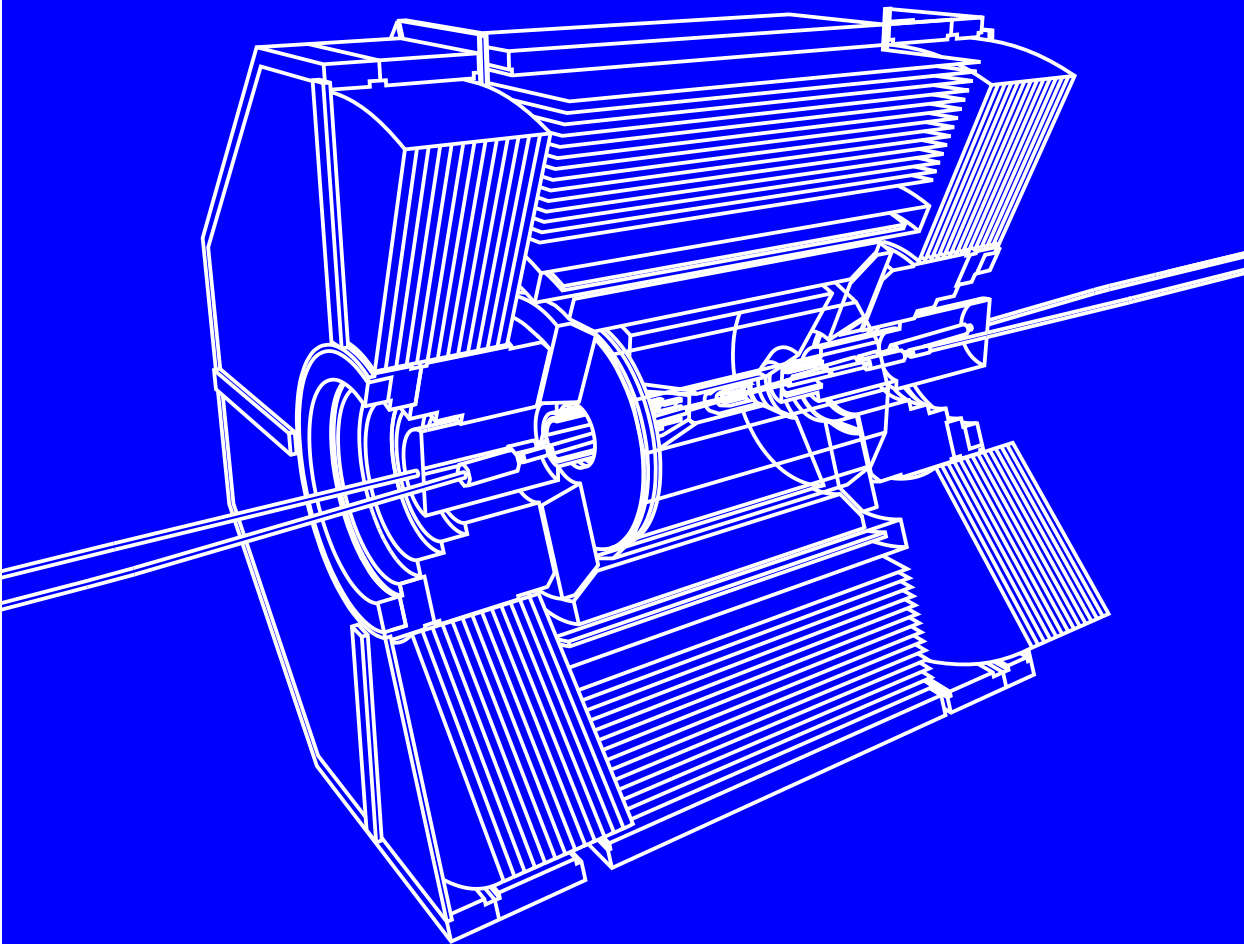




Belle II Computing and requirement of the network



LHCONE Asia-Pacific workshop
@ Nantou, Taiwan

Belle II computing
resource, design, network

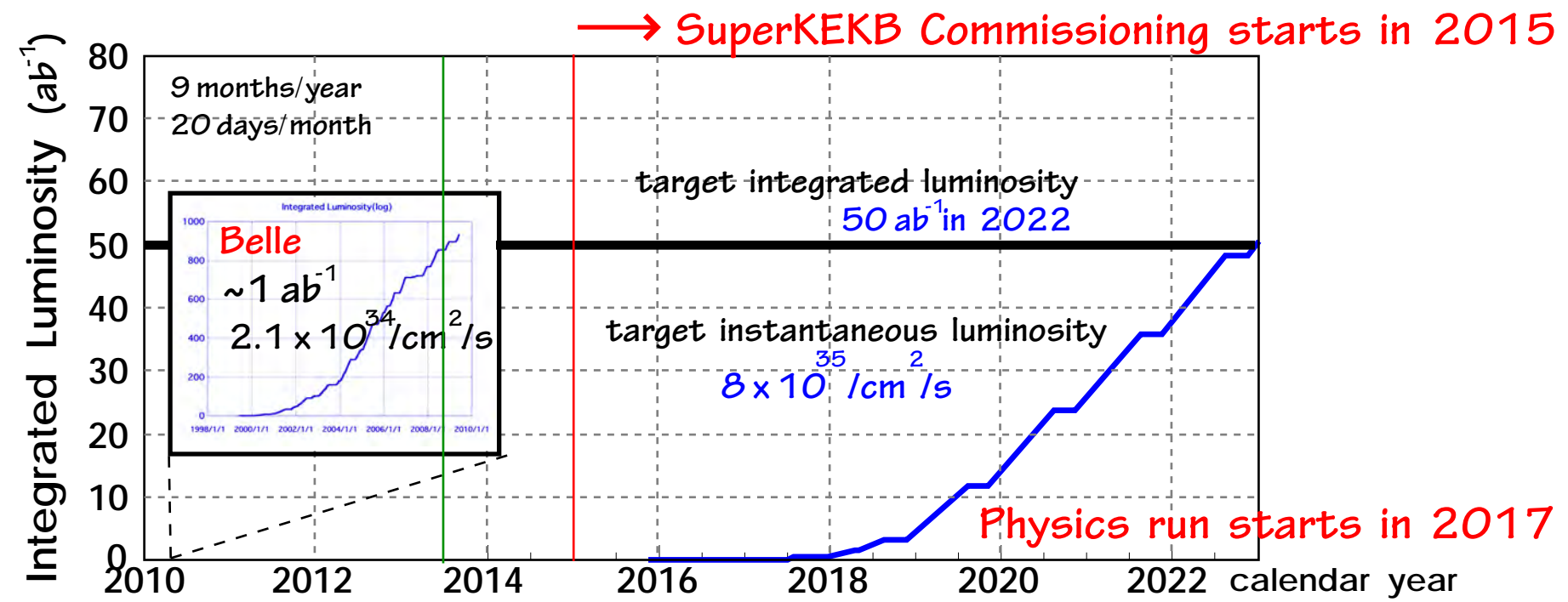
Network data challenge
Trans-Pacific
Trans-Atlantic

LHCONE(-like layer) for Belle II ?

