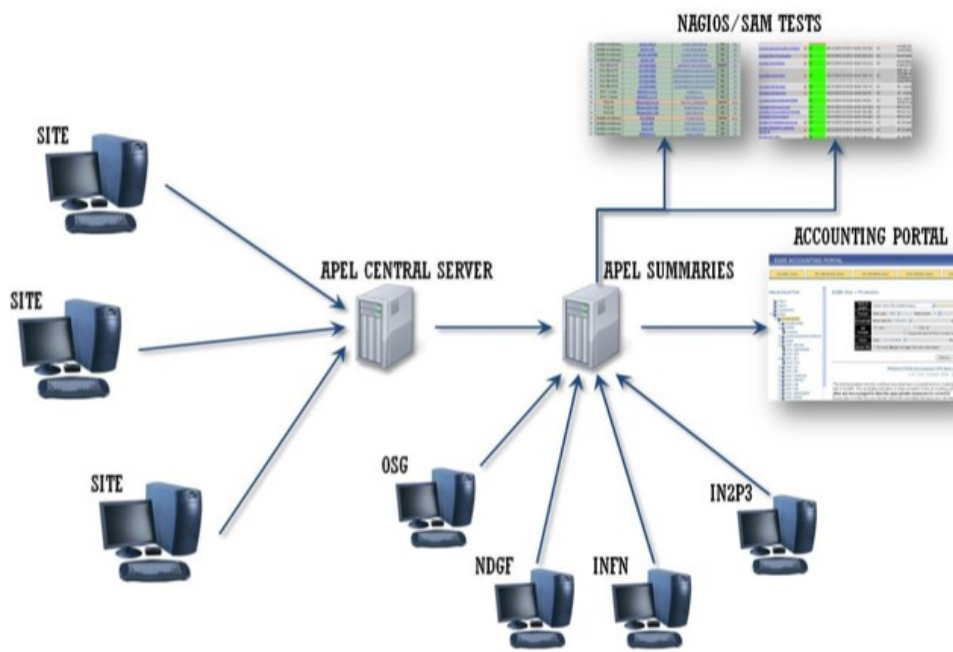
 Add Service

A word on



Accounting

- Accounting for the computing and storage resources
 - Sites share VOs
 - MoU signed with “pledges”



EGI ACCOUNTING PORTAL

GLOBAL View | VO MANAGER View | VO MEMBER View | SITE ADMIN View | USER View | REPORTS | METRICS PORTAL | LINKS

Tier1 Tier2 Countries Production OSG Unregistered VO_Discipline VO_Metrics CUSTOM

Hierarchical Tree | EGI View --> Production

Data to graph: Norm. Sum CPU (kSI2K-hours) Normalised CPU time to a reference value of 1000 Spectint2000

Period: Start year: 2013 Start month: 4 End year: 2014 End month: 3

Groupings: Show data for: REGION as a function of: VO

VO Groups: LHC TOP 10 ALL Custom

Chart: Type: GROUP BAR Scale: LINEAR

dteam VO: Exclude dteam and ops VO's jobs information

Local Jobs: Grid Jobs Only Grid Jobs and Local Jobs Local Jobs Only

Refresh

Normalised CPU time (kSI2K) by REGION and VO.
LHC VOs. April 2013 - March 2014.

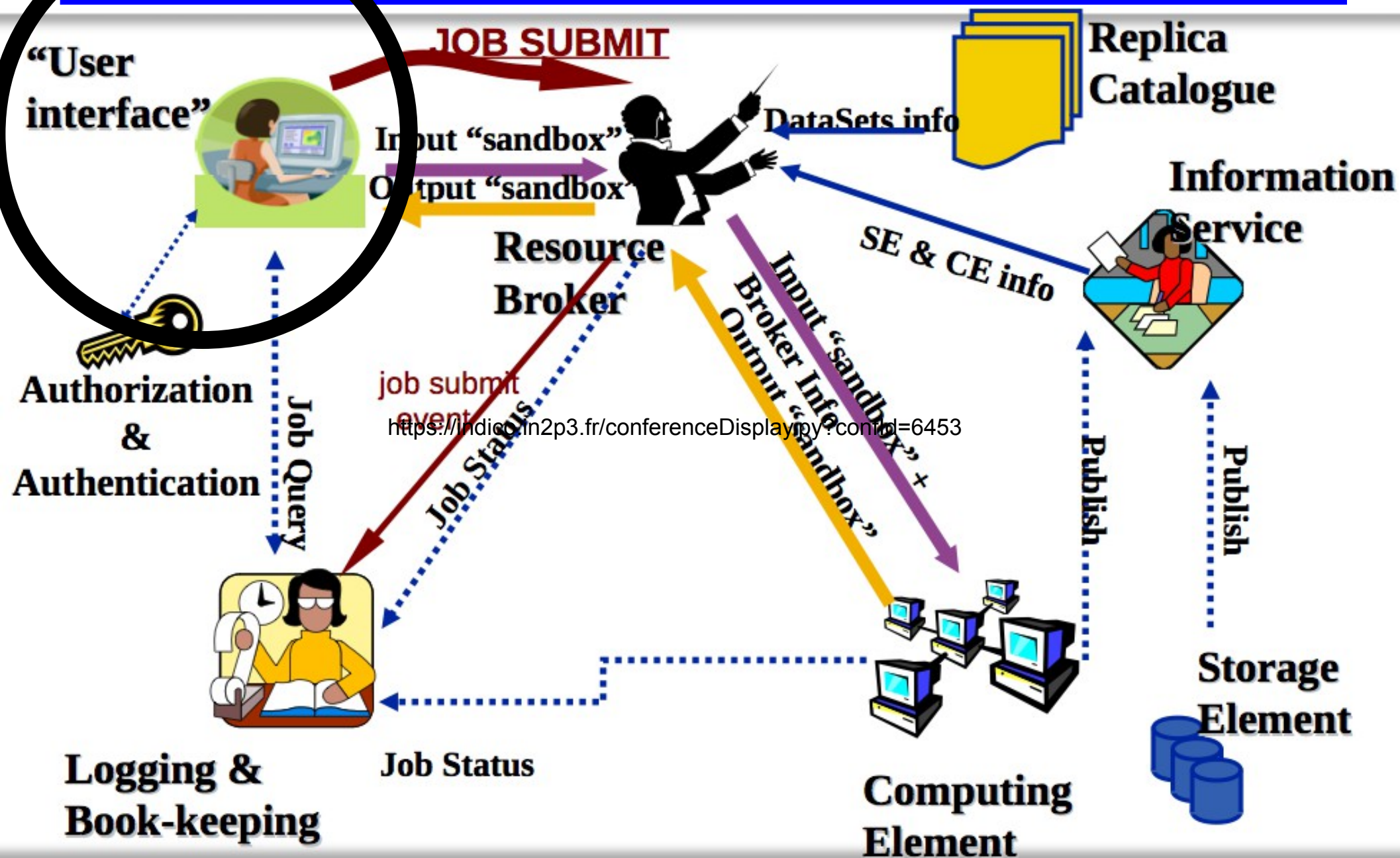
The following table lists the sites that have not published accounting data to GOC during the last 3 months. This probably indicates a major problem in the accounting system of the site so listed sites are encouraged to take the appropriate measures to correct it.

