

The CERN Network

Openlab Summer 2012
CERN, 6th August 2012
edoardo.martelli@cern.ch

Summary



- IT-CS
- CERN networks
- LHC Data Challenge
- WLCG
- LHCOPN and LHCONE
- Openlab
- Conclusions

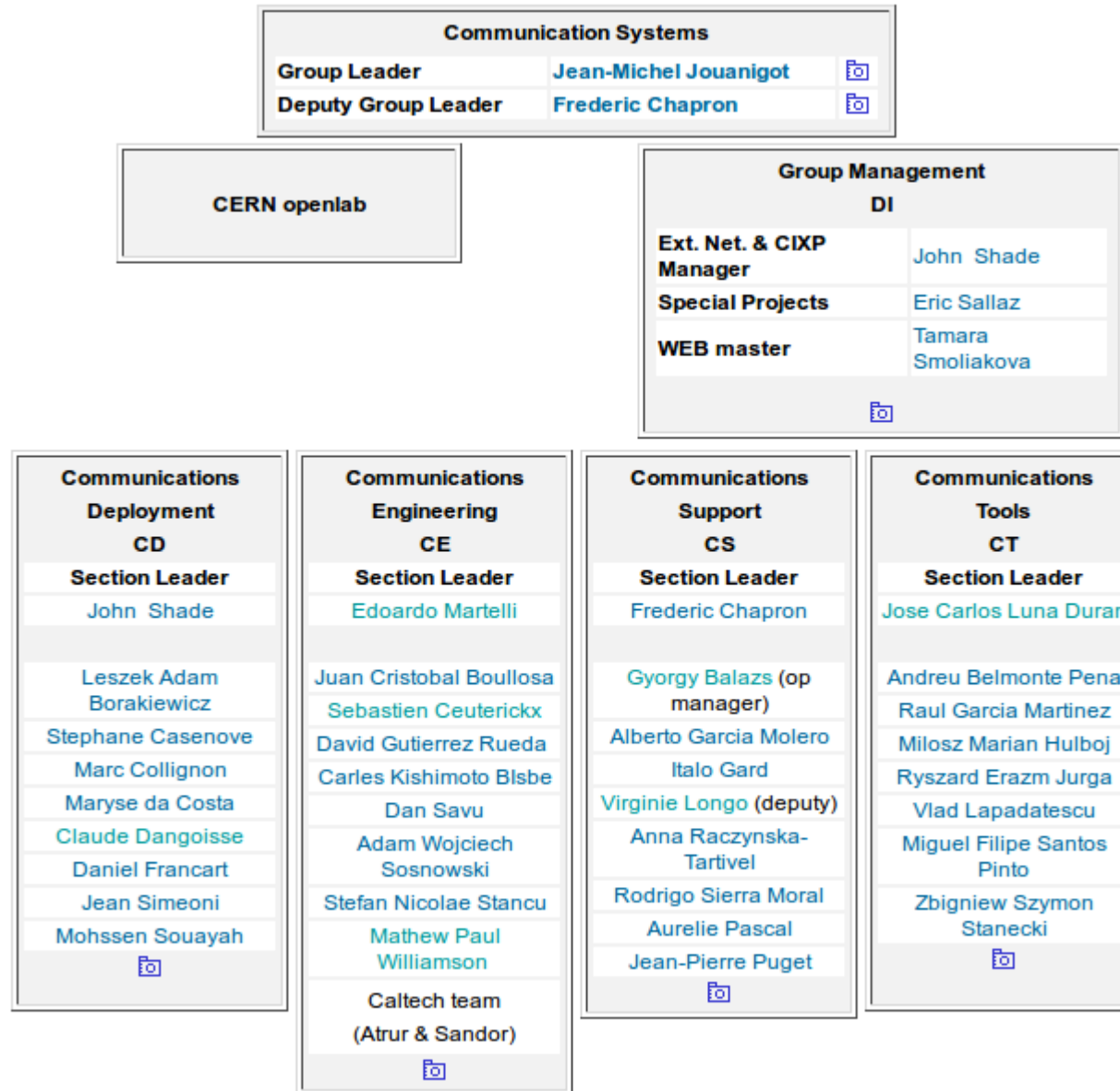
IT-CS

Communication systems

The IT-CS group is responsible for all communication services in use at CERN for data, voice and video

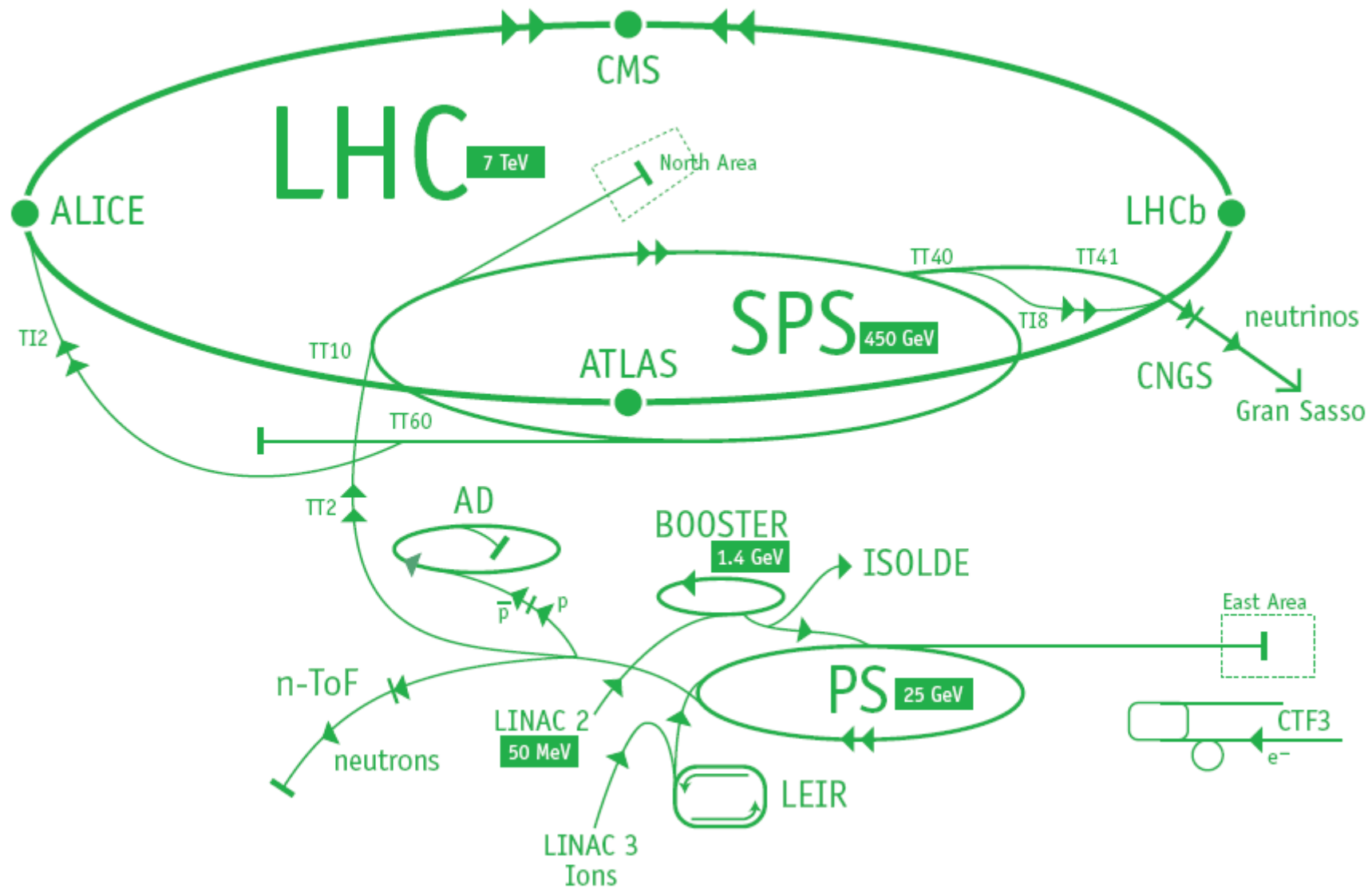
<http://it-cs.web.cern.ch/it-cs/>

IT-CS organization



Networks at CERN

CERN accelerator complex



High Energy Physics over IP

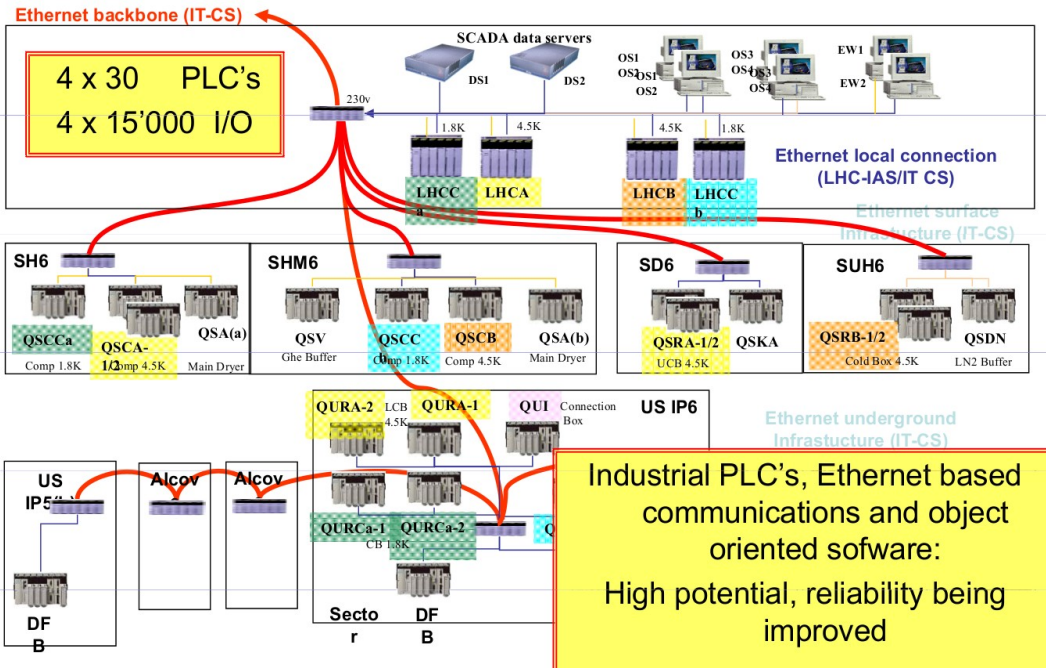
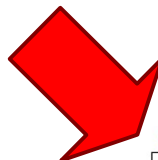
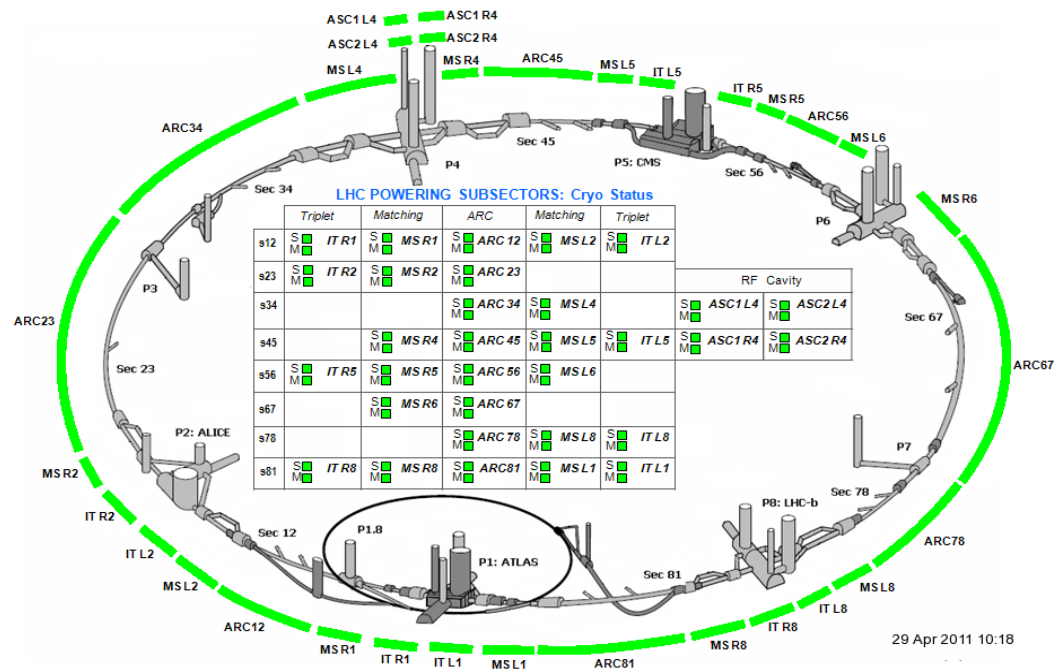


Most of the CERN infrastructure is controlled and managed over a pervasive **IP network**

Cryogenics



27Km of pipes at
-271.11° C by means of
700.000 litres of Helium:
controlled over IP