Takanori Hara (KEK)
takanori.hara@kek.jp

13 Aug., 2014

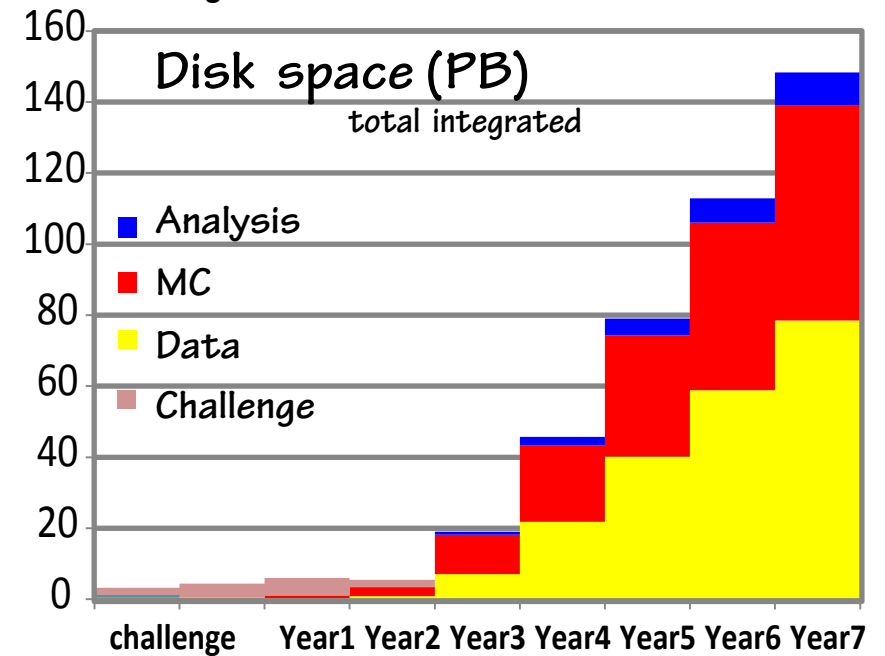
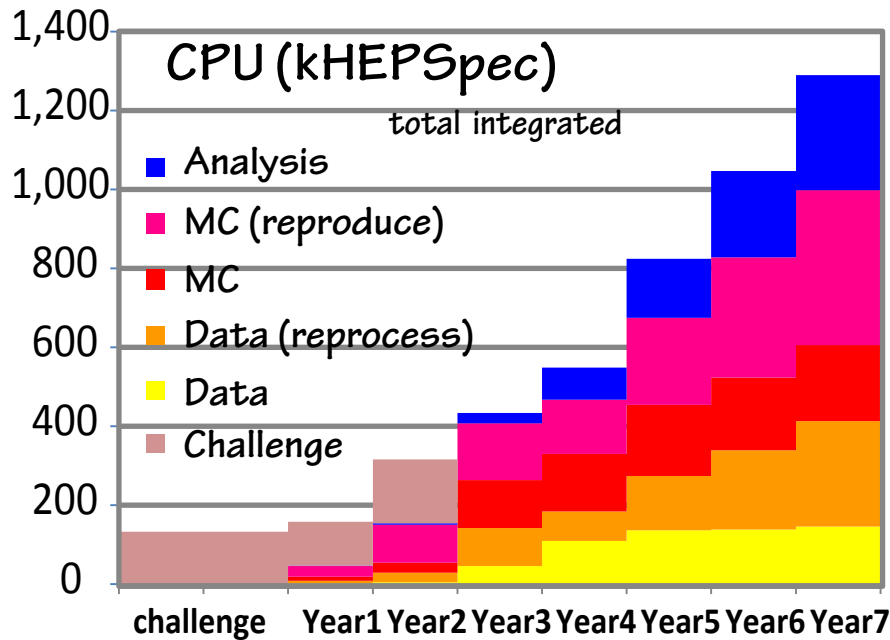