sites NOT publishing accounting data to GOC in the last 3 months	
Region	Sites
AsiaPacific	IR-IPM-HEP, JP-KEK-CRC-02, PK-CIT, TW-EMI-PPS, TW-NTU-HEP
NGI_DE	mainzgrid, UNI-BONN, UNI-SIEGEN-HEP
NGI_FI	CSC, FI_AA, FI_Aalto, FI_Helsinki, FI_JYU, FI_LUT, FI_Oulu, FI_TUT, FI_UEF, FI_UTU
NGI_GRNET	HG-04-OTI-CEID
NGI_IBERGRID	CIEMAT-TIC
NGI_IL	HRL_KZ, IL_COMP, IL_IUCC_IG
NGI_ME	MREN-01-CIS
NGI_NDGF	EENet, IMCSUL, RTUETF, UNICPH-NBI
NGI_NL	RUG-CIT
NGI_RO	RO-13-ISS
NGI_UA	UA-IRE, UA_ICMP_ARC, UA_ICYB_ARC

European Grid Infrastructure

<http://accounting.egi.eu/egi.php>

A word on



User Interface

- For the user to access to the Middleware
- Sites generally provides a UI machine
- Installed:
 - Command lines
 - Applications

```
voms-proxy-init --voms vo.france-grilles.fr
glite-ce-job-submit -a -r lpsc-ce.in2p3.fr:8443/cream-pbs-france_grilles test.jdl
glite-ce-job-list lpsc-ce.in2p3.fr:8443
glite-ce-job-status https://lpsc-ce.in2p3.fr:8443/CREAM050473948
```

#with

```
bash-3.2$ cat test.jdl
Executable = "/bin/sleep";
arguments="180";
StdOutput = "std.out";
StdError = "std.err";
bash-3.2$
```

Monitoring

Examples to show

Nagios

gstat

Jobs atlas

Jobs lpsc

Alice

Alice problems

Running jobs: 236092
Transfer rate: 11.41 GiB/sec



Going forward

ographer
/BKG
NGA, GEBCO

Still improving the grid

- Behind the success of the grid
 - High complexity
 - Lot of human power
- We have gained in experience and we are still improving
 - Very stable network → new data model
 - **Federated storage**
 - Building common tools/interfaces
- Meanwhile we were developing the grid
 - Other areas do treat enormous amount of data
 - Standards tools have been developed
 - The “cloud” is born



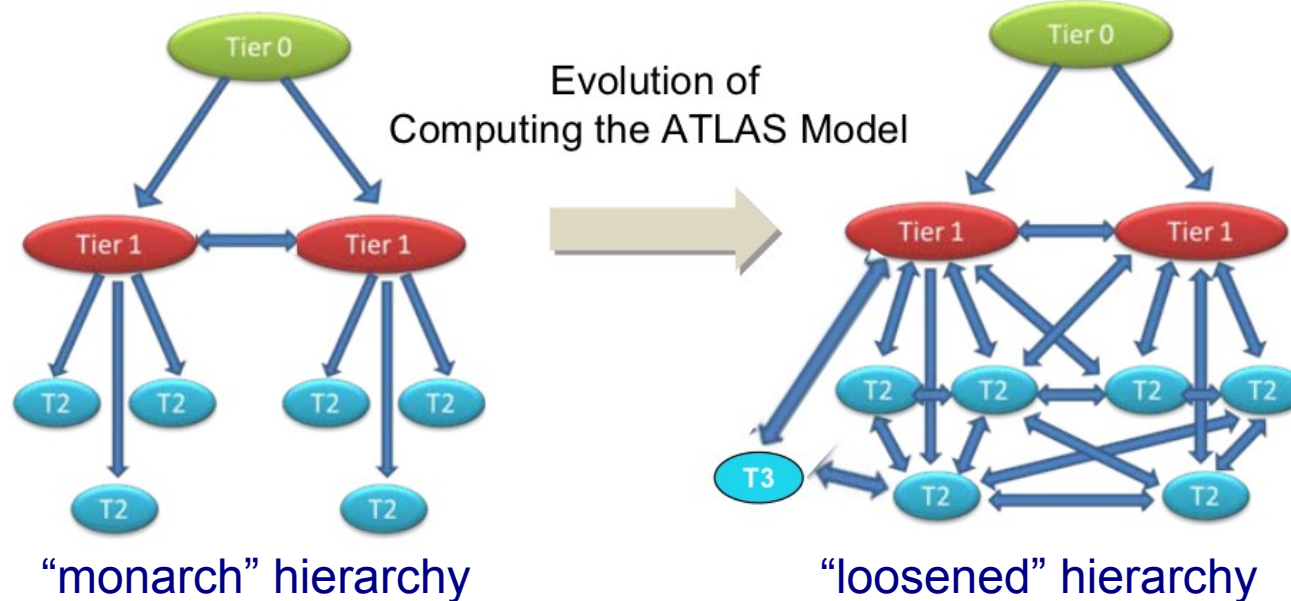
Network

“The Network infrastructure is the most reliable service we have”

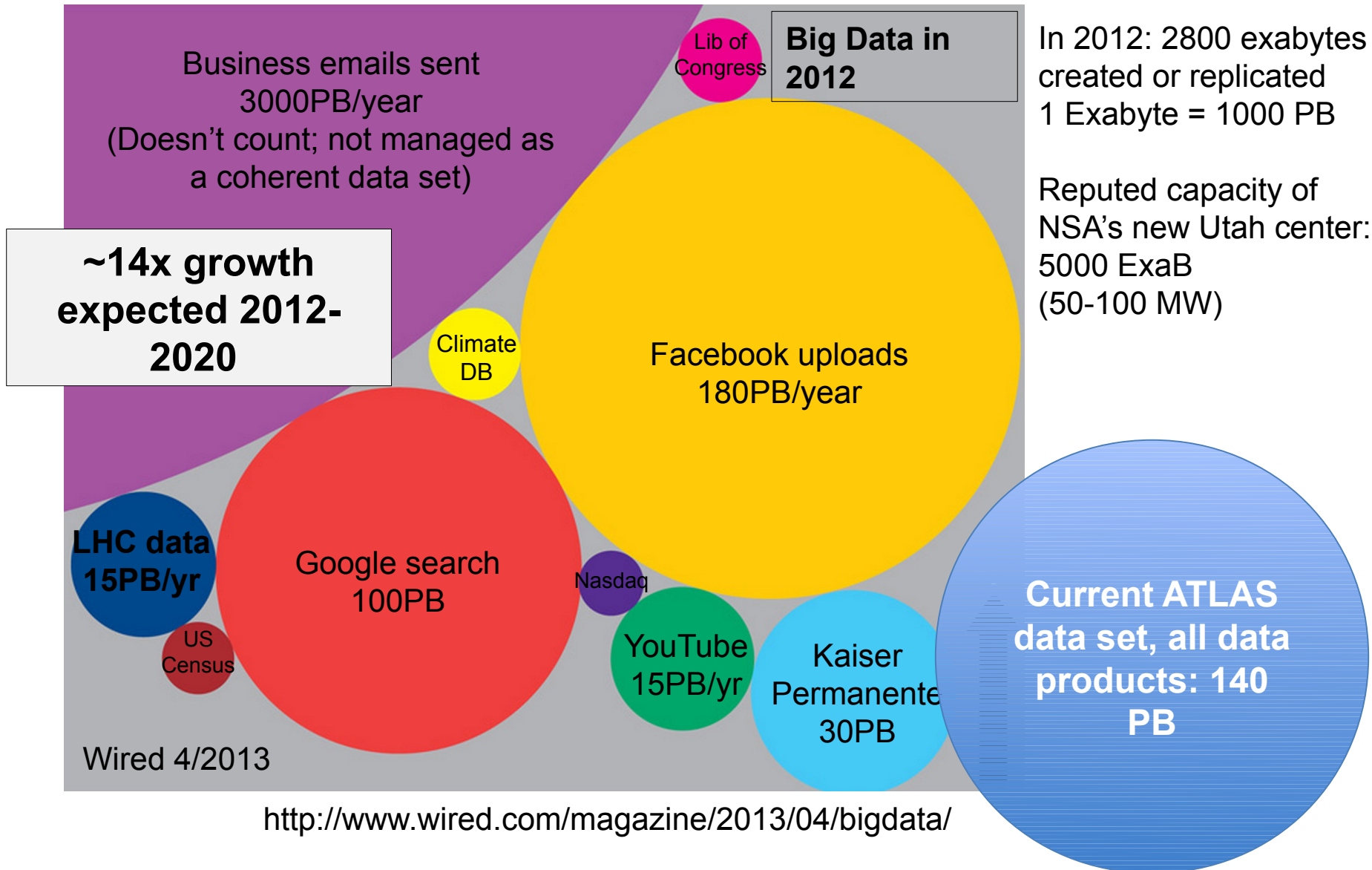
“Network Bandwidth (rather than disk) will need to scale more with users and data volume”