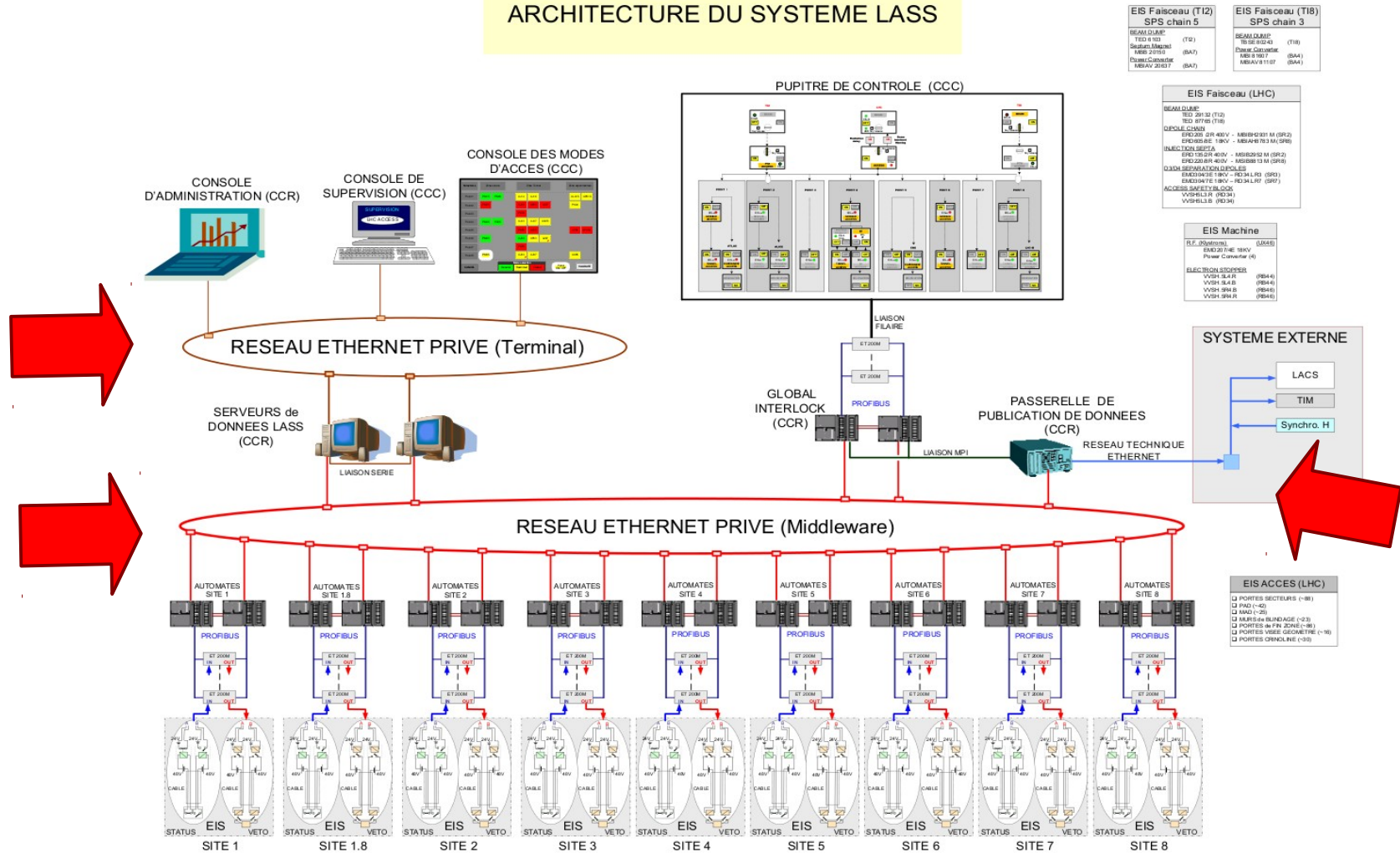


Industrial PLC's, Ethernet based communications and object oriented software:
High potential, reliability being improved

Access control

Safety and Security: made over IP

ARCHITECTURE DU SYSTEME LASS



Source: https://edms.cern.ch/file/931641/1/LASS-LACS_IHM.pdf

Remote inspections

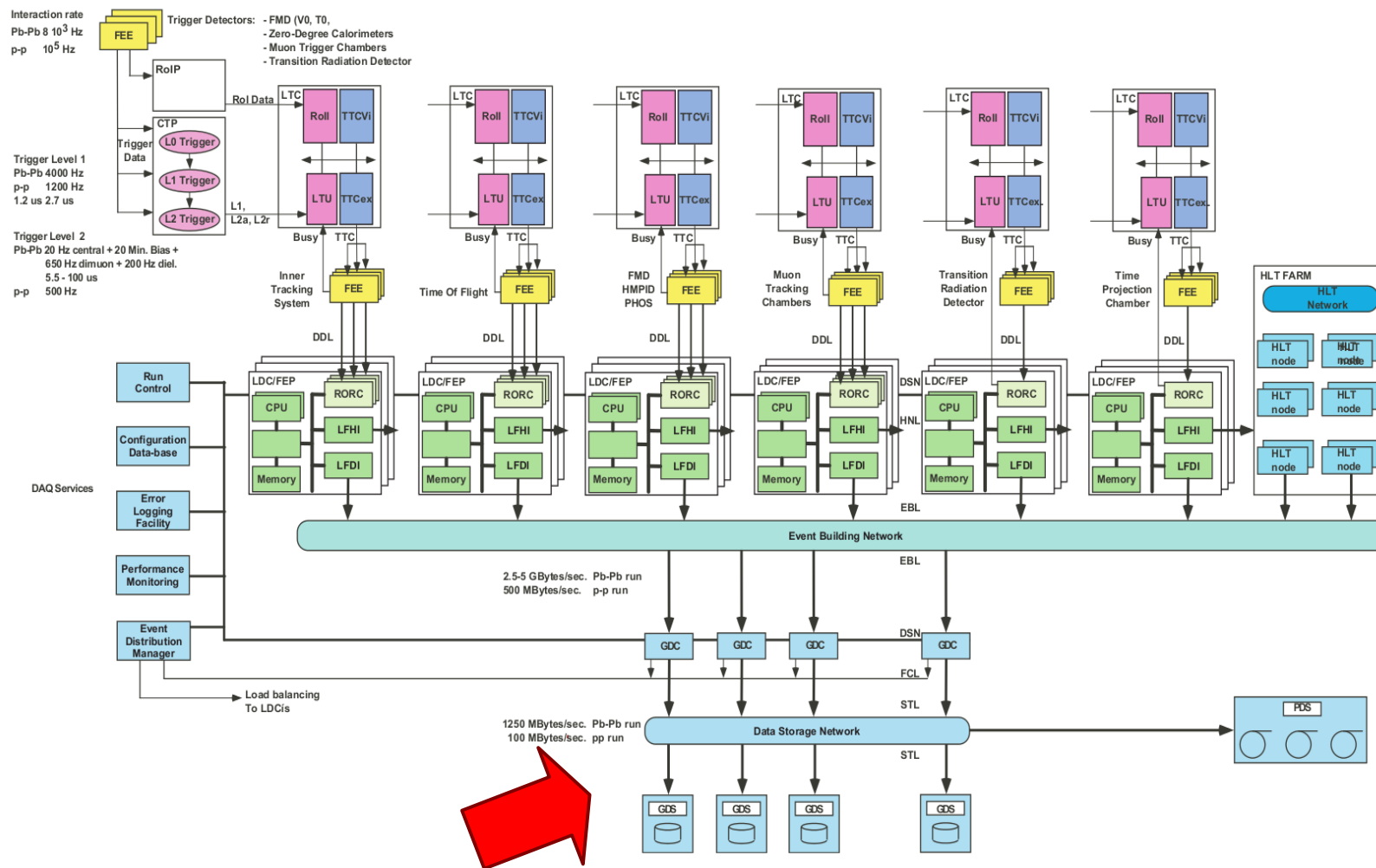


Remote inspection of dangerous areas: robots controlled and giving feedback over **WiFi and GSM IP networks**



DAQ: Data Acquisition

A constant stream of data from the four Detectors to disk storage



Source: http://aliceinfo.cern.ch/Public/Objects/Chapter2/DetectorComponents/daq_architecture.pdf

CCC: CERN Control Centre



The neuralgic centre of the particle accelerator: over IP

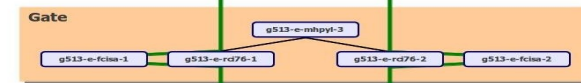
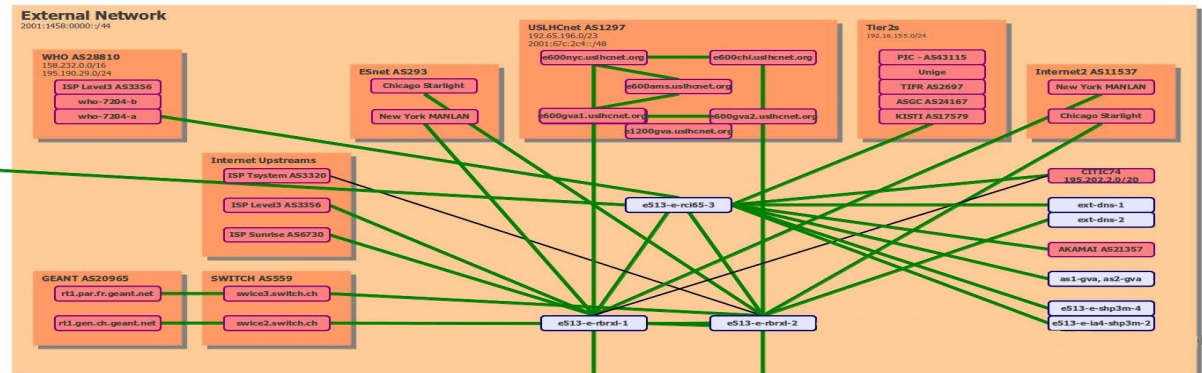
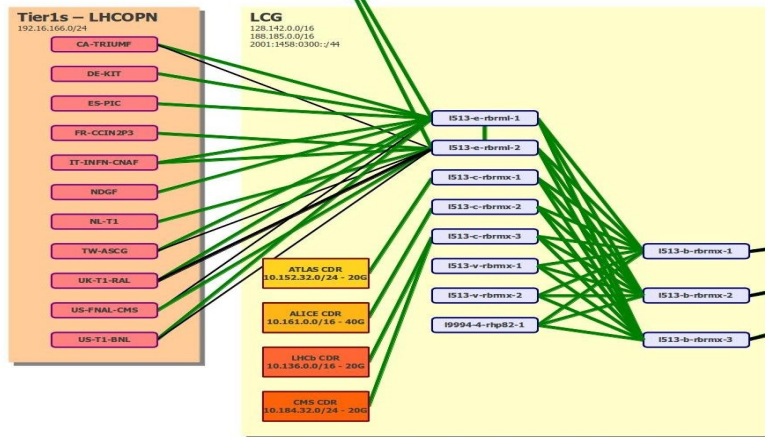
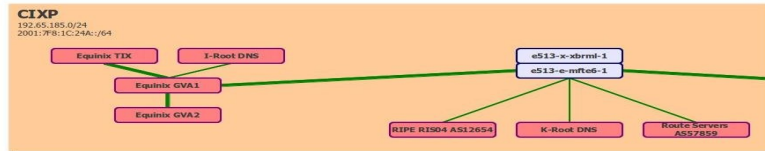


CERN data network

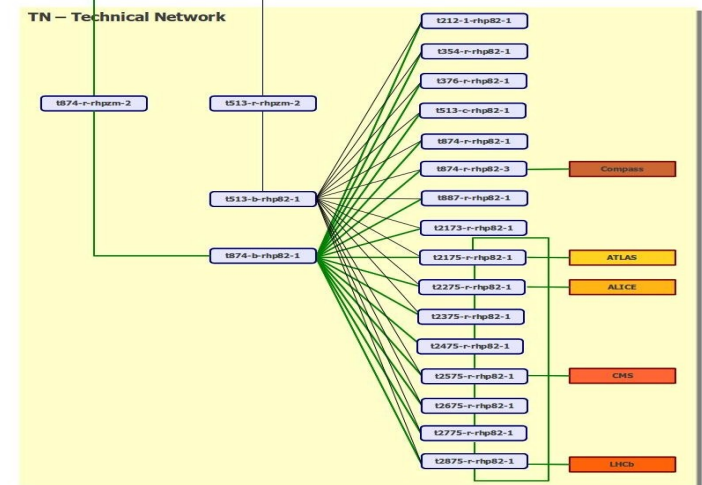
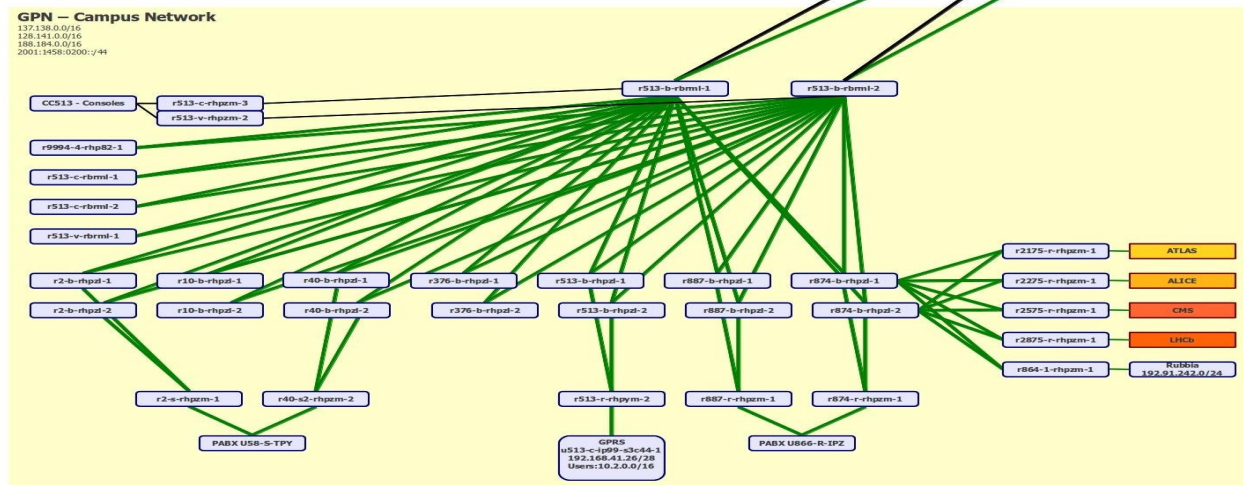


CERN – AS513

Last update: 20120730



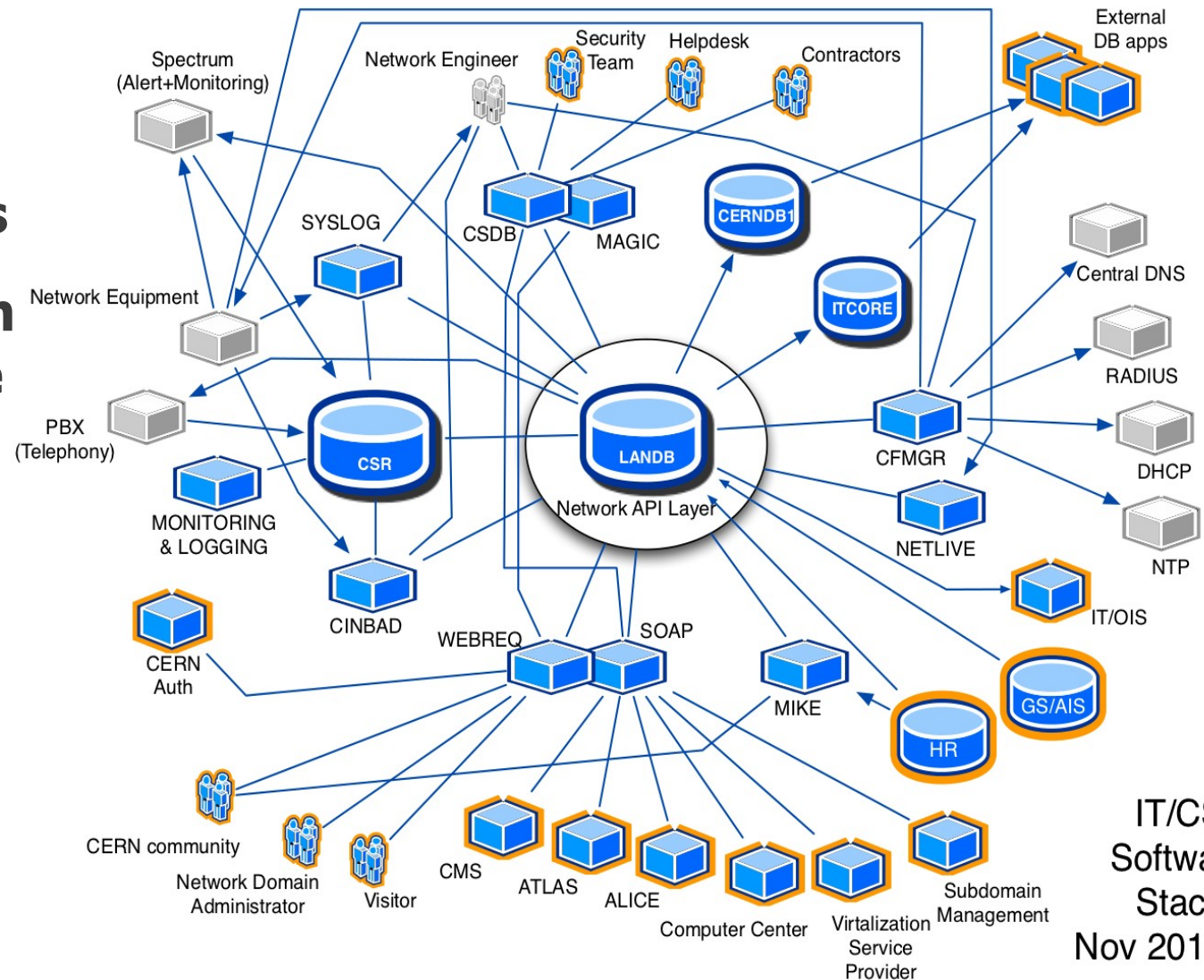
- 150 routers
- 2200 Switches
- 50000 connected devices
- 5000km of optical fibres