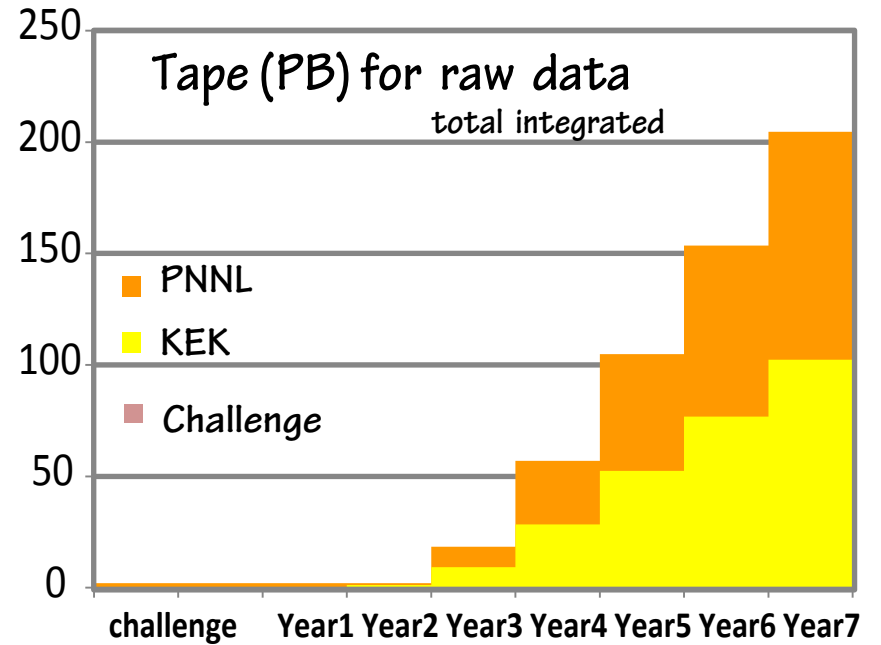
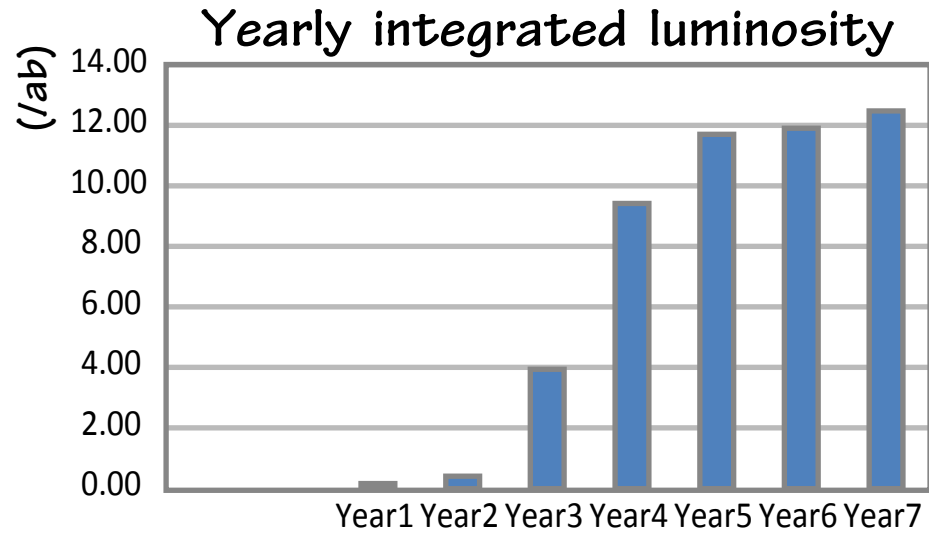
Luminosity Prospect