Ian Bird, WLCG project leader

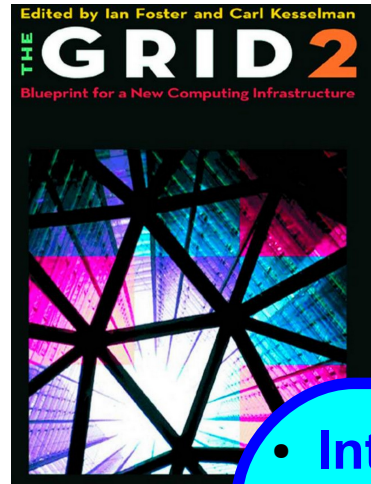
LHC Open Network Environment (LHCONE)



Data set sizes



The “Cloud”



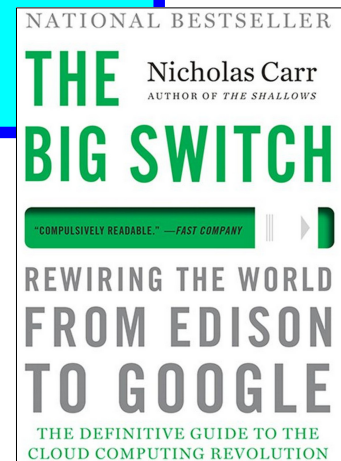
The Grid: 1998 and 2003 (2nd Ed.)

*Grid is used by analogy with the electric power grid...
has had a dramatic impact on human capabilities...*

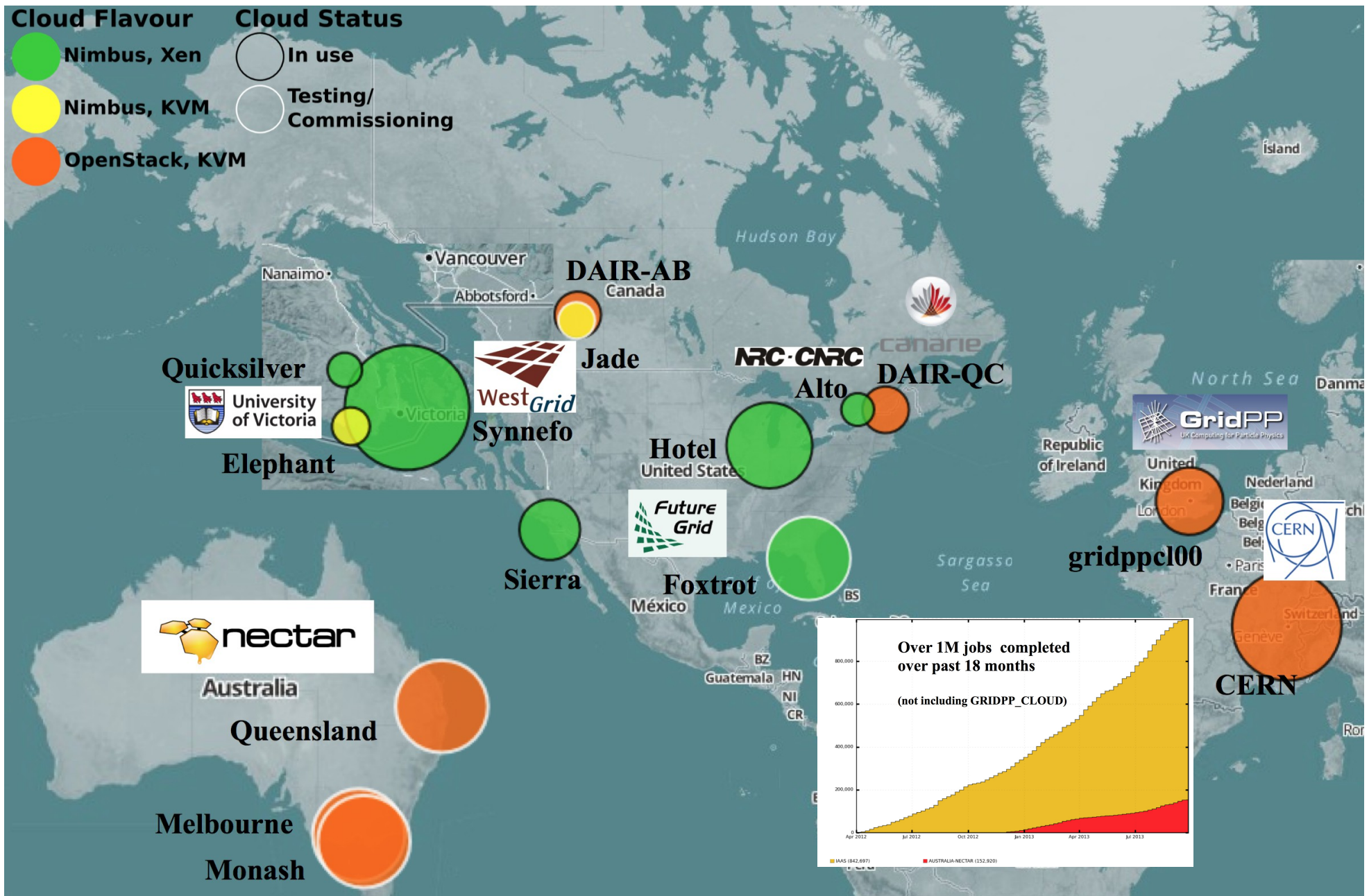
- **Introduces the concept of virtualisation**
 - Isolates the user software (virtual) from physical hardware (real)
 - Remember that grid sites are heterogenous
- Flexible and dynamic resource sharing
- On-demand usage of resources
- Virtualisation already proved to be helpful “on the grid” (CVMFS)
- Clouds already successfully used by experiments
 - In many domains

The Big Switch [to the Cloud]: 2009

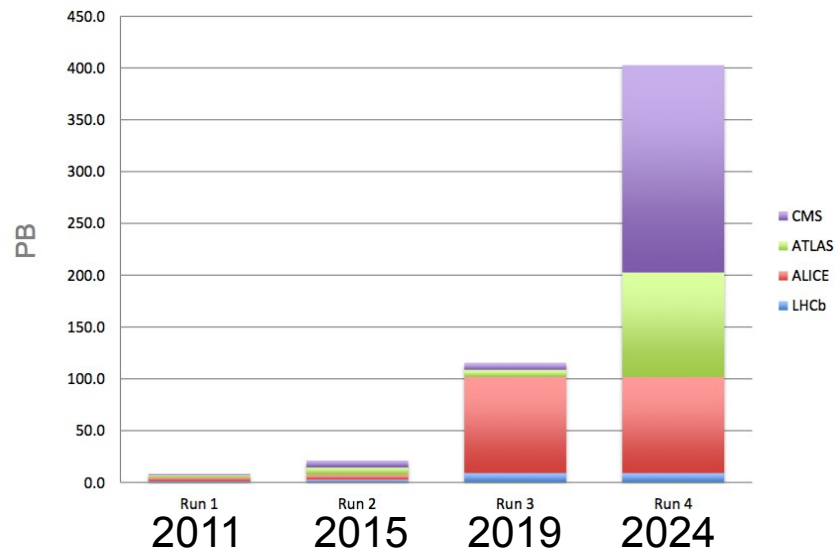
*Computing is turning into a utility... will ultimately
change society as completely as cheap electricity
did...*