Network Provisioning and Management System



- 250 Database tables
- 100,000 Registered devices
- 50,000 hits/day on web user interface
- 1,000,000 lines of codes
- 11 years of development



IT/CS
Software
Stack
Nov 2010 V.3

Monitoring and Operations



The whole network is monitored and operated by the CERN NOC (Network Operation Centre)

The screenshot displays the Spectrum network management interface. On the left is a navigation tree with a table of network elements. The main area shows a topology diagram of the General Purpose Network (GPN) with various areas and components. Below the diagram is a 'Component Detail' panel for the GPN network, showing its status and configuration.

Name	14	13	81
My CA Spectrum	14	13	81
Global Collections (51)	10	10	78
Global Collection Hierarchy (2)			1
Configuration Manager (3)	6	10	71
eHealth Manager			
Service Performance Manager			
cs-srv-40 (0x2000000)	14	13	79
Policy Manager (7)			
Service Manager (3)	8	7	39
TopOrg			
Universe (9)	14	13	77
CNIC (4)			
CORE (2)			
cs-srv-40 (1)			
EXTNET (11)			3
FIREWALL (4)			
GPN (21)	12	11	60
Internet Upstreams (3)			
LCG (21)	2	2	14
cs-srv-42			
World			
Correlation Manager			
LostFound			
Remote Operations Manager			
Virtual Host Manager			
cs-srv-42 (0x3000000)			
cs-srv-44 (0x4000000)			2

General Purpose Network

513 Area Centre /613

CORE

Meyrin Area

External Network Temporary Exhibition

Preveessin / SPS / LHC Area

Component Detail: GPN of type Network

Information | Host Configuration | Root Cause | Interfaces | Performance | Neighbors | Alarms | Events | Attributes

Condition Normal

Subnet Address

Subnet Mask 0.0.0.0

Creation Time Jan 27, 2010 4:32:41 PM CET

Landscape cs-srv-40 (0x2000000)

Security String [set](#)

Child Count 21

Rollup Condition Critical

Value When Yellow 1 [set](#)

Value When Orange 3 [set](#)

Value When Red 7 [set](#)

Yellow Threshold 3 [set](#)

Orange Threshold 6 [set](#)

Red Threshold 10 [set](#)

IPv6 dual stack network deployment on going:
ready in 2013

Already available: dual-stack testbed

More information:

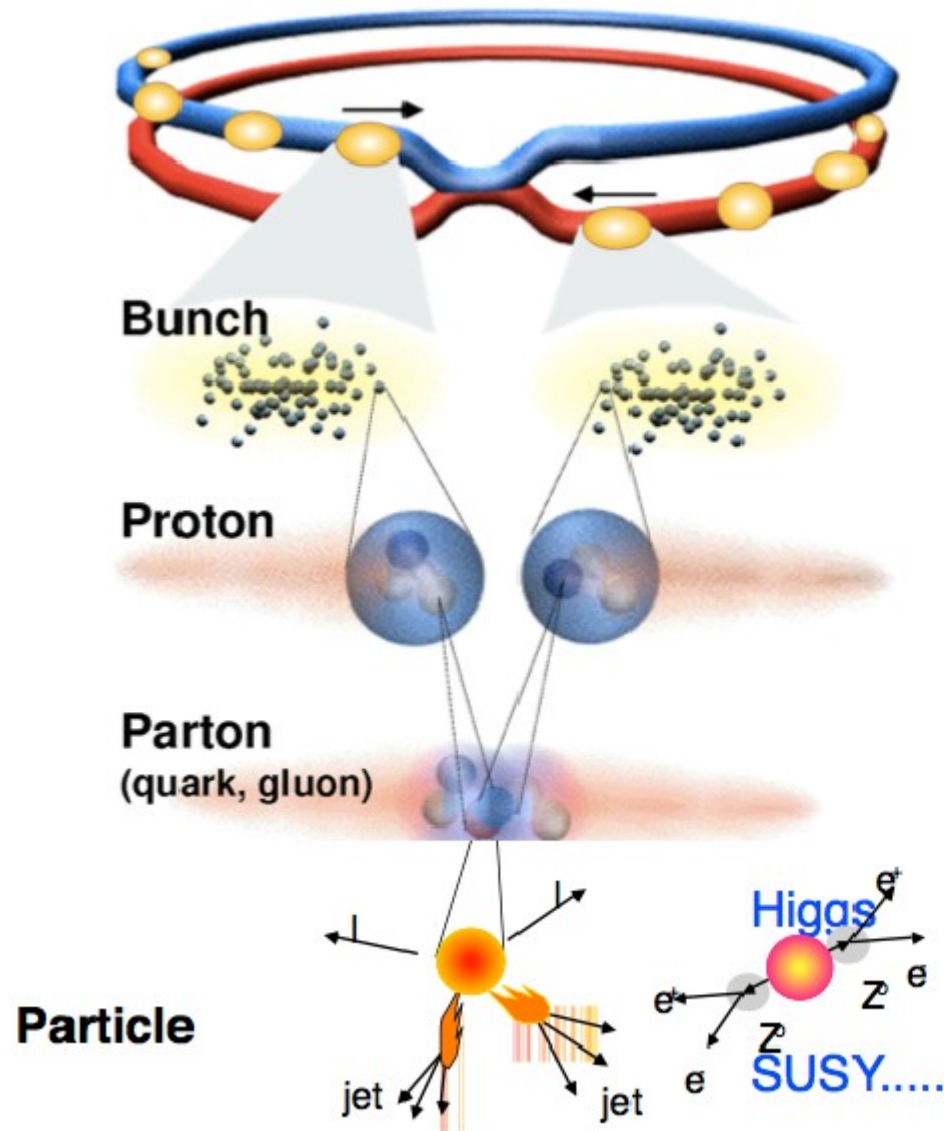
<http://cern.ch/ipv6>

almost



LHC Data Challenge

Collisions in the LHC



Proton - Proton 2808 bunch/beam
Protons/bunch 10^{11}
Beam energy 7 TeV (7×10^{12} eV)
Luminosity $10^{34} \text{cm}^{-2} \text{s}^{-1}$

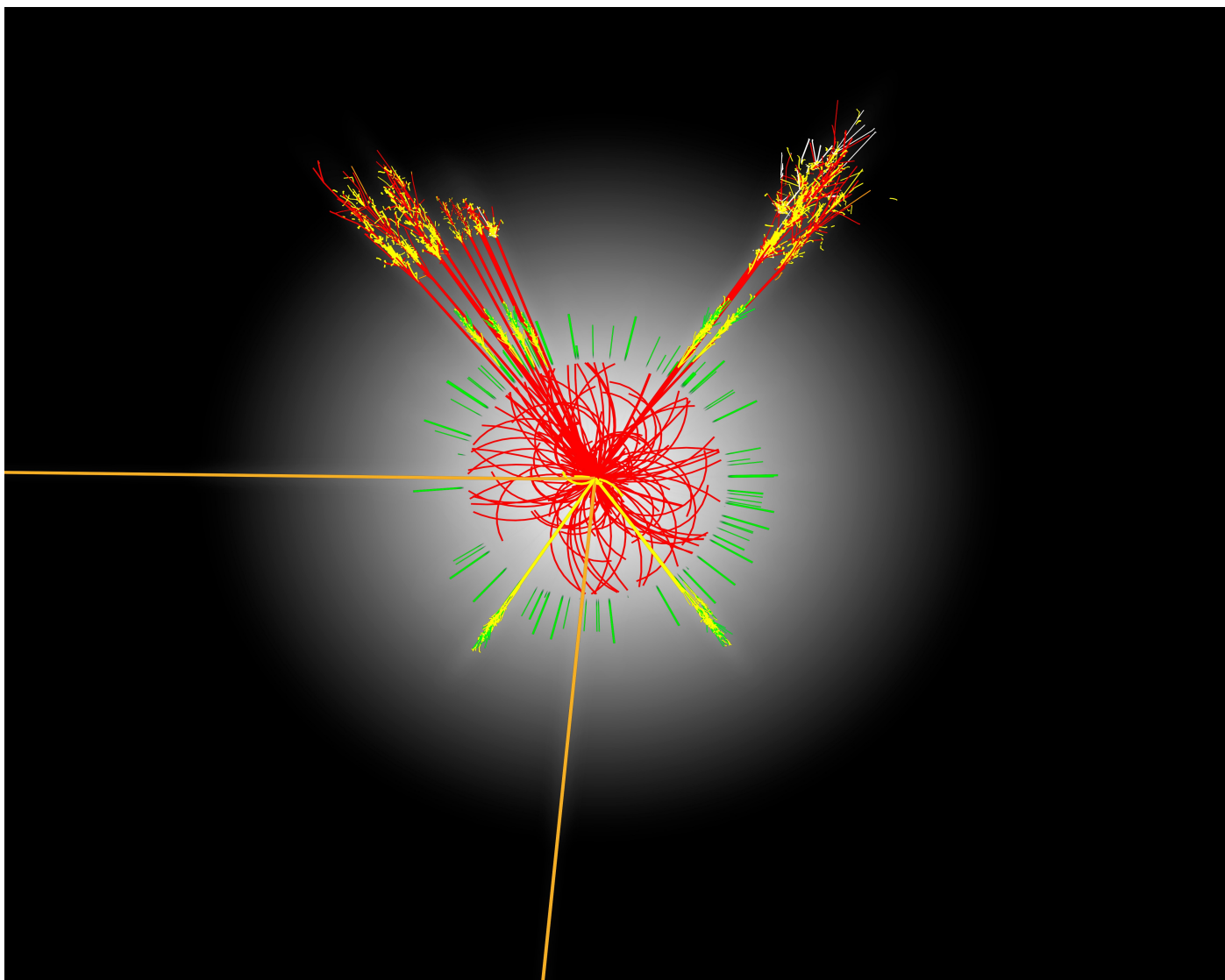
Crossing rate 40 MHz

Collision rate $\approx 10^7 - 10^9$

New physics rate $\approx .00001$ Hz

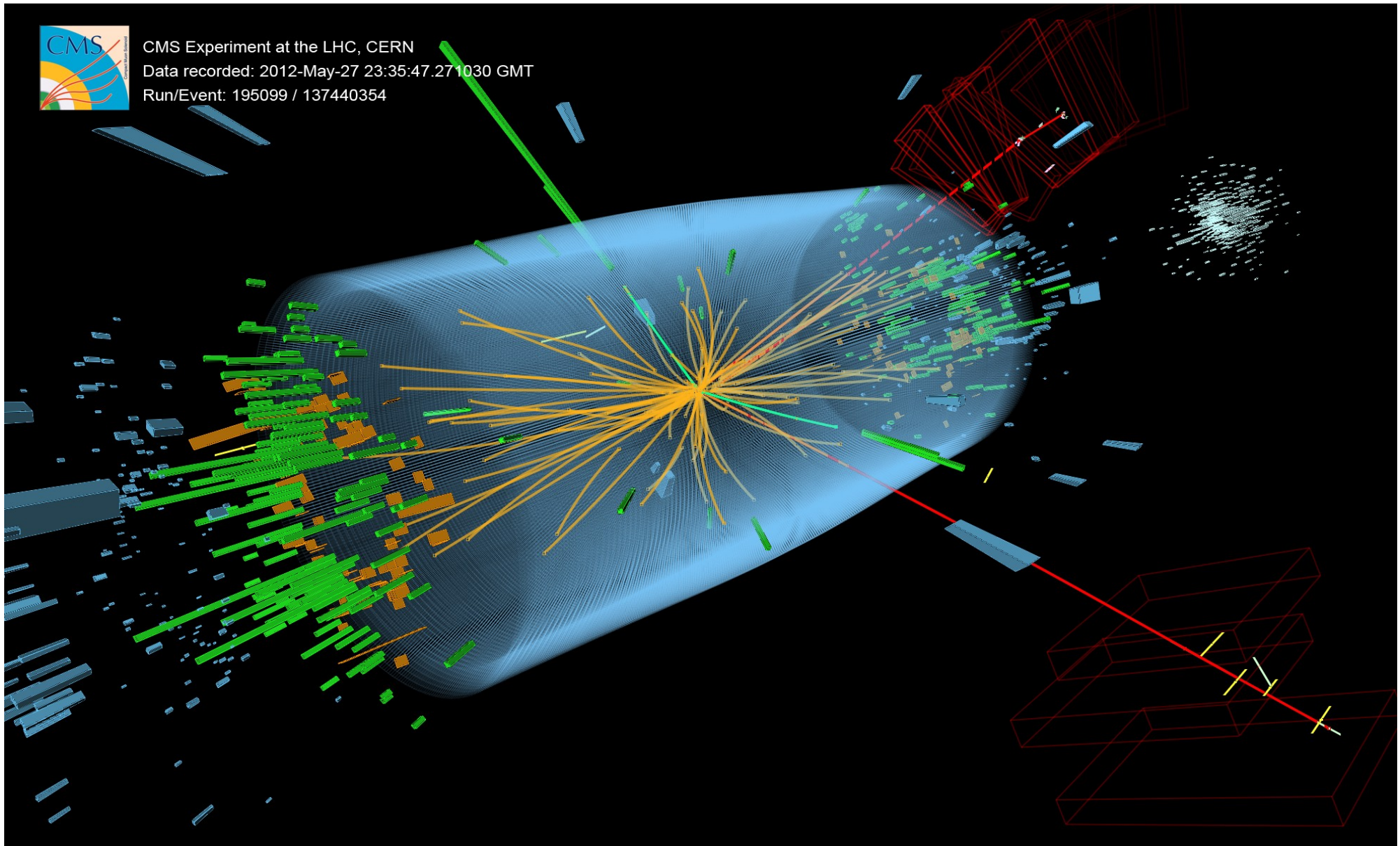
Event selection:
1 in 10,000,000,000,000

Comparing theory...