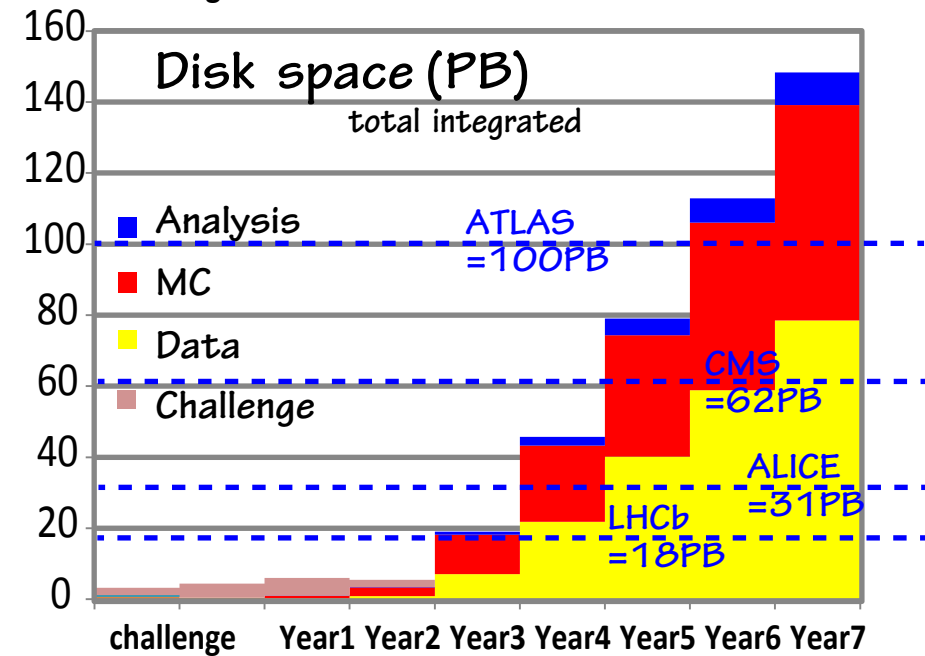
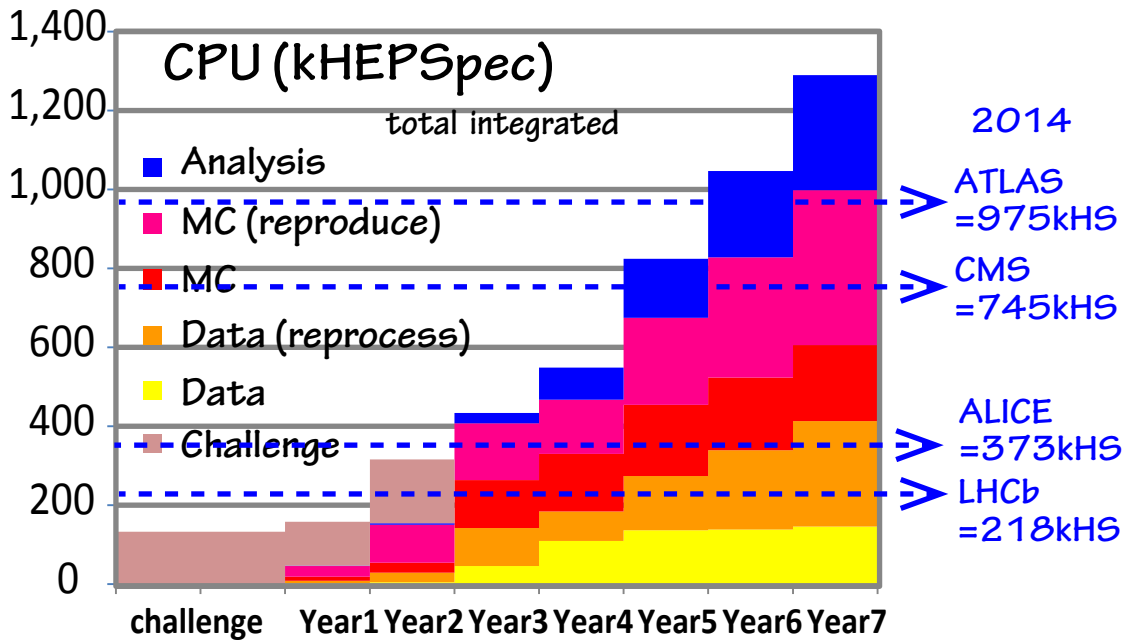