Grid-of-clouds used by ATLAS (LHC)



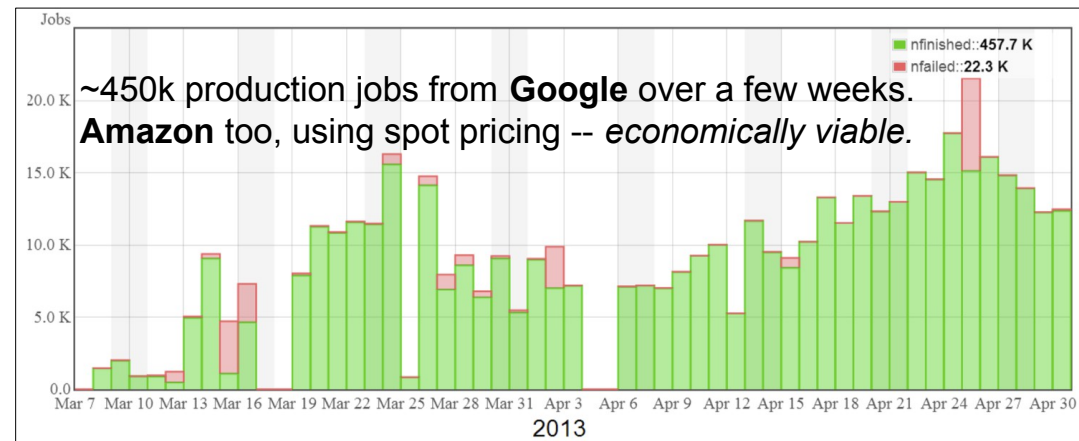
Opportunistic computing by LHC



Evolution of RAW data set size for the LHC experiments

- derived data is not included
- cleaning of deprecated data is included

- **Use of existing clouds**
 - Private
 - Academic
- **Use of supercomputers**
 - Empty cycles
 - Idea: cloud-like site



Titan (Oak Ridge)

Supercomputers

Remember

- Grid is ideal for a large number of independent jobs (**High Throughput Computing**)

Needs for supercomputers

- One task must be executed on hundreds of cores
 - Very fast communication between cores
 - Large available memory
- HPC (**High Performance Computing**)
- Pyramidal Tiers-0/1/2 in Europe
 - France: “Curie” with 92 000 cores
- Seismology, mathematics, chemistry, aerodynamics ... + astro particles