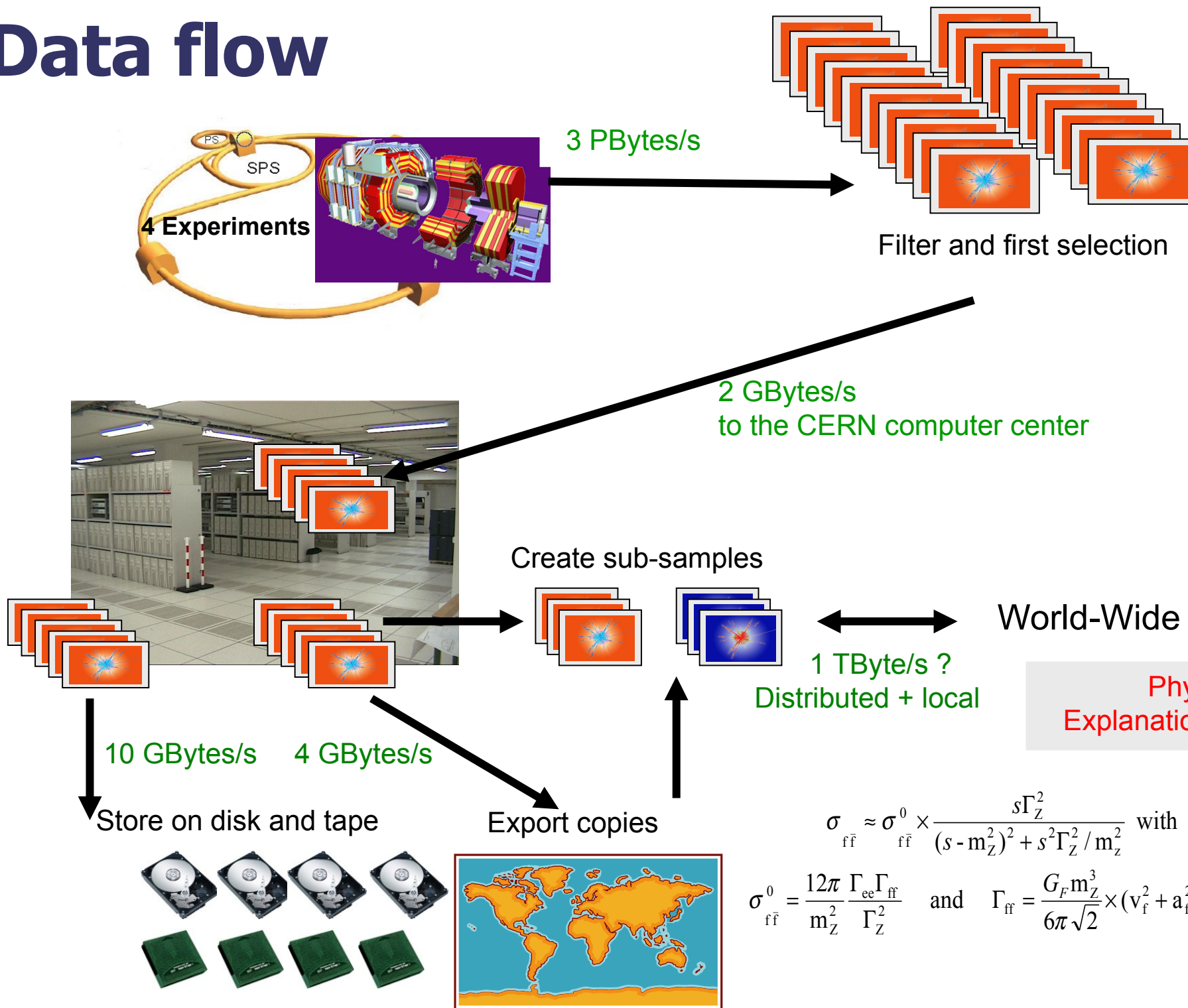
Simulated production of a Higgs event in ATLAS

.. to real events



Higgs event in CMS

Data flow



$$\sigma_{f\bar{f}} \approx \sigma_{f\bar{f}}^0 \times \frac{s\Gamma_Z^2}{(s-m_Z^2)^2 + s^2\Gamma_Z^2/m_Z^2} \text{ with}$$

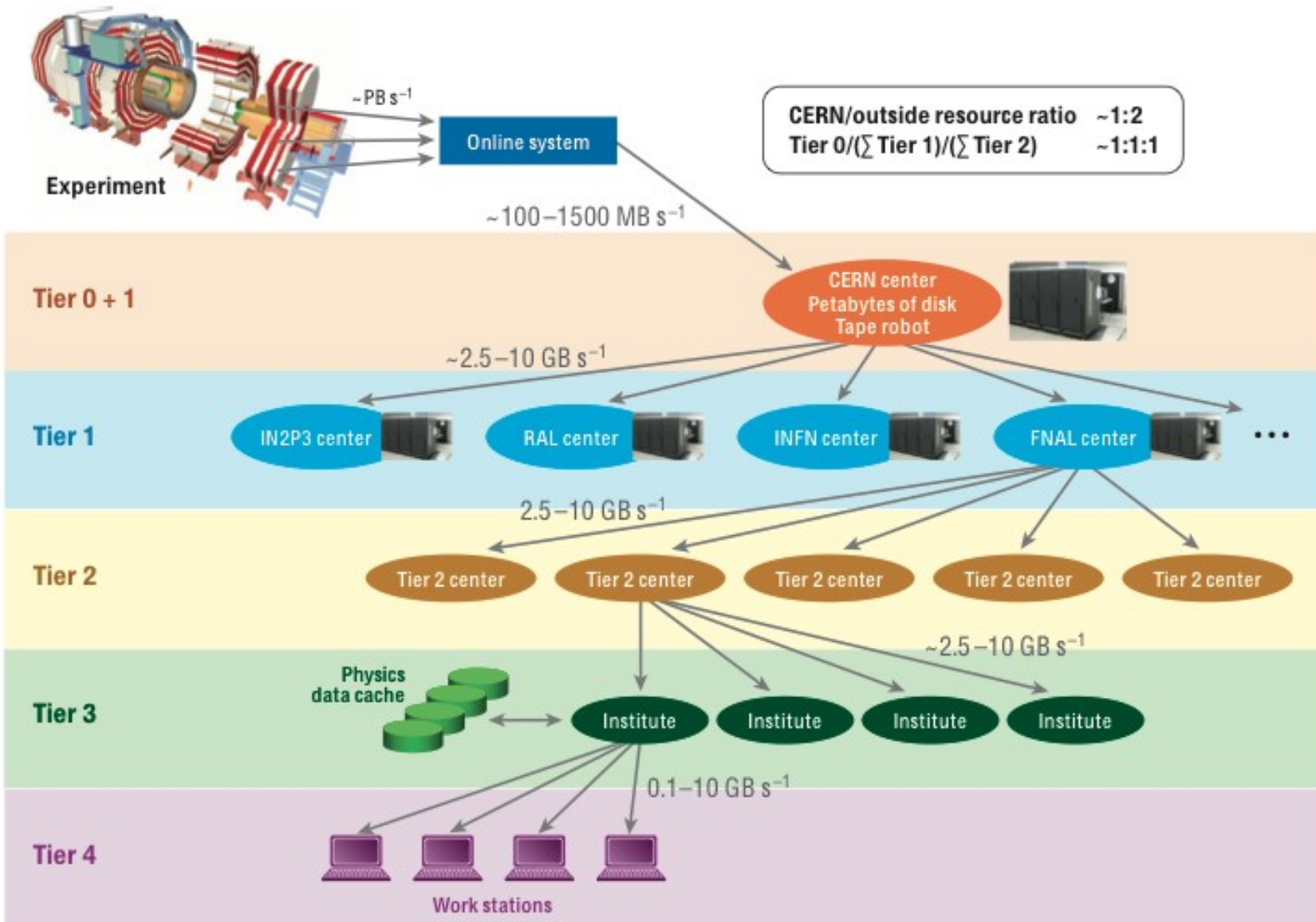
$$\sigma_{f\bar{f}}^0 = \frac{12\pi}{m_Z^2} \frac{\Gamma_{ee}\Gamma_{ff}}{\Gamma_Z^2} \text{ and } \Gamma_{ff} = \frac{G_F m_Z^3}{6\pi\sqrt{2}} \times (v_f^2 + a_f^2)$$

Data Challenge

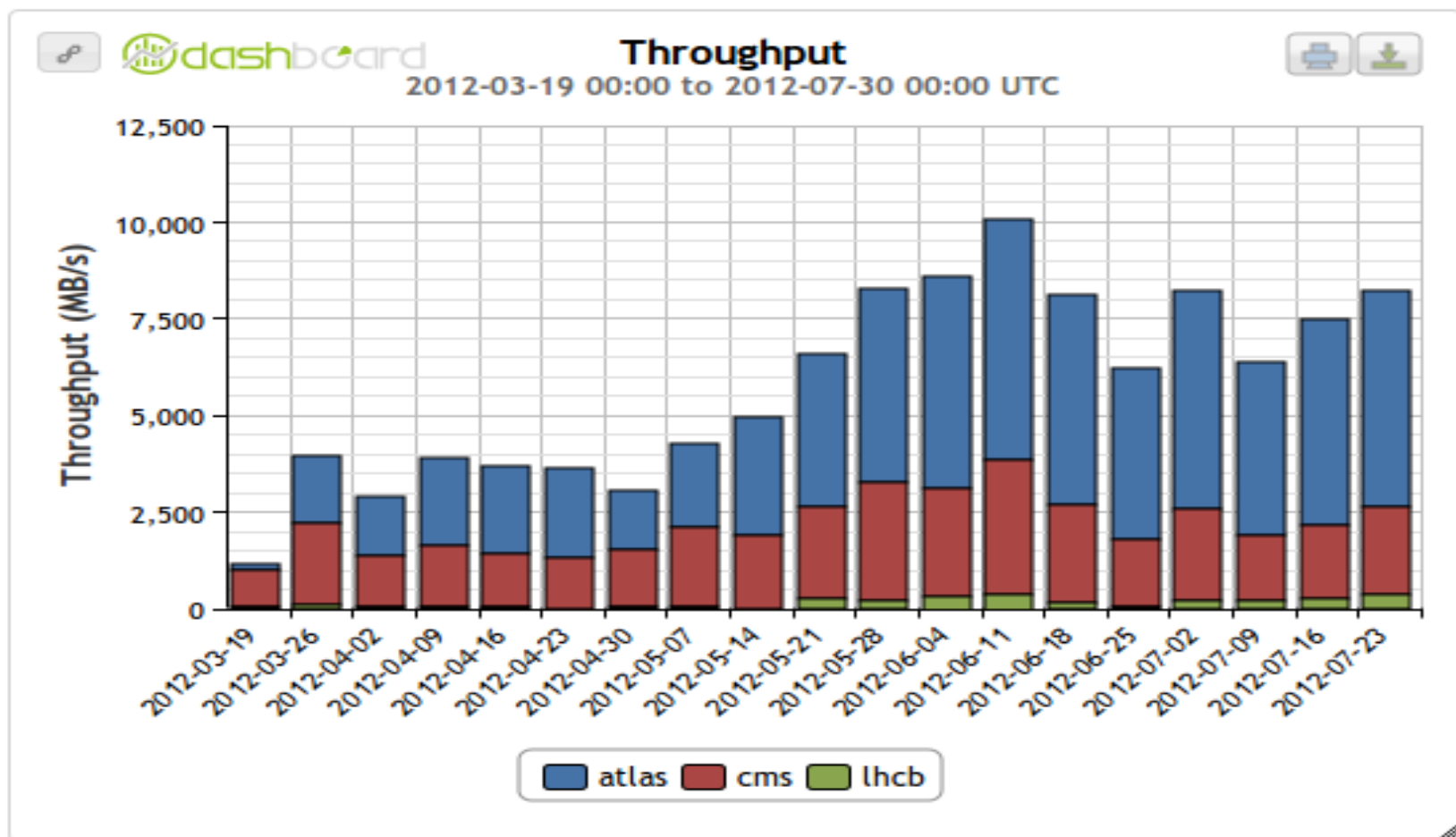


- **40 million collisions per second**
- **After filtering, 100 collisions of interest per second**
- **10^{10} collisions recorded each year**
= 15 Petabytes/year of data