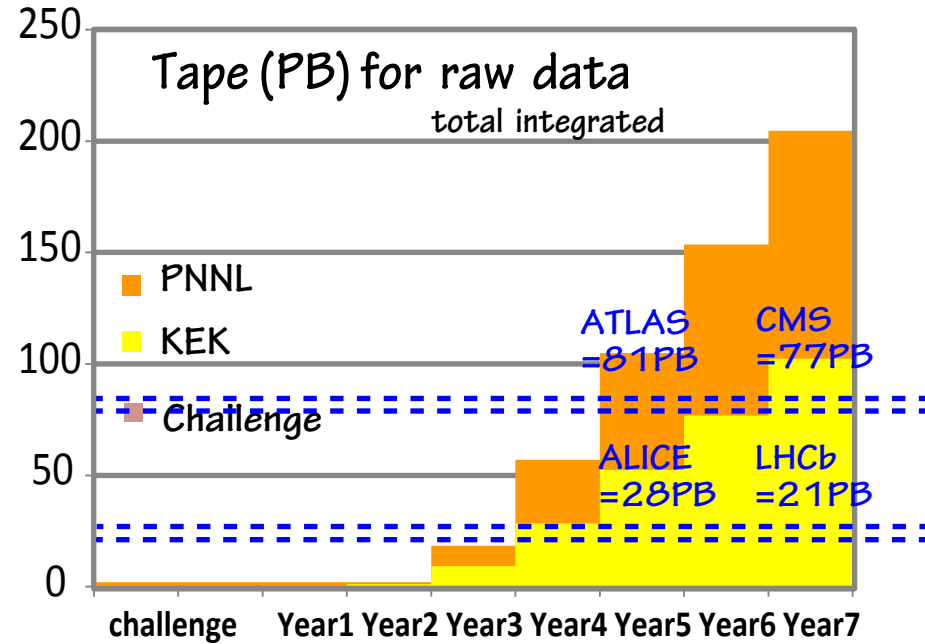
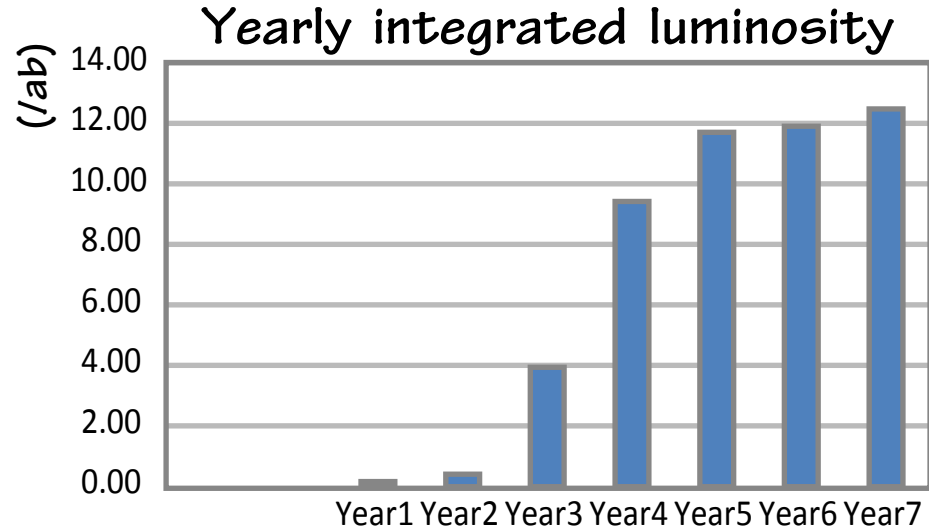