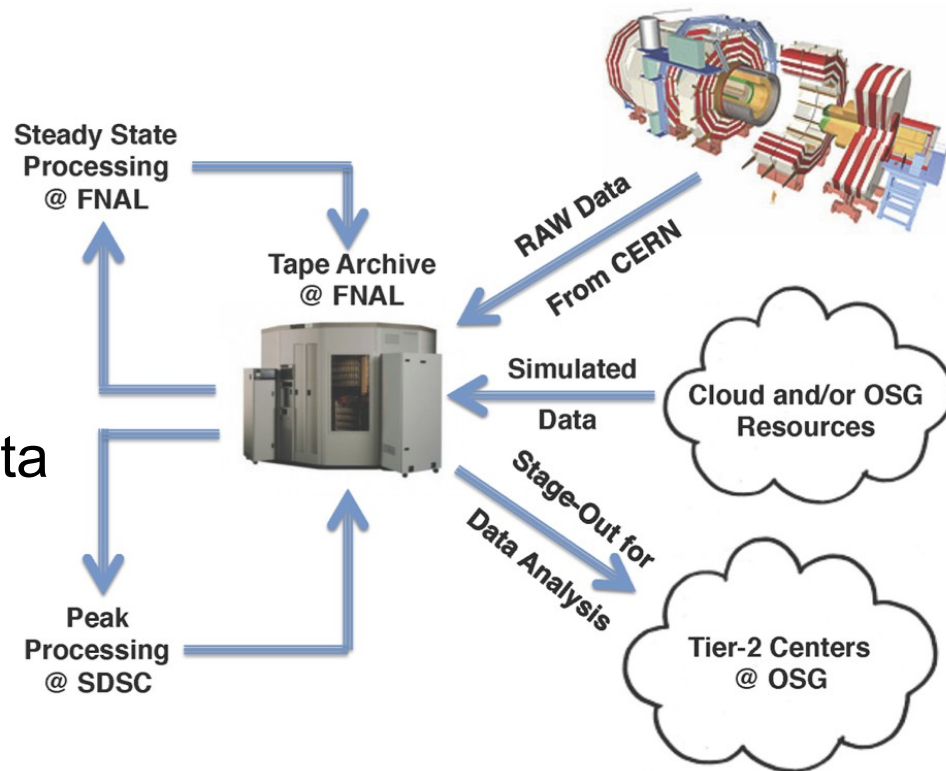


Usage of HPC centres for LHC

CMS using SDSC's Gordon Supercomputer

1024 Sandy Bridge nodes, 16 cores/node, 4 GB/core
Plus large memory supernodes
300 TB SSD storage

- early processing of 2012 parked data
- 125 TB in → 150 TB out

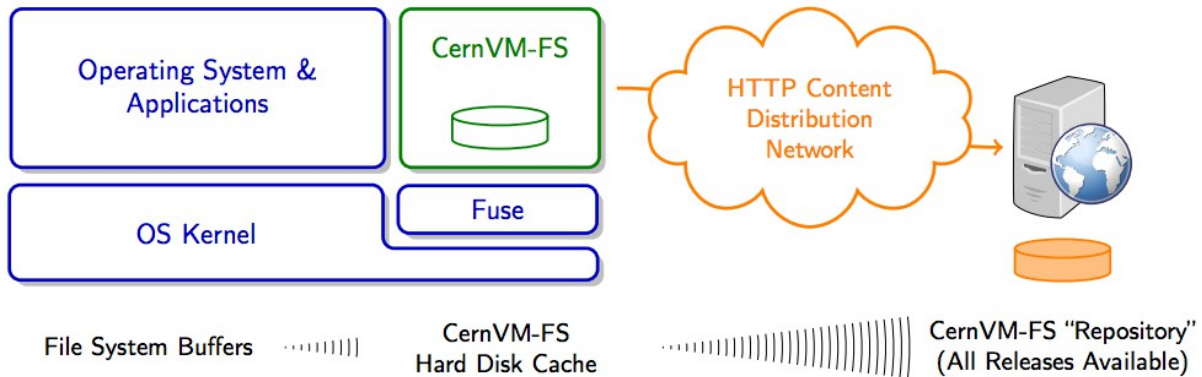


Number of HPC T2s

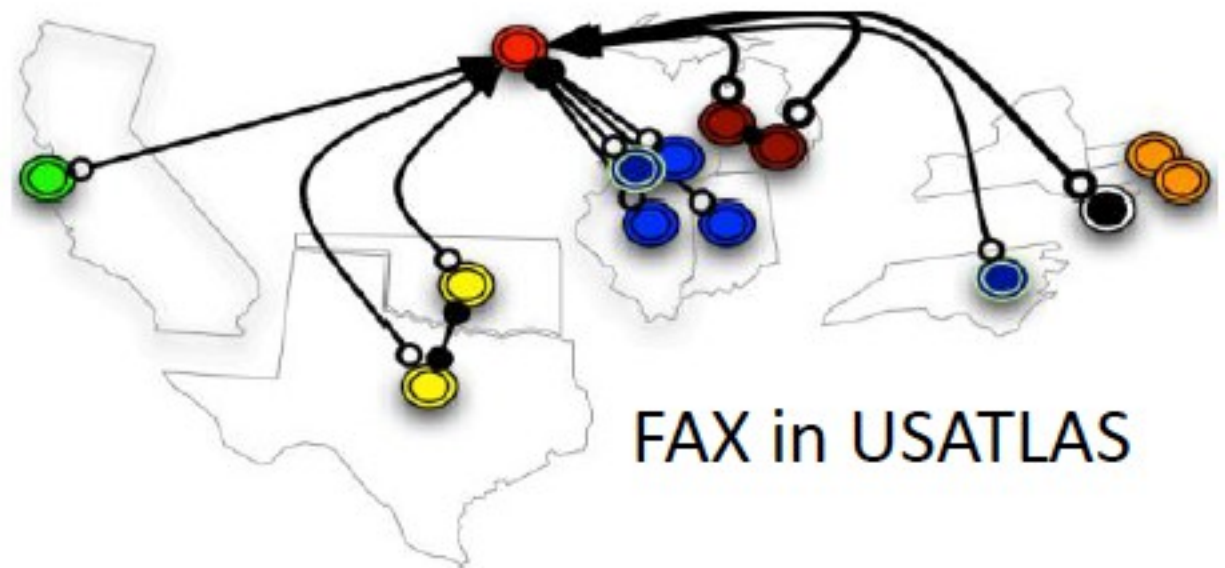
→ many other initiatives in this area

LHC Computing Evolution

Growing uniformity counters growing scale and complexity.

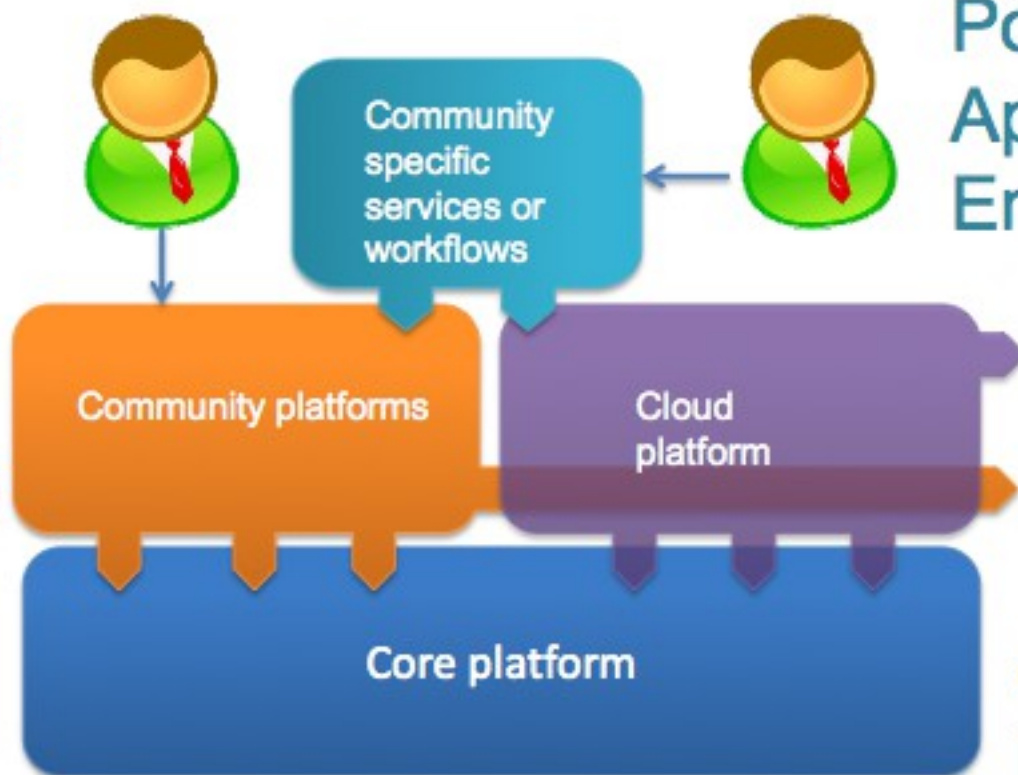


Data federation



EGI platforms

ARC,
dCache
gLite
Globus
QCG
Unicore



Portals, gateways,
Applications, wflow
Engines



Compute, Storage
Data archives,
Desktop Grids,
HPC

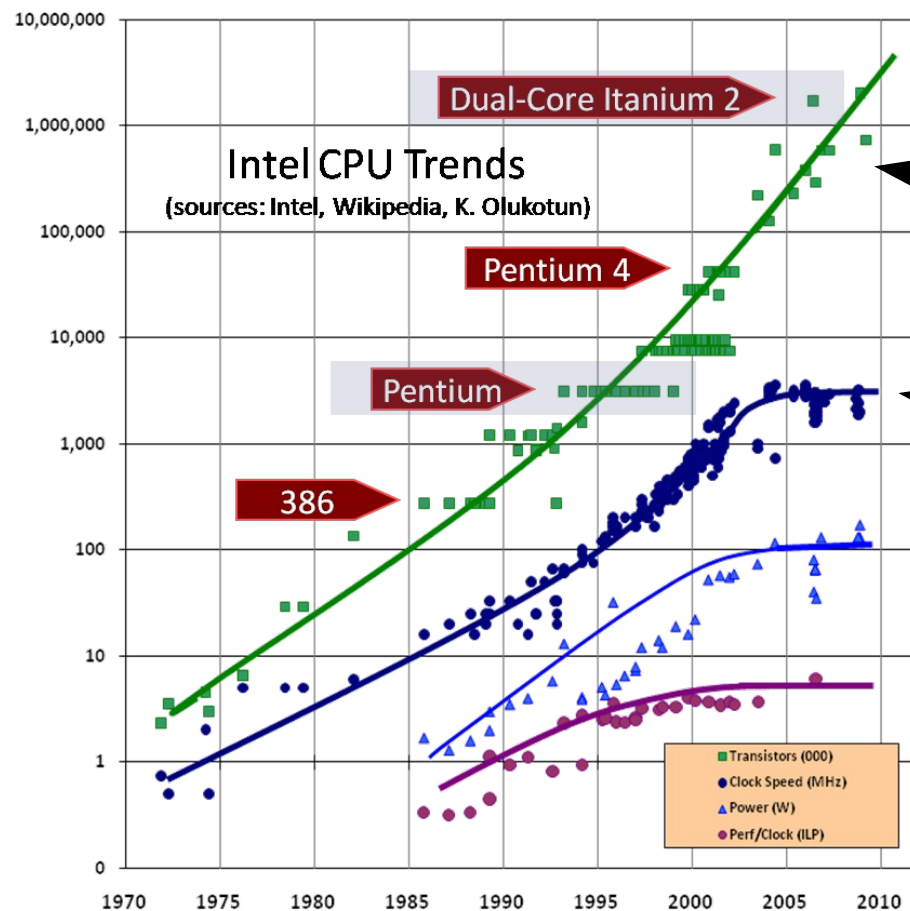
Monitoring, Accounting,
EGI Helpdesk
Service registry and discovery

EGI service evolution

- 'Old' Model
 - Isolated Technology Platforms
 - High Throughput Computing
- 'Current' Model
 - Integrated Technology Platforms
 - HPC, HTC, Data
- 'Future' Model
 - Federated Cloud Platform
 - Community Platforms

2005: “The free lunch is over”

- Les performances des CPUs conventionnelles **ne progressent plus**
- Si les besoins en HEP augmentent, il faut **s'adapter aux nouvelles architectures** : multi-cores, many-cores, GPU (console de jeux)