Computing model

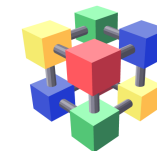


Last months data transfers



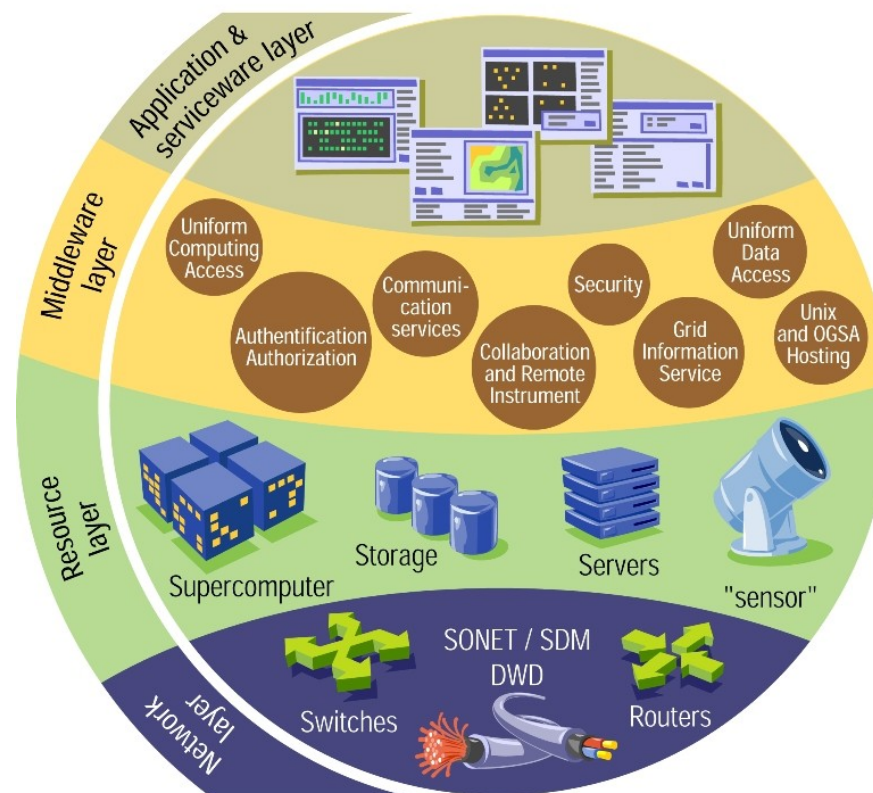
WLCG

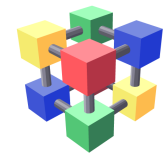
Worldwide LHC Computing Grid



Distributed Computing Infrastructure for LHC experiments

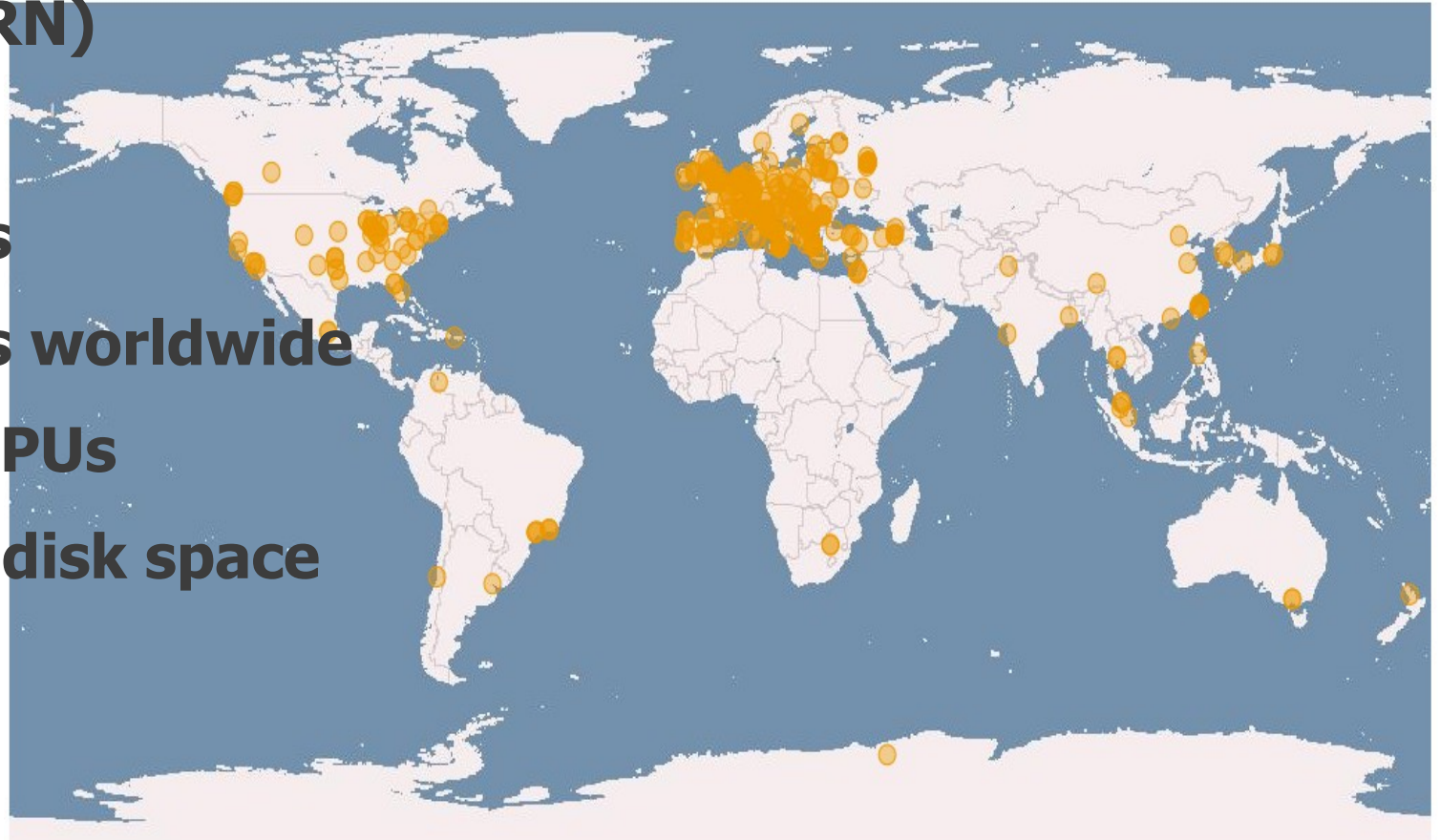
Collaborative effort of the HEP community





WLCG sites:

- 1 Tier0 (CERN)
- 11 Tier1s
- ~140 Tier2s
- >300 Tier3s worldwide
- ~250,000 CPUs
- ~ 150PB of disk space



CERN Tier0 resources



Servers	11000
Processors	15000
Cores	64000
HEPspec06	480000

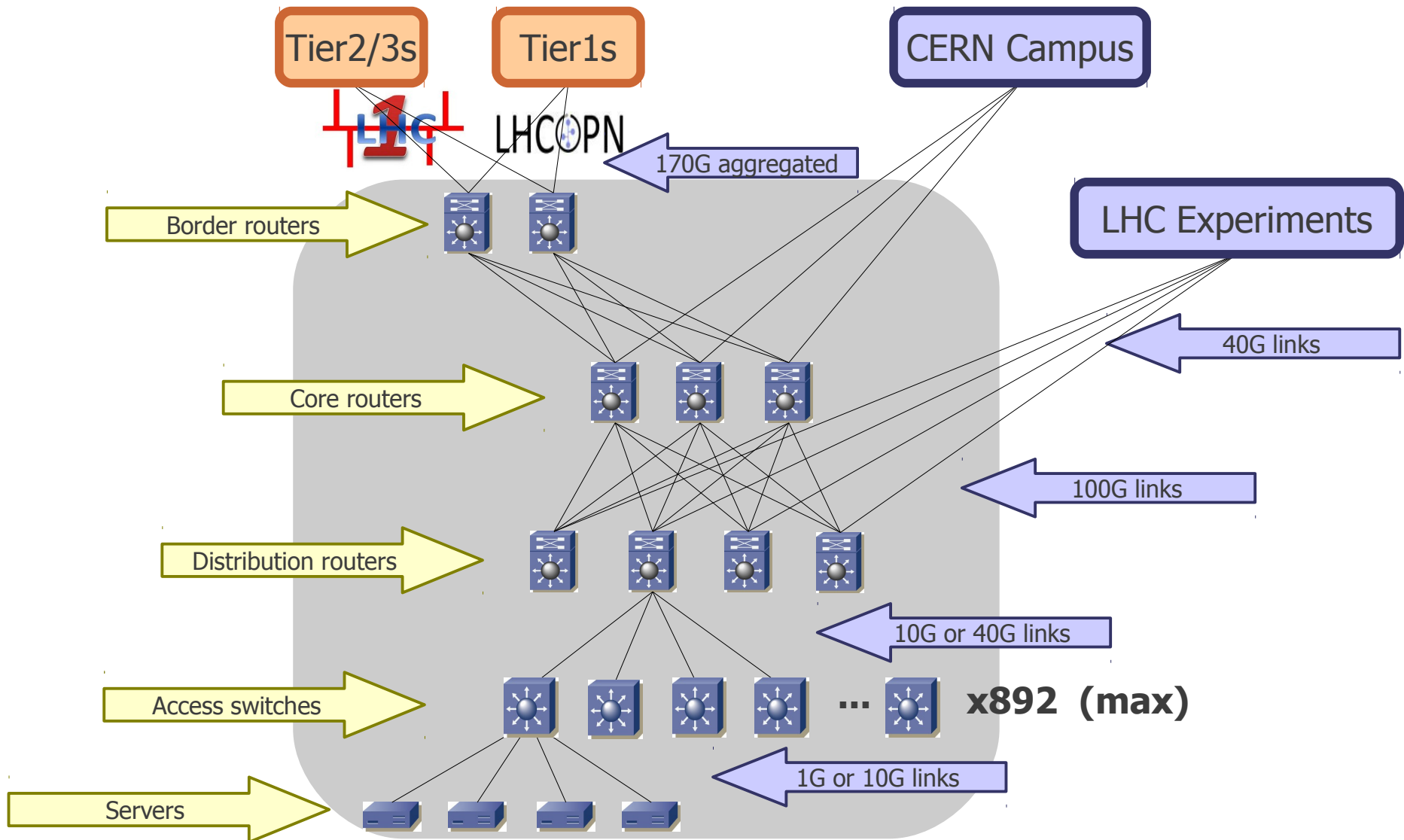
Disks	64000
Raw Disk Capacity (TB)	63000
Memory Modules	56000
RAID controllers	3750

Tape drives	160
Tape cartridges	45000
Tape slots	56000
Tape capacity(TB)	34000

High Speed routers	23
Ethernet switches	500
10Gbps ports	3000
100Gbps ports	48

March 2012

CERN Tier0 LCG network



Virtualization mobility (Software Defined Networks)

Commodity Servers with 10G NICs

High-end Servers with 40G NICs

40G and 100G interfaces on switches and routers

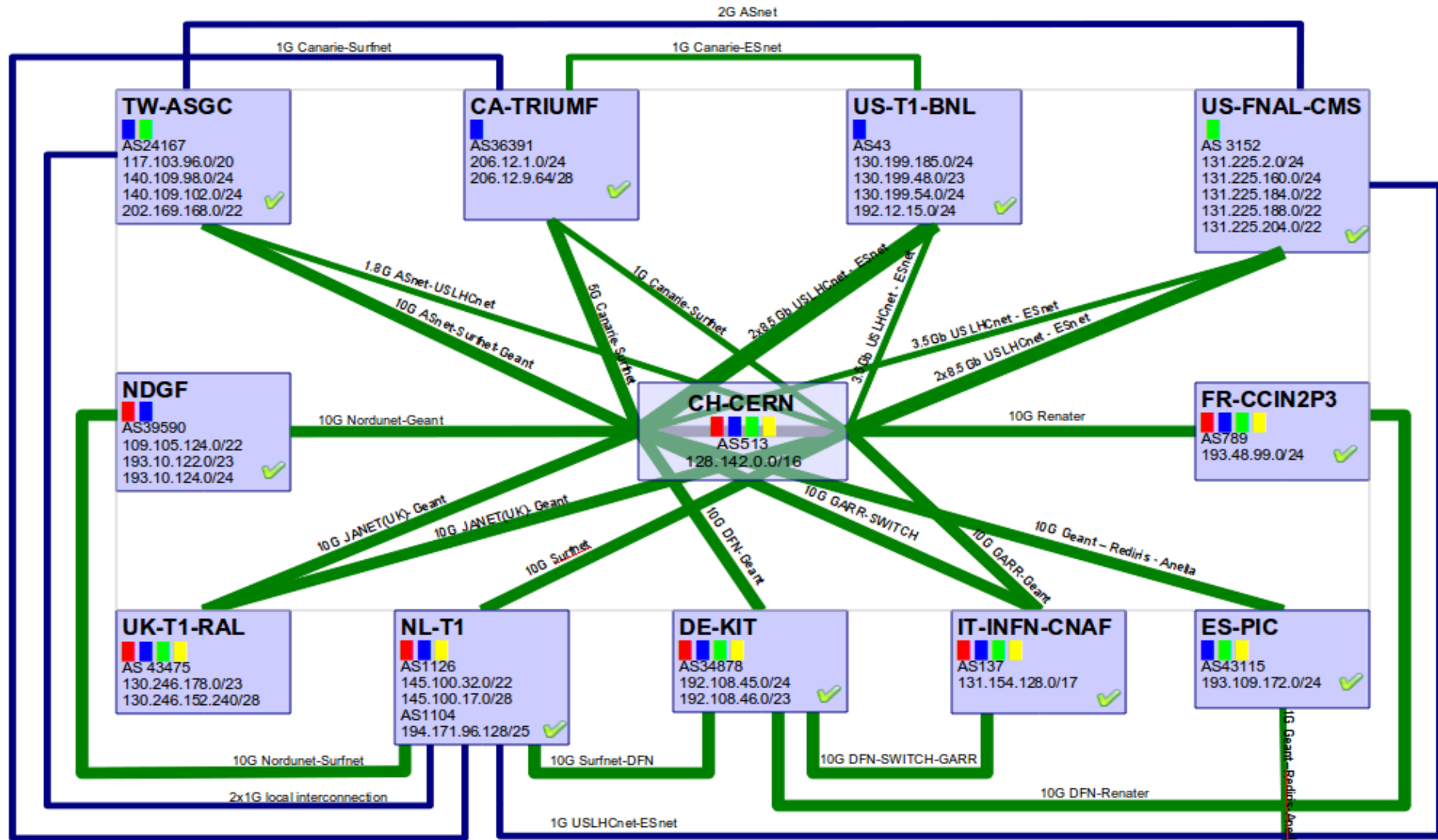
LHCOPN

LHC Optical Private Network

Tier0-Tier1s network



LHCOPN



- T0-T1 and T1-T1 traffic
- T1-T1 traffic only
- - - Not deployed yet
- (thick) >= 10Gbps
- (thin) <10Gbps
- = Alice
- = Atlas
- = CMS
- = LHCb
- ✓ = internet backup available
- p2p prefix: 192.16.166.0/24
- edoardo.martelli@cern.ch 20 110208



A collaborative effort



Designed, built and operated by the Tier0-Tier1s community

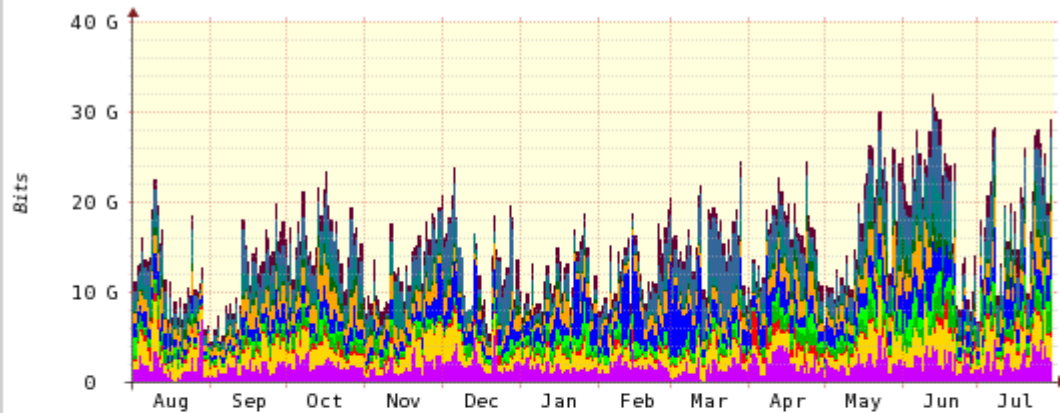
Links provided by the Research and Education network providers: Geant, USLHCnet, Esnet, Canarie, ASnet, Nordunet, Surfnet, GARR, Renater, JANET.UK, Rediris, DFN, SWITCH

- Single and bundled long distance 10G ethernet links
- Multiple redundant paths. Star+PartialMesh topology
- BGP routing: communities for traffic engineering, load balancing.
- Security: only declared IP prefixes can exchange traffic.