Experiment	Event size	Rate @ Storage	Rate @ Storage
	[kB]	[event/sec]	[MB/sec]
Belle II	300	6,000	1,800
ALICE (Pb-Pb)	50,000	100	4,000
ALICE (p-p)	2,000	100	200
ATLAS	1,500	600	700
CMS	1,500	150	225 (<~1000)
LHCb	55	4,500	250

(@ max. luminosity)

(LHC experiments : as seen in 2011/2012 runs)





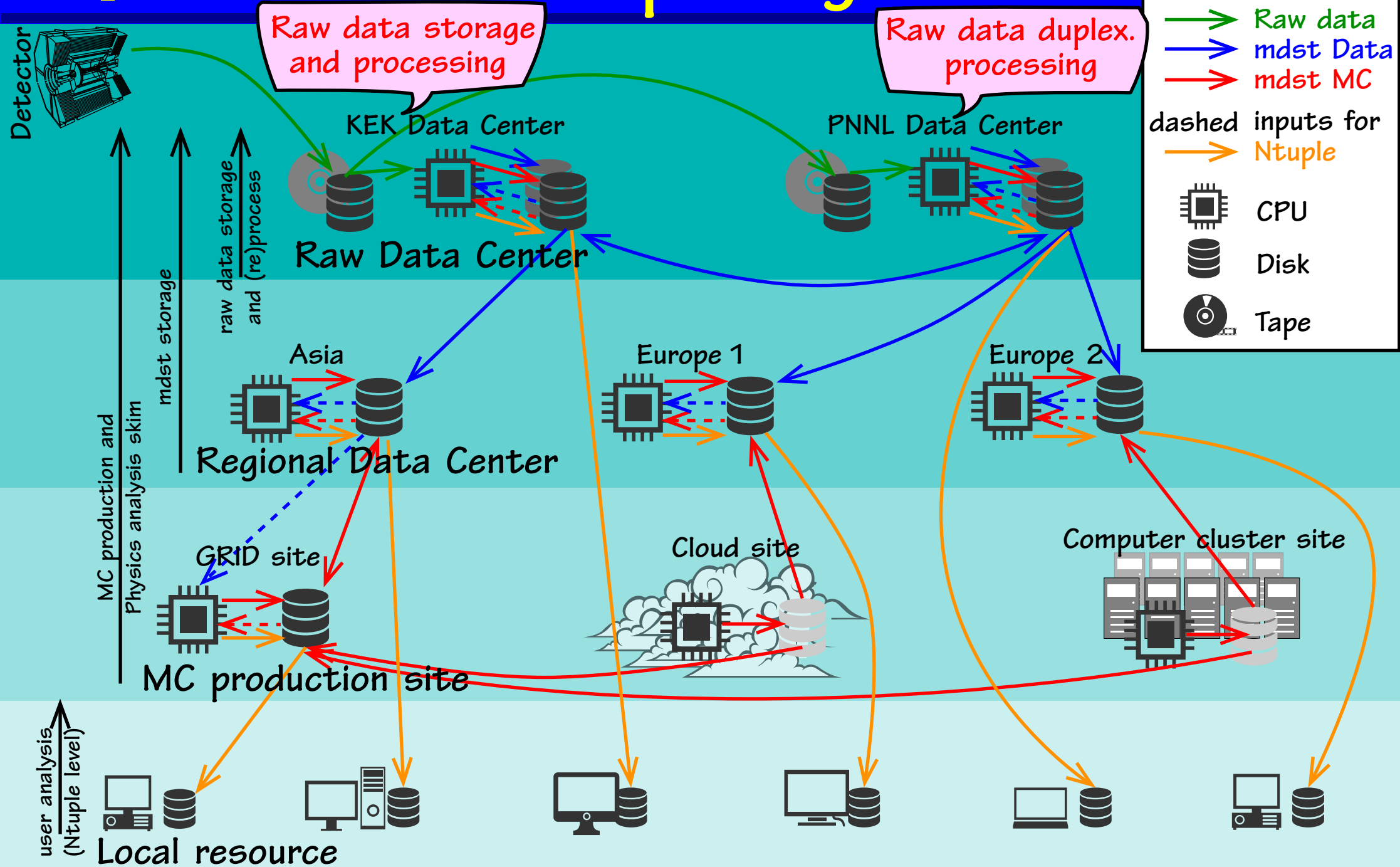
Belle II Collaboration



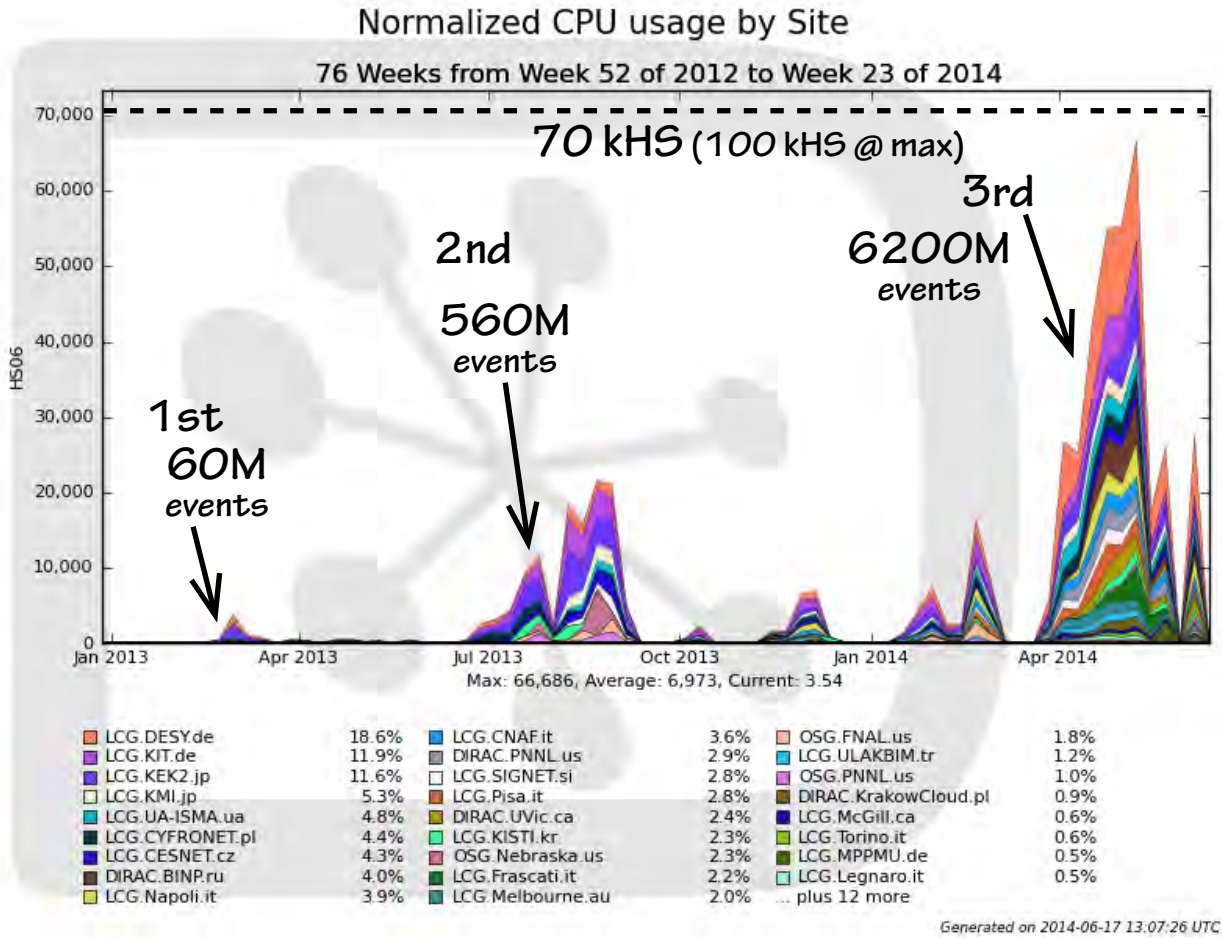
c.f.
 ATLAS, 38 countries, 177 institutes, ~3000 members
 CMS : 42 countries, 182 institutes, 4300 members
 ALICE : 36 countries, 131 institutes, 1200 members
 LHCb : 16 countries, 67 institutes, 1060 members