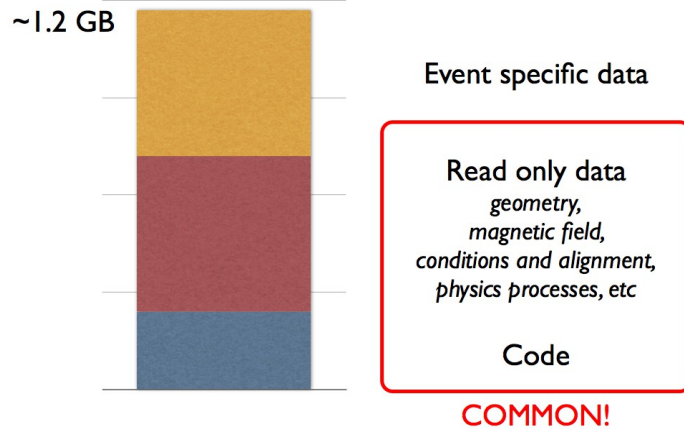
Multi-cores **actuelles**

- **Limitation mémoire et IO par coeur**

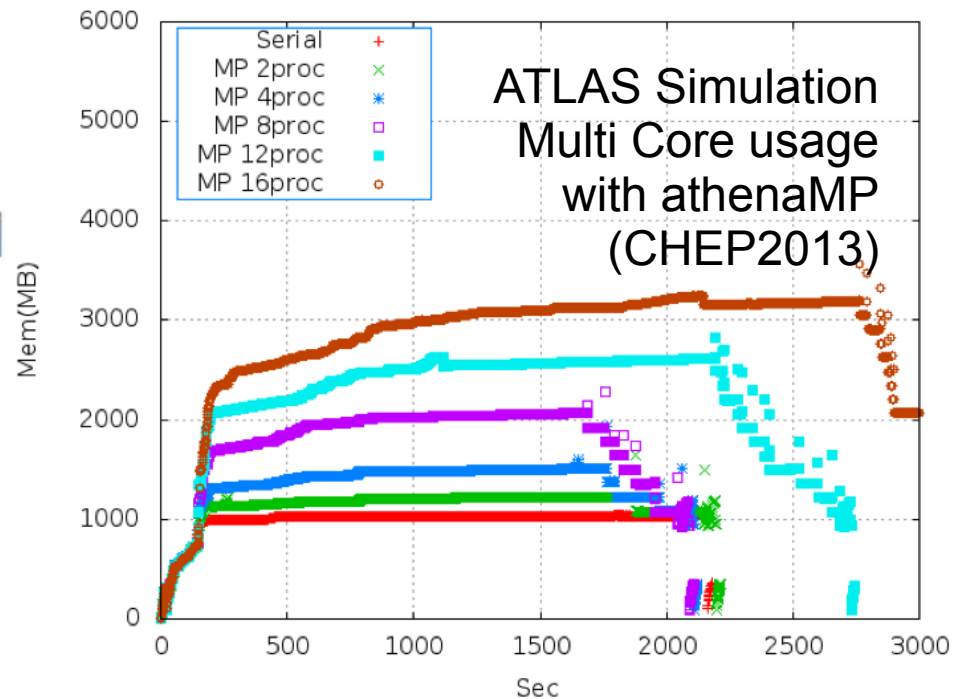
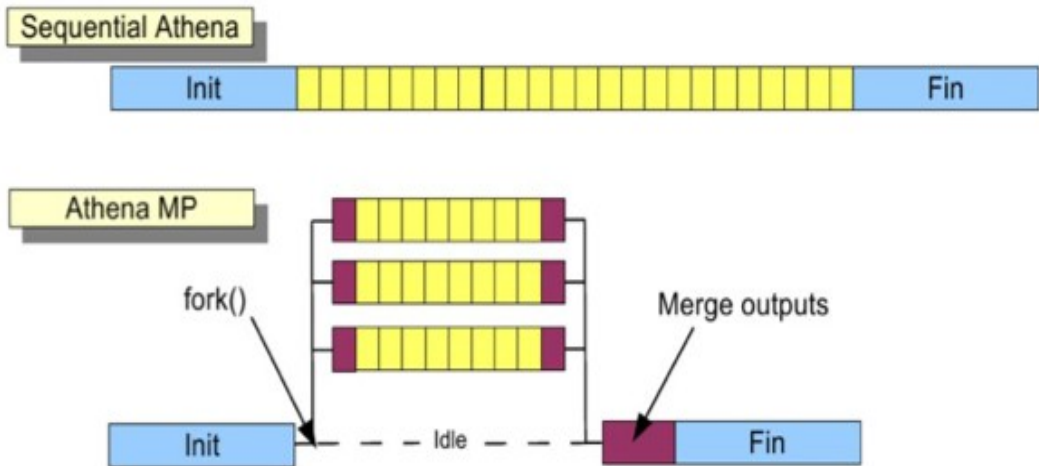
Fréquence d'horloge atteint un plateau