Traffic to the Tier1s



LHCOPN TOTAL Traffic (CERN -> Tiers1)

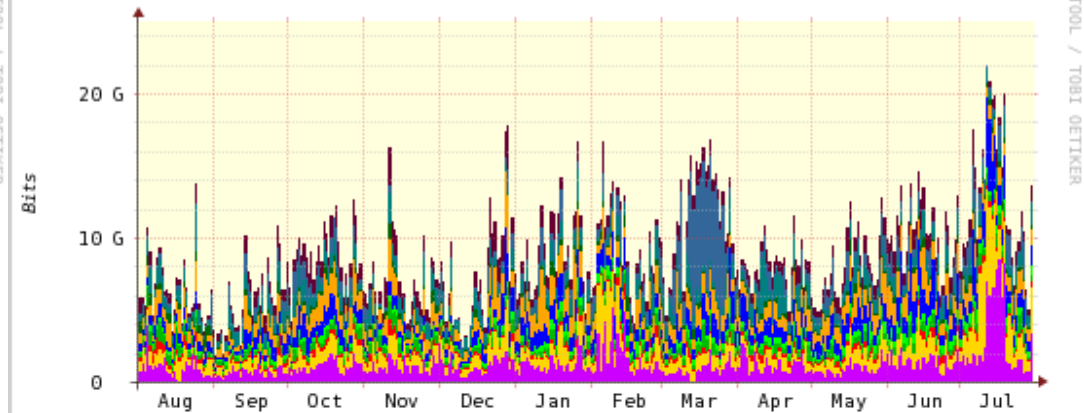


	Avg	Max	Peak	Curr
To DE-KIT	1.89G	6.93G	9.91G	2.17G
To IN2P3	1.89G	5.63G	9.55G	2.78G
To NDGF	423.62M	3.65G	9.72G	756.25 M
To NLT1	844.68M	3.94G	9.89G	3.88 G
To ASGC	727.67M	3.25G	7.89G	1.80G
To CNAF	2.16G	10.50G	12.77G	4.57G
To RAL	1.60G	6.61G	15.23G	3.04G
To TRIUMF	529.38M	2.95G	5.19G	1.81G
To BNL	1.57G	6.64G	13.68G	3.80G
To FNAL	2.08G	8.92G	15.92G	2.59G
To ES-PIC	1.24G	4.48G	9.72G	1.79G

Total to Tiers1 Avg: 14.91G Max: 31.97G Curr: 28.97G
 Last update: Tue Jul 31 2012 11:32:17

CERN 2012

LHCOPN TOTAL Traffic Flow (Tiers1 -> CERN)



	Avg	Max	Peak	Curr
From DE-KIT	1.26G	8.49G	9.90G	3.32G
From IN2P3	1.12G	6.83G	9.88G	3.31G
From NDGF	282.01M	1.51G	9.73G	310.85M
From NLT1	466.06M	1.64G	9.00G	447.82M
From ASGC	369.67M	1.54G	8.50G	638.11M
From CNAF	1.06G	5.10G	10.63G	1.52G
From RAL	1.21G	5.85G	10.25G	1.09G
From TRIUMF	265.91M	1.47G	2.79G	190.80M
From BNL	935.08M	3.74G	11.13G	1.54G
From FNAL	922.11M	7.58G	11.13G	194.24M
From ES-PIC	805.33M	3.02G	9.42G	1.09G

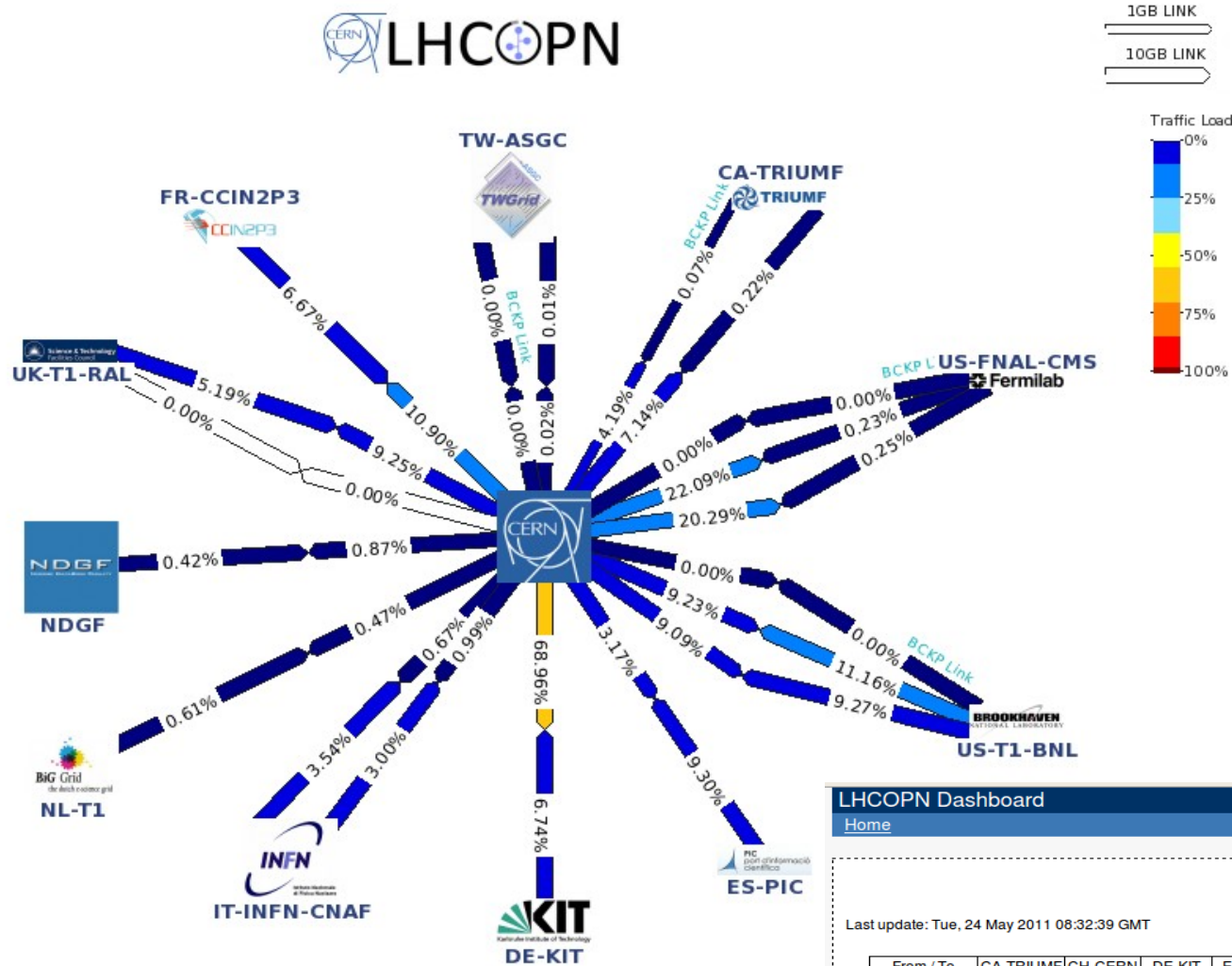
Total from Tiers1 Avg: 8.68G Max: 21.99G Curr: 13.65G
 Last update: Tue Jul 31 2012 11:32:18

CERN 2012

Monitoring



LHCOPN



Created: Apr 28 2011 13:10:10

LHCOPN Dashboard

Home

Current status

Last update: Tue, 24 May 2011 08:32:39 GMT

From / To	CA-TRIUMF	CH-CERN	DE-KIT	ES-PIC	FR-CCIN2P3	IT-INFN-CNAF	NDGF	NL-T1	TW-ASGC	UK-T1-RAL	US-FNAL-CMS	US-T1-BNL
CA-TRIUMF	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
CH-CERN	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
DE-KIT	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
ES-PIC	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
FR-CCIN2P3	OK	OK	OK	Critical	OK	OK	OK	OK	OK	OK	OK	OK
IT-INFN-CNAF	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
NDGF	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
NL-T1	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
TW-ASGC	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
UK-T1-RAL	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
US-FNAL-CMS	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
US-T1-BNL	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

JSON feed

Legend: OK Deviation from Baseline Critical Unknown

LHCONE

LHC Open Network Environment

Driving the change



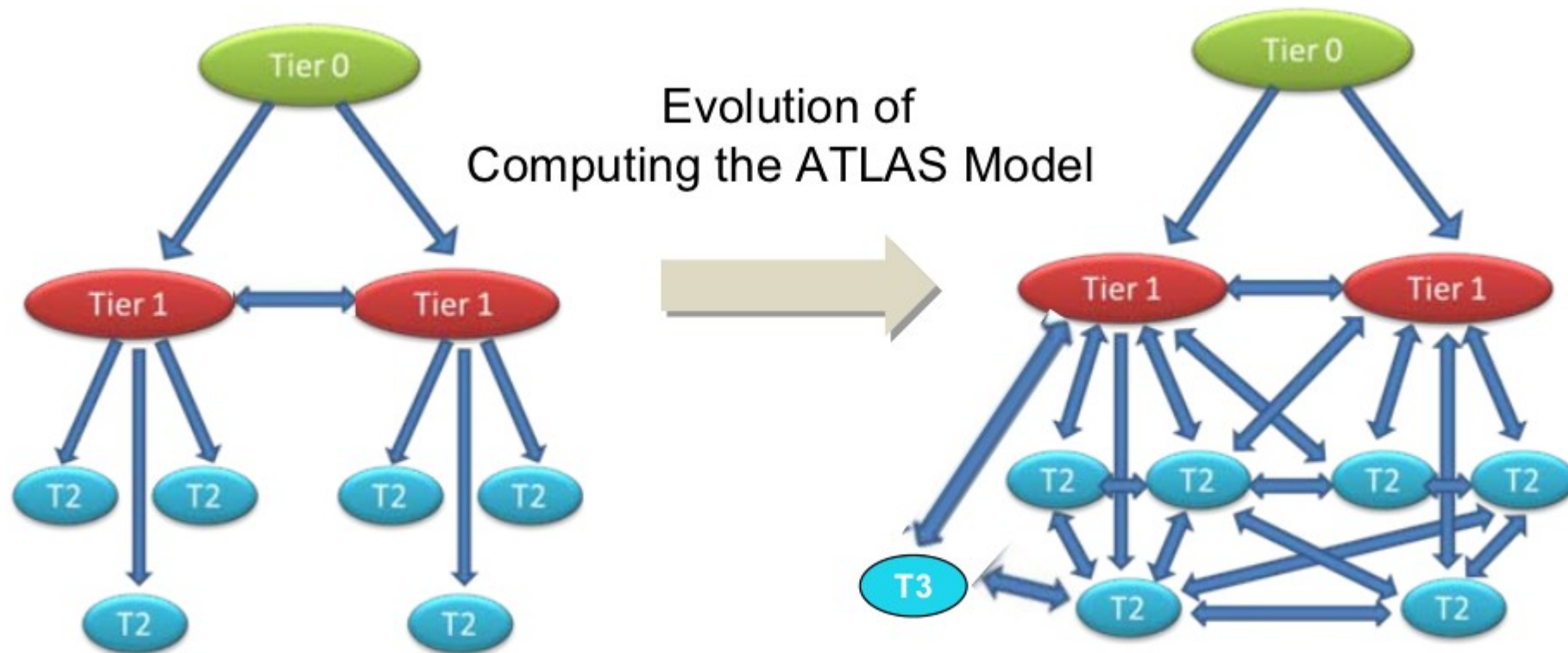
“The Network infrastructure is the most reliable service we have”

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

“Data placement will be driven by demand for analysis and not pre-placement”

Ian Bird, WLCG project leader

Change of computing model (ATLAS)



New computing model



- Better and more dynamic use of storage
- Reduce the load on the Tier1s for data serving
- Increase the speed to populate analysis facilities

Needs for a faster, predictable, pervasive network connecting Tier1s and Tier2s

Requirements from the Experiments



- Connecting any pair of sites, regardless of the continent they reside
- Bandwidth ranging from 1Gbps (Minimal), 5Gbps (Nominal), 10G and above (Leadership)
- Scalability: sites are expected to grow
- Flexibility: sites may join and leave at any time
- Predictable cost: well defined cost, and not too high

Needs for a better network



- more bandwidth by federating (existing) resources
- sharing cost of expensive resources
- accessible to any TierX site

=



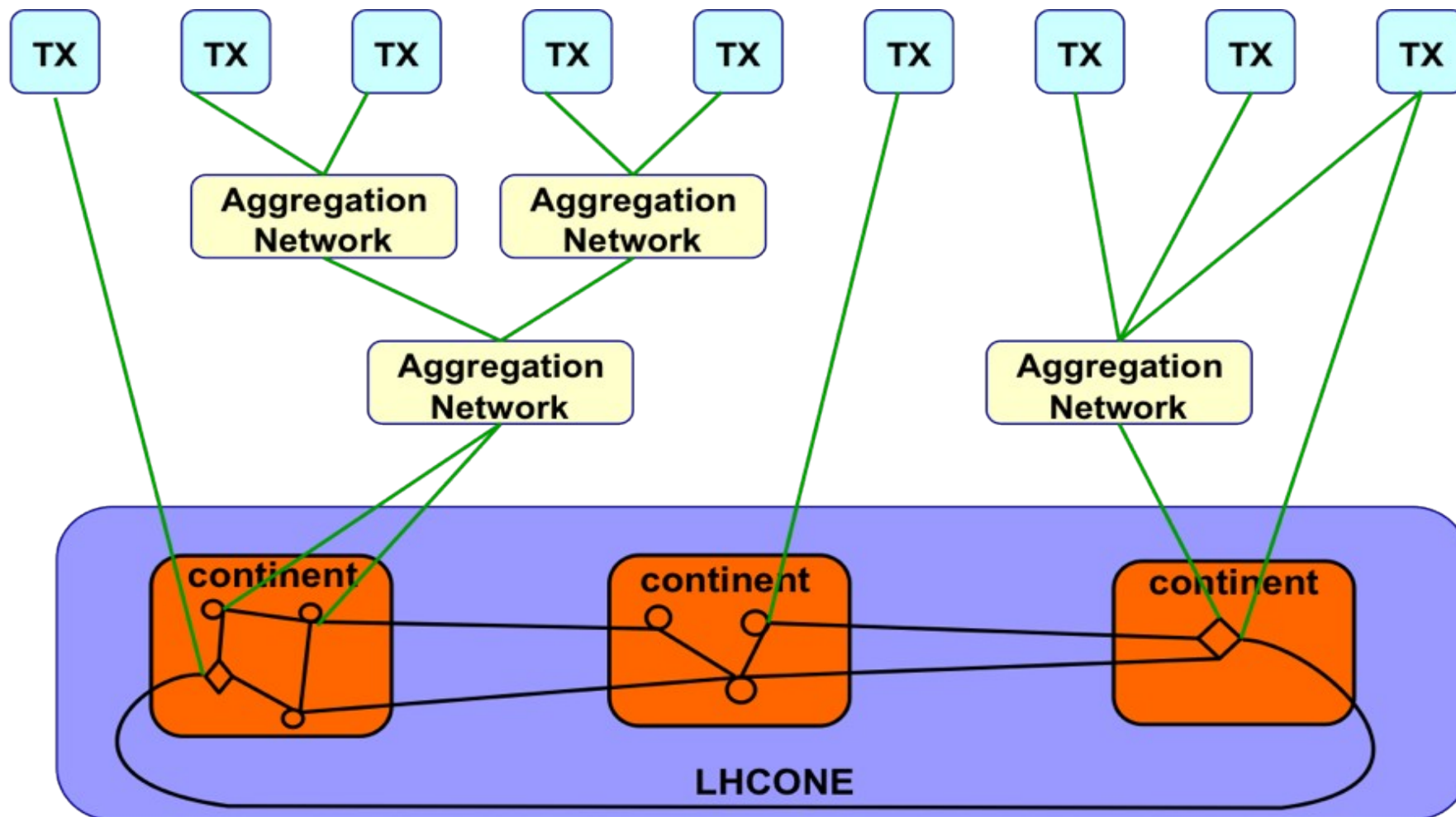
LHC Open Network Environment

LHCONE concepts



- Serves any LHC sites according to their needs and allowing them to grow
- A collaborative effort among Research & Education Network Providers
- Based on Open Exchange Points: easy to join, neutral
- Multiple services: one cannot fit all
- Traffic separation: no clash with other data transfer, resource allocated for and funded by HEP community