Saving Memory

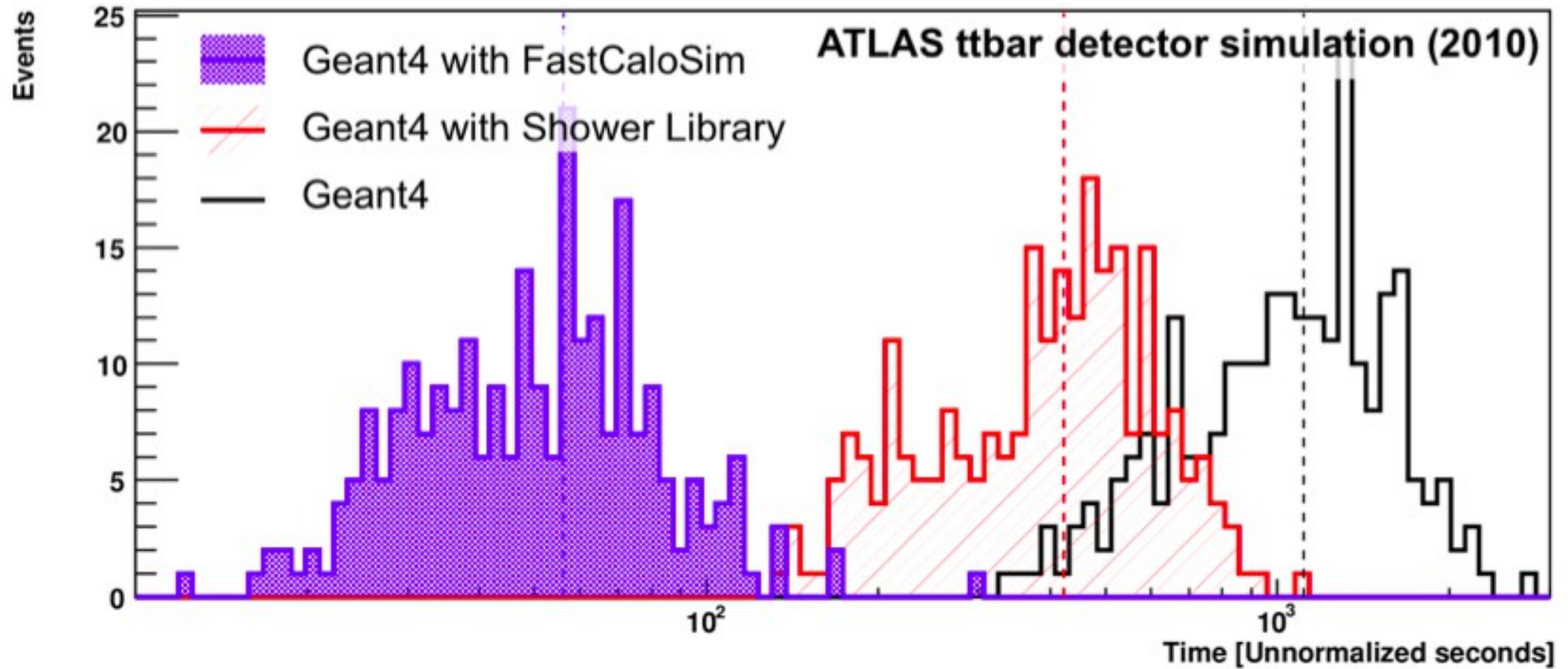
CMS offline software memory budget



ATLAS



Saving time



CHEP2013

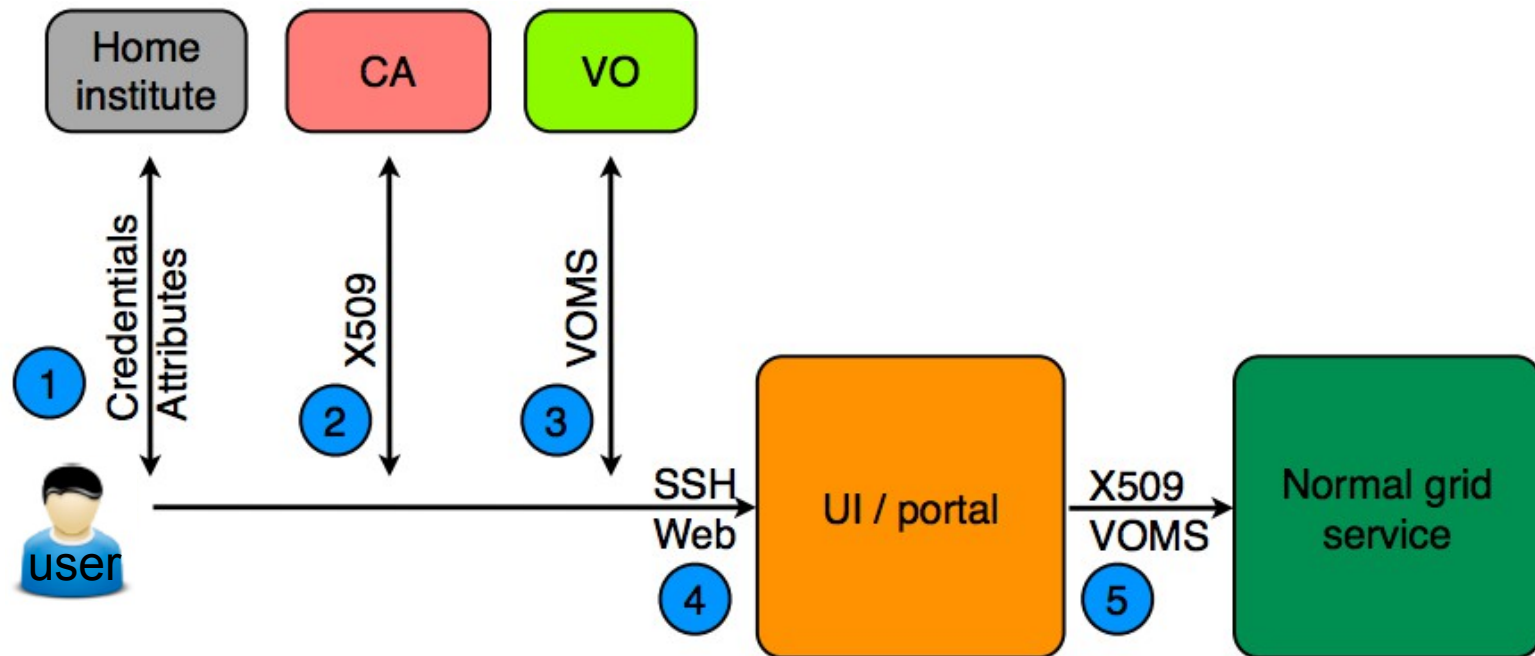
Running jobs: 236092
Transfer rate: 11.41 GiB/sec



Accessing “the” grid

ographer
/BKG
NGA, GEBCO

Authentication



User Interface

- For the user to access to the Middleware
- Sites generally provides a UI machine
- Installed :
 - Command lines
 - Applications

```
voms-proxy-init --voms vo.france-grilles.fr
glite-ce-job-submit -a -r lpsec-ce.in2p3.fr:8443/cream-pbs-france_grilles test.jdl
glite-ce-job-list lpsec-ce.in2p3.fr:8443
glite-ce-job-status https://lpsec-ce.in2p3.fr:8443/CREAM050473948
```

#with

```
bash-3.2$ cat test.jdl
Executable = "/bin/sleep";
arguments="180";
StdOutput = "std.out";
StdError = "std.err";
bash-3.2$
```

Global Grid User Support



Did you know...

[Documentation](#)

[Registration](#)



[Search ticket](#)

[Submit ticket](#)

[Support staff](#)

Hello Catherine Biscarat
on prod-on.ggus.eu



Tickets

- ▶ [Submit a new ticket via browser](#)
- ▶ [Submit a new ticket via email](#)

Tickets from Catherine Biscarat (access via certificate)

ID	Status	Last Update	Info
----	--------	-------------	------

You don't have tickets in the system

- ▶ [Show my complete ticket list](#) (open/closed/subscribed)
- ▶ [Search ticket database](#)

Open tickets of all users

ID	VO	Info
▶ 101871	ops	[Rod Dashboard] Issues detected at IN2P3-LPSC
▶ 101870	ops	[Rod Dashboard] Issues detected at IN2P3-LPSC
▶ 101869	ops	[Rod Dashboard] Issues detected at IN2P3-LPSC
▶ 101868	other	CREAM_Job Errors on Bologna
▶ 101867	ops	[Rod Dashboard] Issues detected at IN2P3-IRES
▶ 101866	ops	[Rod Dashboard] Issues detected at IN2P3-SUBATECH
▶ 101865	ops	[Rod Dashboard] Issues detected at IN2P3-SUBATECH
▶ 101864	ops	RO-14-ITIM down apel gap
▶ 101863	ops	NAGIOS *emi.cream.glexec.WN-gLExec-/ops/Role=pilo...
▶ 101862	ops	NAGIOS *emi.cream.glexec.WN-gLExec-/ops/Role=pilo...
▶ 101861	ops	NAGIOS *emi.cream.CREAMCE-JobSubmit-ops* failed o...
▶ 101860	ops	NAGIOS *org.apel.APEL-Pub* failed on ce01.mosigri...
▶ 101859	ops	NAGIOS *org.sam.WN-Rep-ops* failed on grid03.spac...
▶ 101858	ops	NAGIOS *org.sam.SRM-GetURLs-ops* failed on grid0...
▶ 101857	dteam	update email address in operations-portal

- ▶ [Show all open tickets](#)

News

No news at the moment.

Info

GGUS tools/reports

- [GGUS ticket timeline tool - TTT](#)
- [Report Generator](#)
- [WLCG Reports](#)

GGUS development plans

- [Browse current open features](#)
- [Description of development procedures](#)
- [Ongoing worklist & Release Notes](#)
- [Submit a request for a new feature to GGUS](#)

GGUS Search

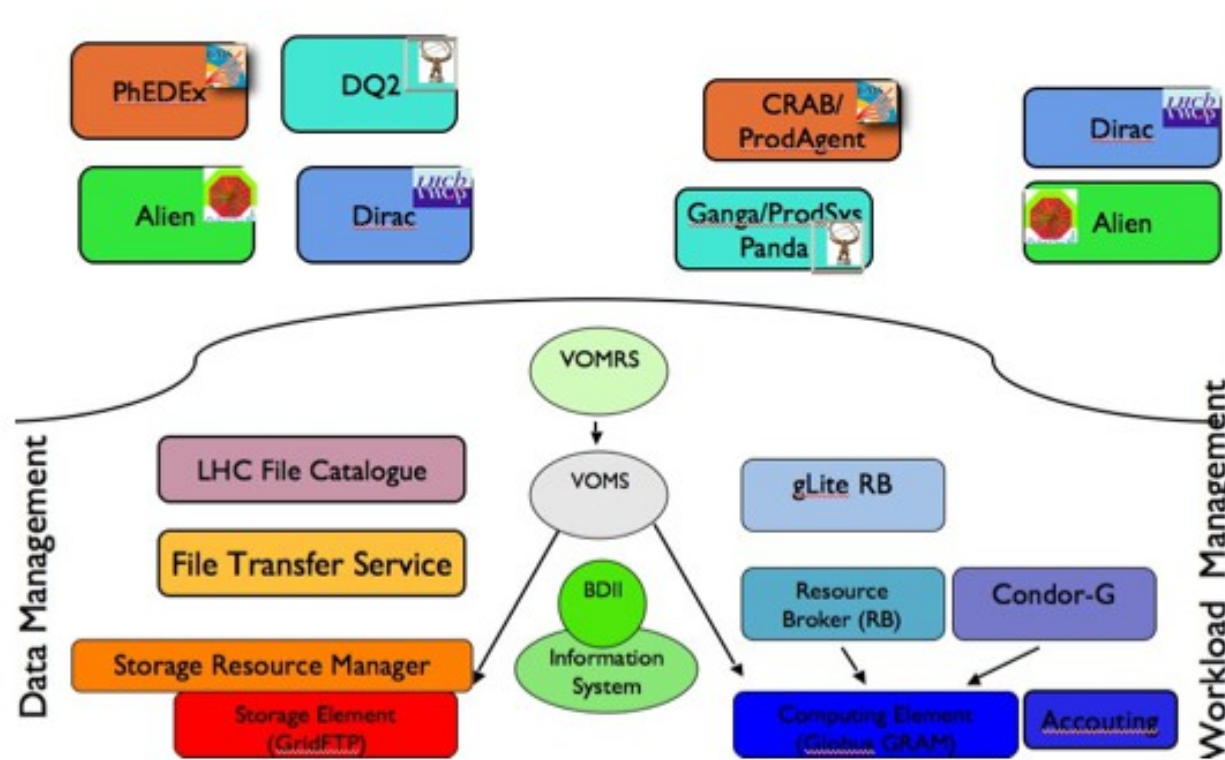
- [Documentation](#)
- [GGUS ticket search](#)
- [Special GGUS hints](#)

Ticket Search

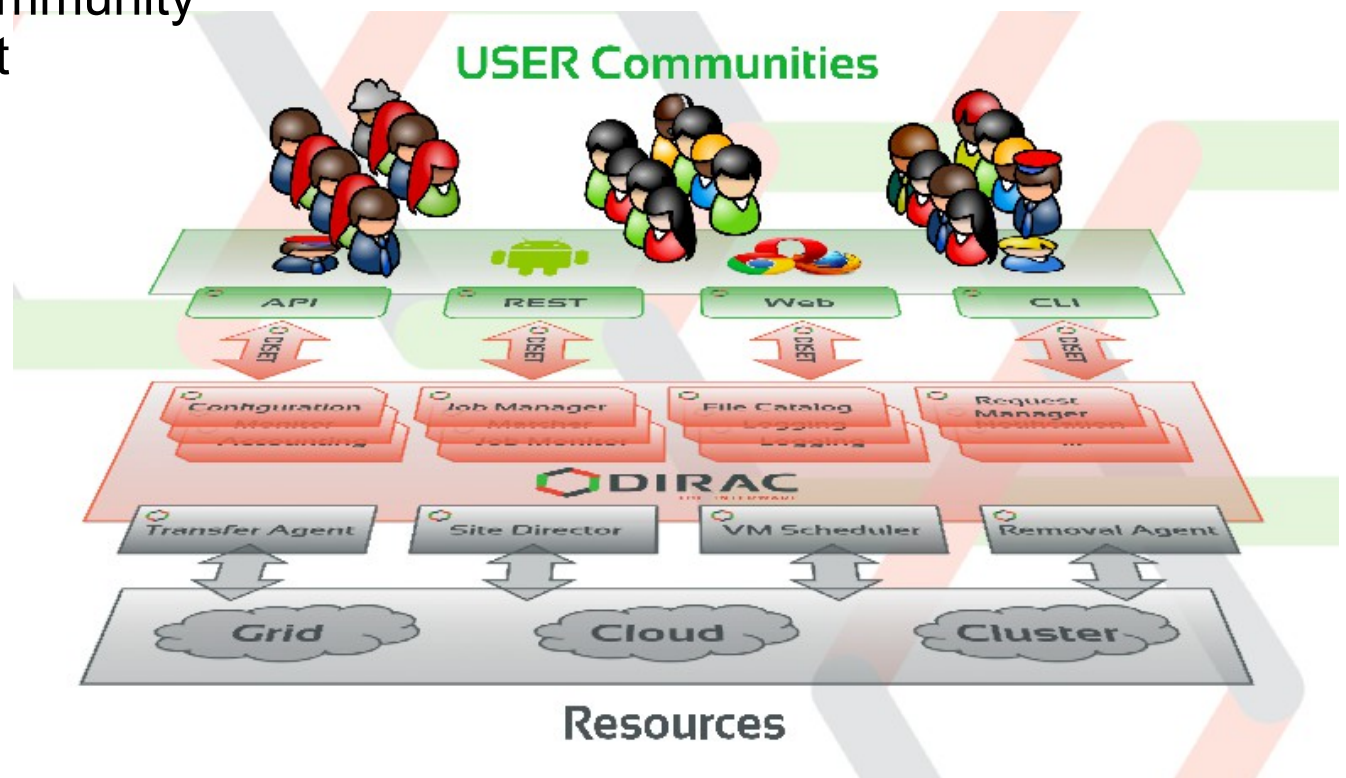
▶ Search ticket by ID:

Experiments own Middleware

- HEP experiments pioneered the massive use of grids
- CERN Director General Rolf Heuer about the Higgs discovery:
*"It was a global effort and it is a global success. The results today are only possible because of the extraordinary performance of the accelerators, including the infrastructure, the experiments, and the **Grid computing**."*
- Large VOs have developed their own Middleware
 - Easy to use interfaces
 - Better control



- Initially developed by the LHCb experiment (LHC)
- In 2009 decision to generalise the software
 - Separate specific LHCb functionalities
- 2013: DIRAC consortium
- DIRAC is now a general-purpose Middleware, offering services to any scientific community
 - Fermi-LAT, Glast
 - LSST
 - CTA
 - ...



Running jobs: 236092
Transfer rate: 11.41 GiB/sec

Today, Together

- Why grid computing

A success story : the grid for the LHC

- Other Grids

- Behind the scene

Technical details

- Going forward

Standards, simplicity, clouds

- Accessing the grid



ographer

/BKG

NGA, GEBCO

References

- International Conference on Computing in High Energy and Nuclear Physics (CHEP): <http://www.chep2013.org/>
- French summer school “Rencontres de physique de l'infiniment grand à l'infiniment petit”:
 - <https://indico.in2p3.fr/conferenceDisplay.py?confId=7293>
- French meeting “Les Journées réseaux” : <https://conf-ng.jres.org/2013>
- “Formation utilisateurs France-Grilles”: <https://indico.in2p3.fr/conferenceDisplay.py?confId=6453>
- International Symposium on Grids and Clouds (ISGC) 2013: <http://indico3.twgrid.org/indico/conferenceDisplay.py?confId=370>
- French Tutorial about the EGI usage: <https://indico.in2p3.fr/conferenceDisplay.py?confId=6453>
- CERN Summer Student Lecture 2011: <https://indico.cern.ch/event/134624/>