LHCONE architecture



- ◇ distributed exchange point
- single node exchange point

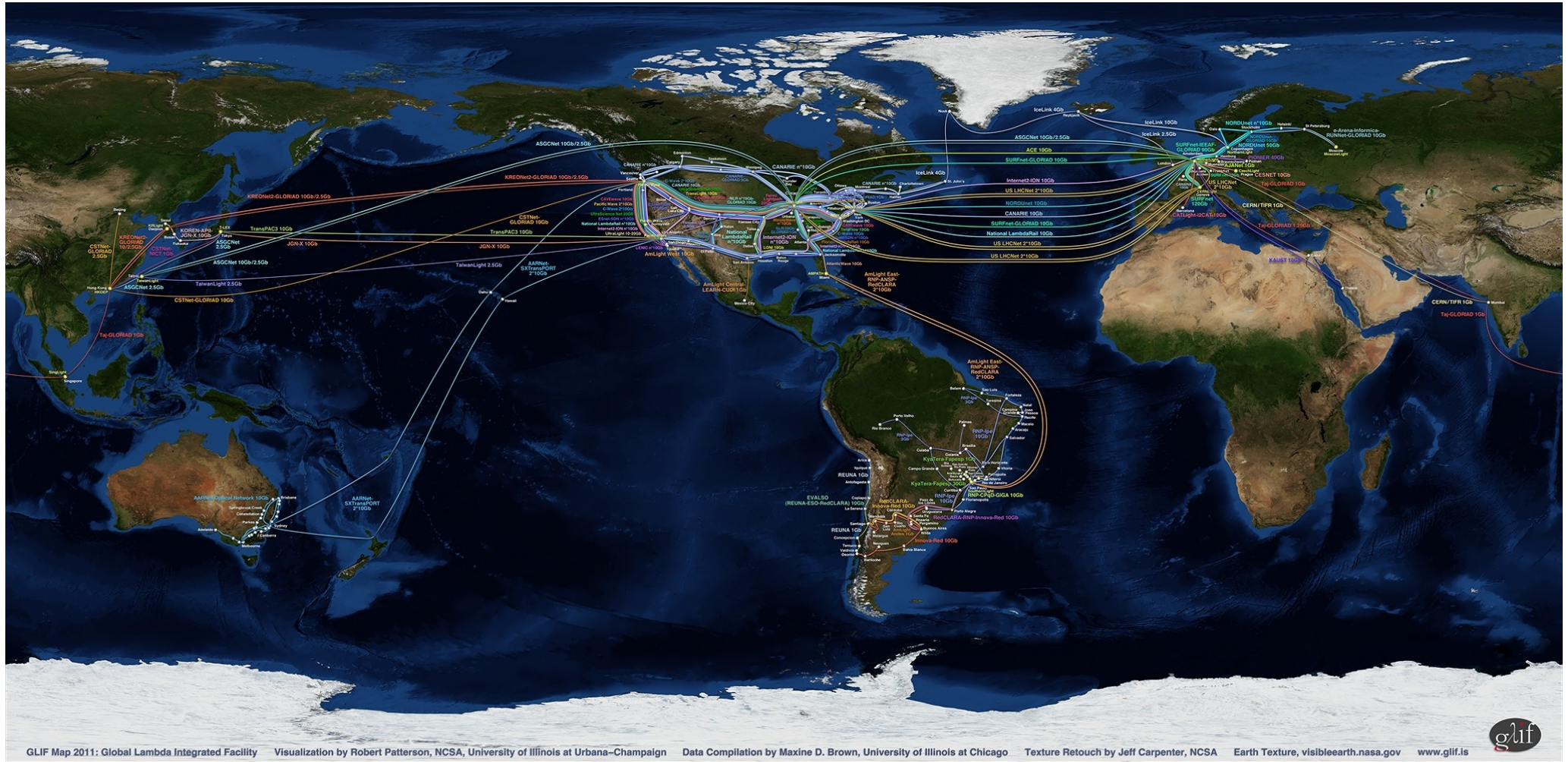
LHCONE building blocks



- **Single node Exchange Points**
- **Continental/regional Distributed Exchange Points**
- **Interconnect circuits between Exchange Points**

These exchange points and the links in between collectively provide LHCONE services and operate under a common LHCONE policy

The underlying infrastructure



GLIF Map 2011: Global Lambda Integrated Facility Visualization by Robert Patterson, NCSA, University of Illinois at Urbana-Champaign Data Compilation by Maxine D. Brown, University of Illinois at Chicago Texture Retouch by Jeff Carpenter, NCSA Earth Texture, visibleearth.nasa.gov www.glif.is



LHCONE services



- **Layer3 VPN**
- **Point-to-Point links**
- **Monitoring**

Openlab and IT-CS

Openlab project: **CINBAD**

CERN **I**nvestigation of **N**etwork **B**ehaviour and **A**nomaly **D**etection

Project Goals:

Understand the behaviour of large computer networks (10'000+ nodes) in High Performance Computing or large Campus installations to be able to:

- detect traffic anomalies in the system
- perform trend analysis
- automatically take counter measures
- provide post-mortem analysis facilities

Resources:

- In collaboration with HP Networking
- Two Engineers in IT-CS

Results



Project completed in 2010

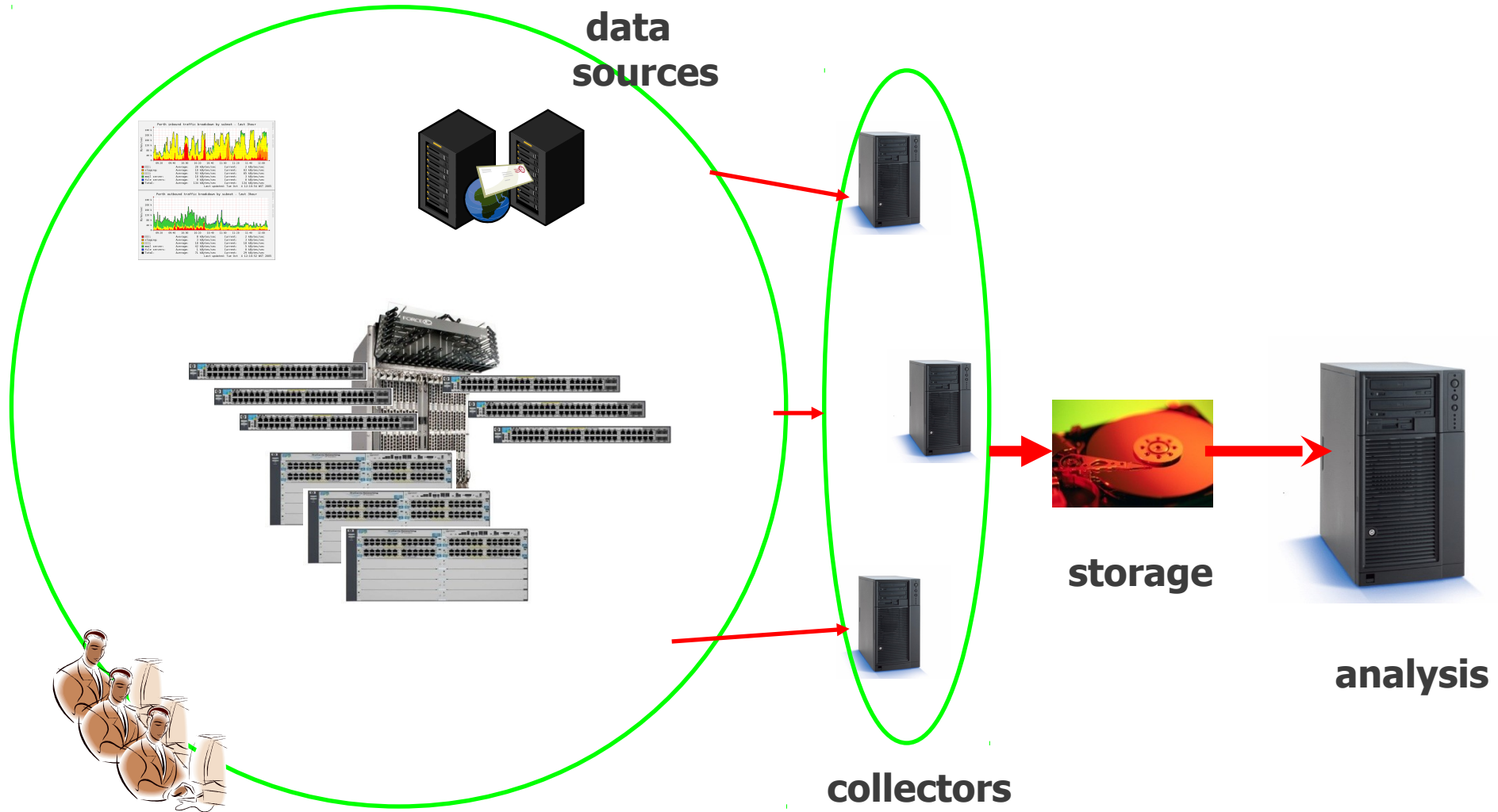
For CERN:

Designed and deployed a complete framework (hardware and software) to detect anomalies in the Campus Network (GPN)

For HP:

Intellectual properties of new technologies used in commercial products

CINBAD Architecture



Openlab project: **WIND**

Wireless Infrastructure Network Deployment

Project Goals

- Analyze the problems of large scale wireless deployments and understand the constraint
- Simulate behaviour of WLAN
- Develop new optimisation algorithms

Resources:

- In collaboration with HP Networking
- Two Engineers in IT-CS
- Started in 2010

Wireless LAN (WLAN) deployments are problematic:

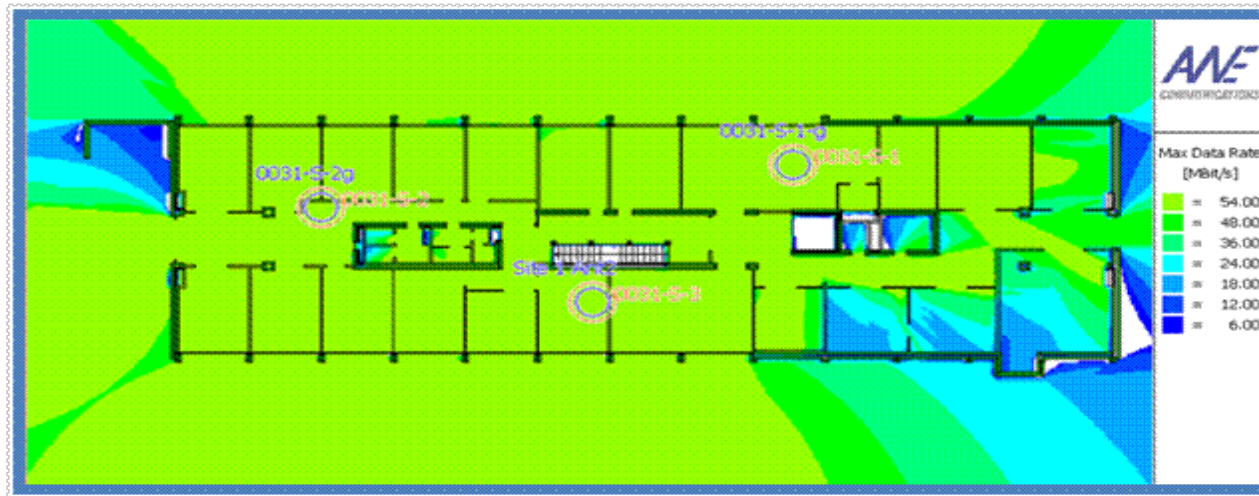
- Radio propagation is very difficult to predict
- Interference is an ever present danger
- WLANs are difficult to properly deploy
- Monitoring was not an issue when the first standards were developed
- When administrators are struggling just to operate the WLAN, performance optimisation is often forgotten

Example: Radio interferences

Max data rate in 0031-S: The APs work on the same channel



Max data rate in 0031-S: The APs work on 3 independent channels



Expected results



Extend monitoring and analysis tools

Act on the network

- smart load balancing
- isolating misbehaving clients
- intelligent minimum data rates

More accurate troubleshooting

Streamline WLAN design

Openlab project: **ViSION**



Project Goals:

- Develop a SDN traffic orchestrator using OpenFlow

Resources:

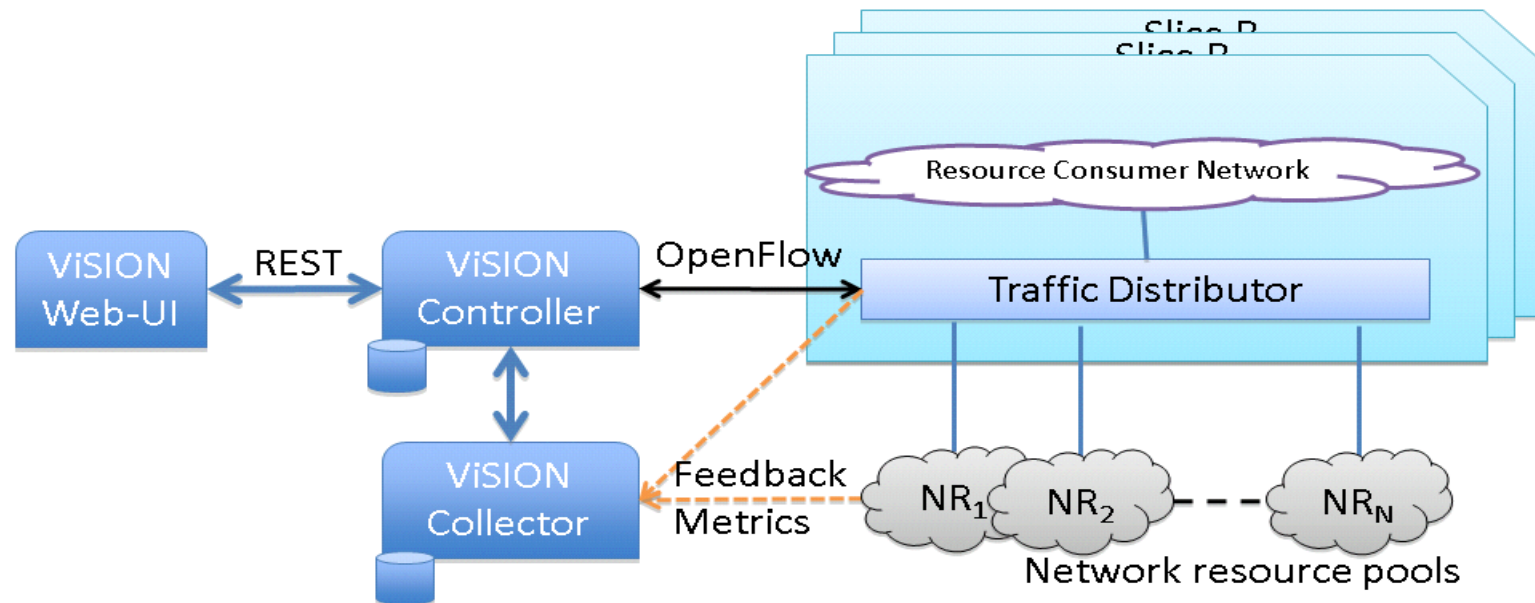
- In collaboration with HP Networking
- Two Engineers in IT-CS
- Started in 2012

SDN traffic orchestrator using OpenFlow:

- distribute traffic over a set of network resources
- perform classification (different types of applications and resources)
- perform load sharing (similar resources).

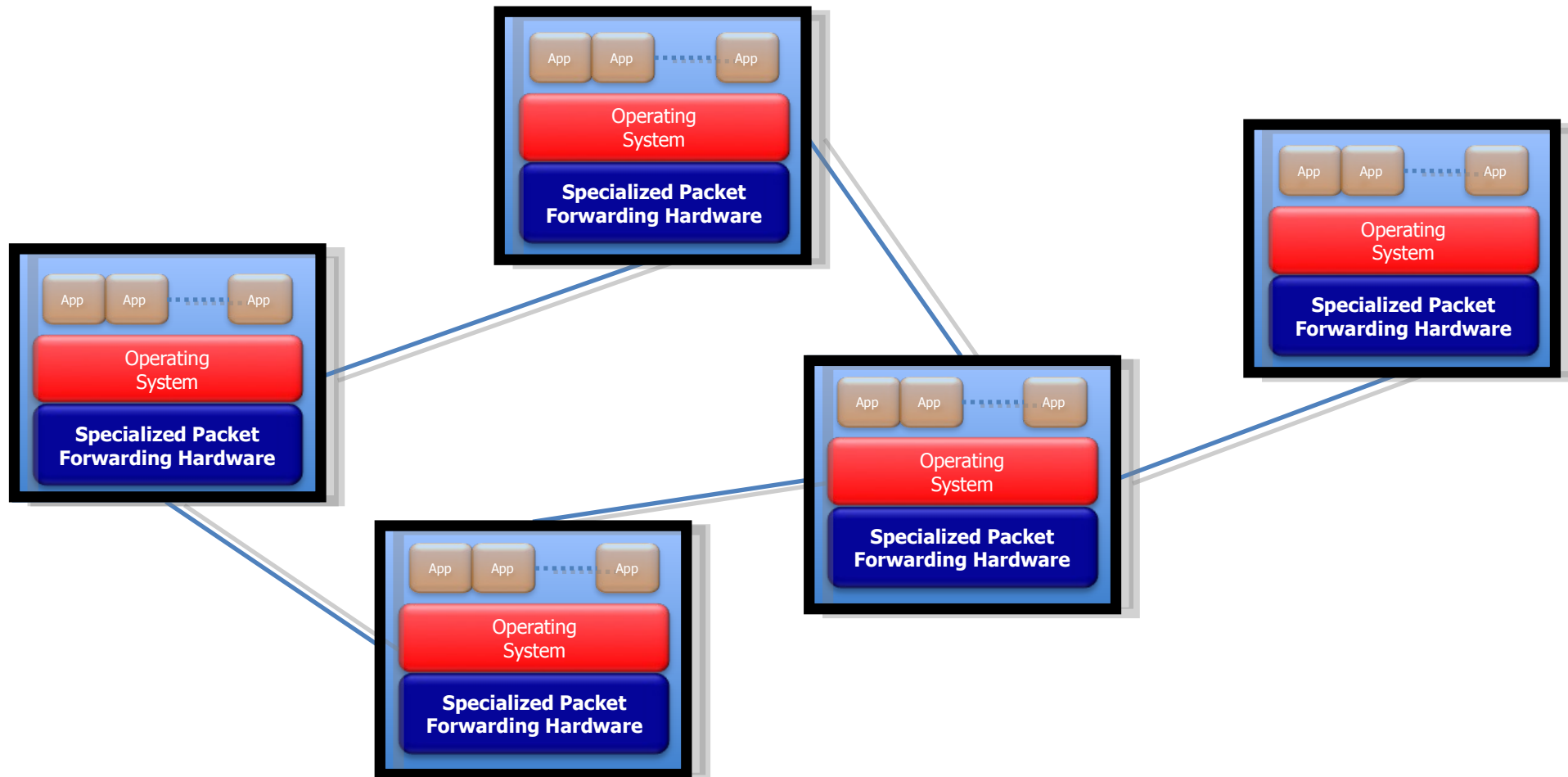
Benefits:

- improved scalability and control than traditional networking technologies



From traditional networks...

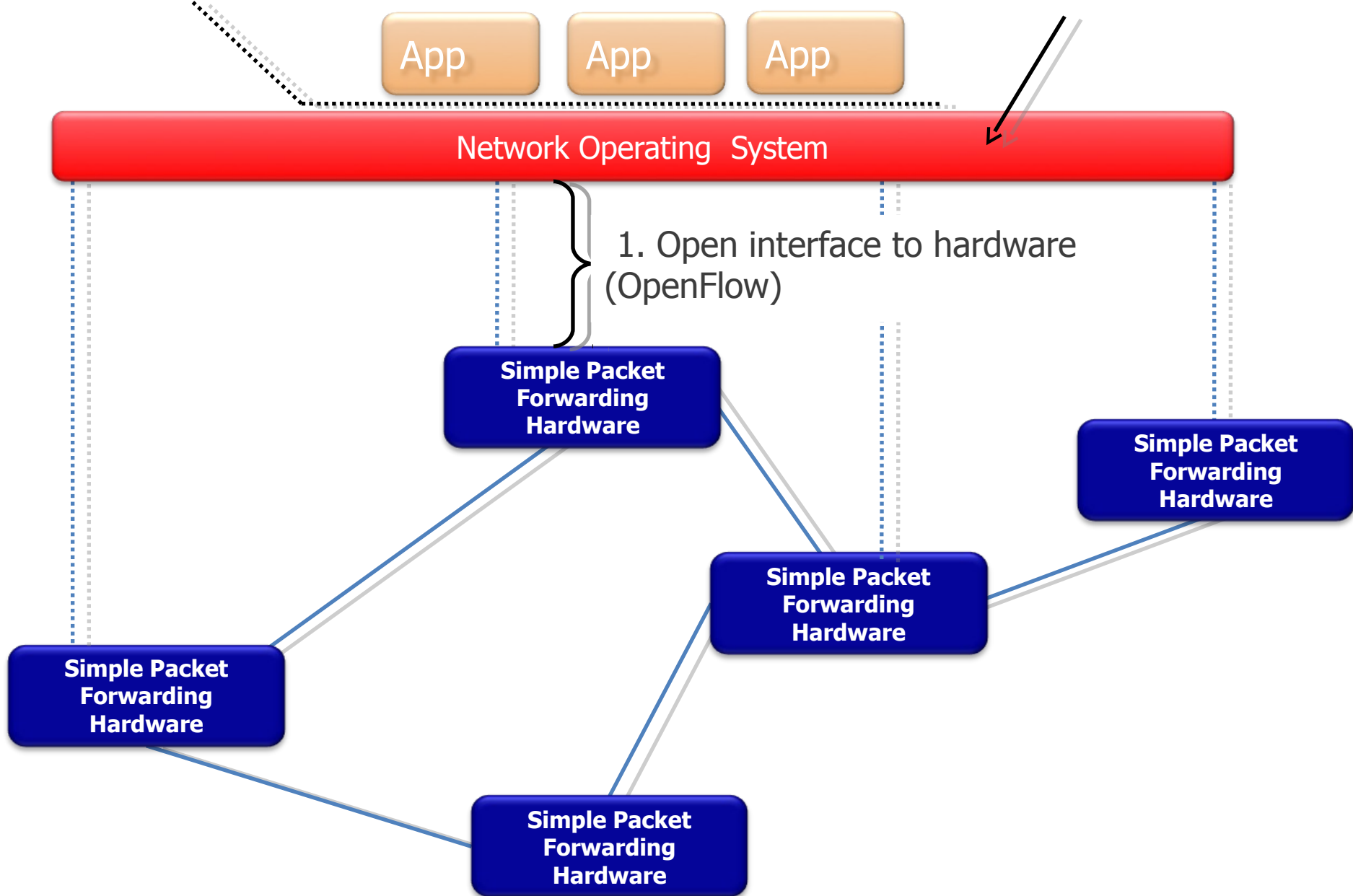
Closed boxes, fully distributed protocols



.. to Software Defined Networks (SDN)

3. Well-defined open API

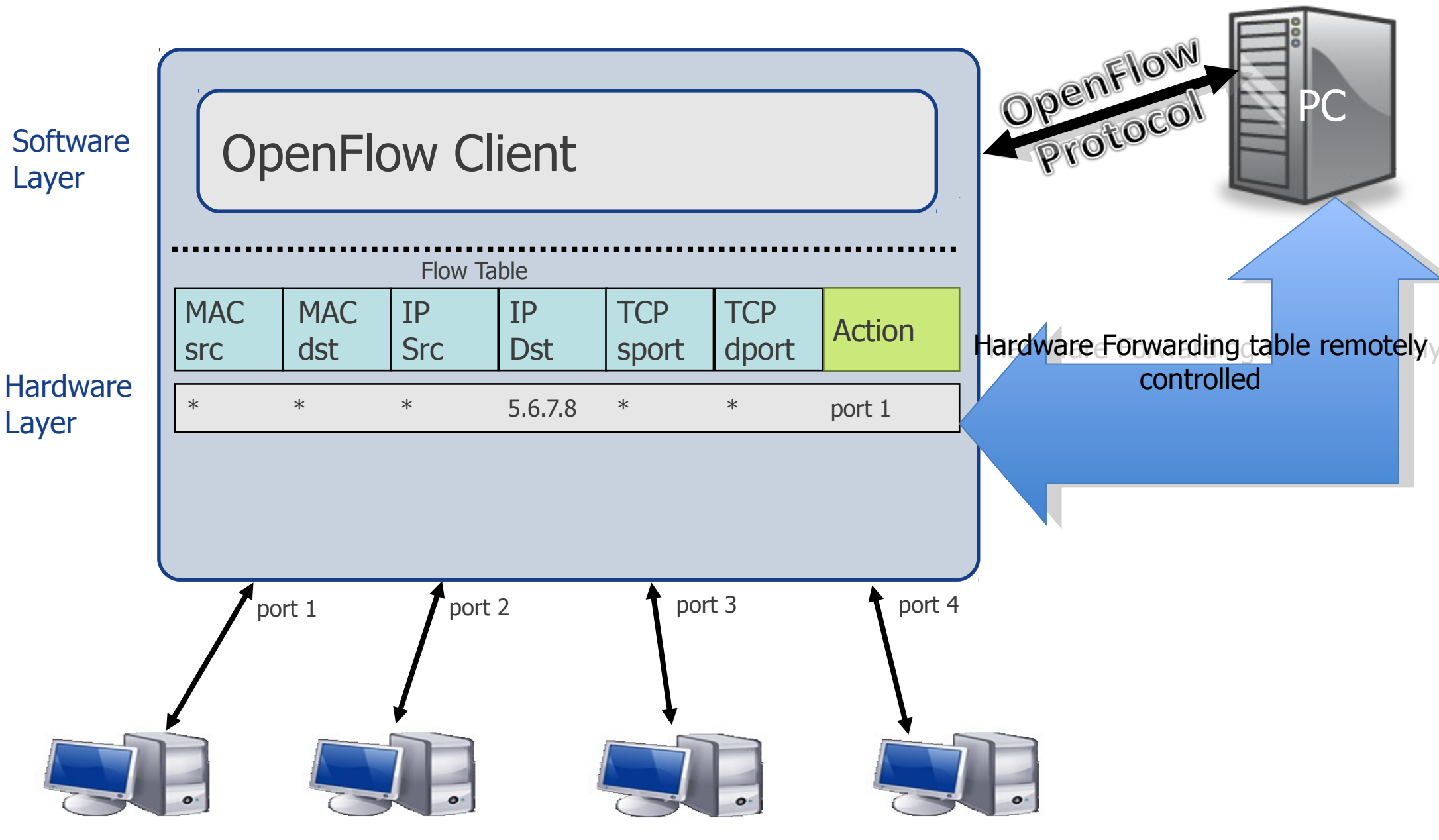
2. At least one good operating system
Extensible, possibly open-source



OpenFlow example



Controller



Conclusions

Conclusions



- The Data Network is an essential component of the LHC instrument
- The Data Network is a key part of the LHC data processing and will become even more important
- More and more security and design challenges to come

Credits



Artur Barczyk (LHCONE)

Dan Savu (VISION)

Milosz Hulboj (WIND and CINBAD)

Ryszard Jurga (CINBAD)

Sebastien Ceuterickx (WIND)

Stefan Stancu (VISION)

Vlad Lapadatescu (WIND)

What's next



SWAN: Space Wide Area Network :-)