Belle II Computing Model



Current status of computing



15 countries/regions

27 sites (+ 2 non-Belle II sites)

HEPHY (Vienna) and MPPMU (Munich) joined recently

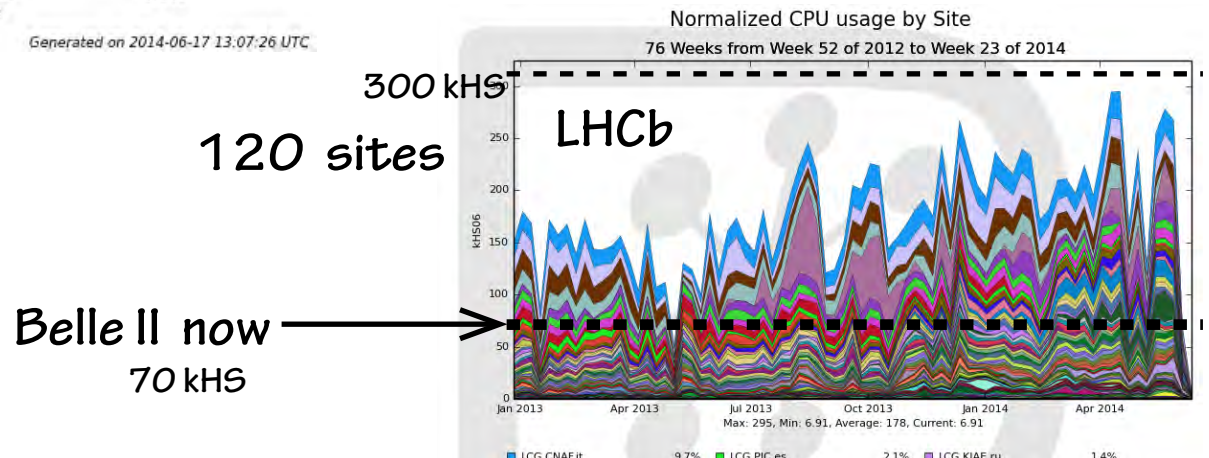
GRID, Cloud, local cluster is available

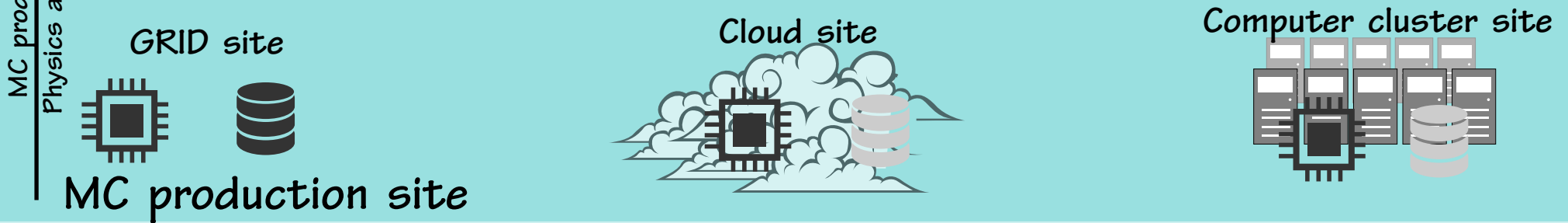
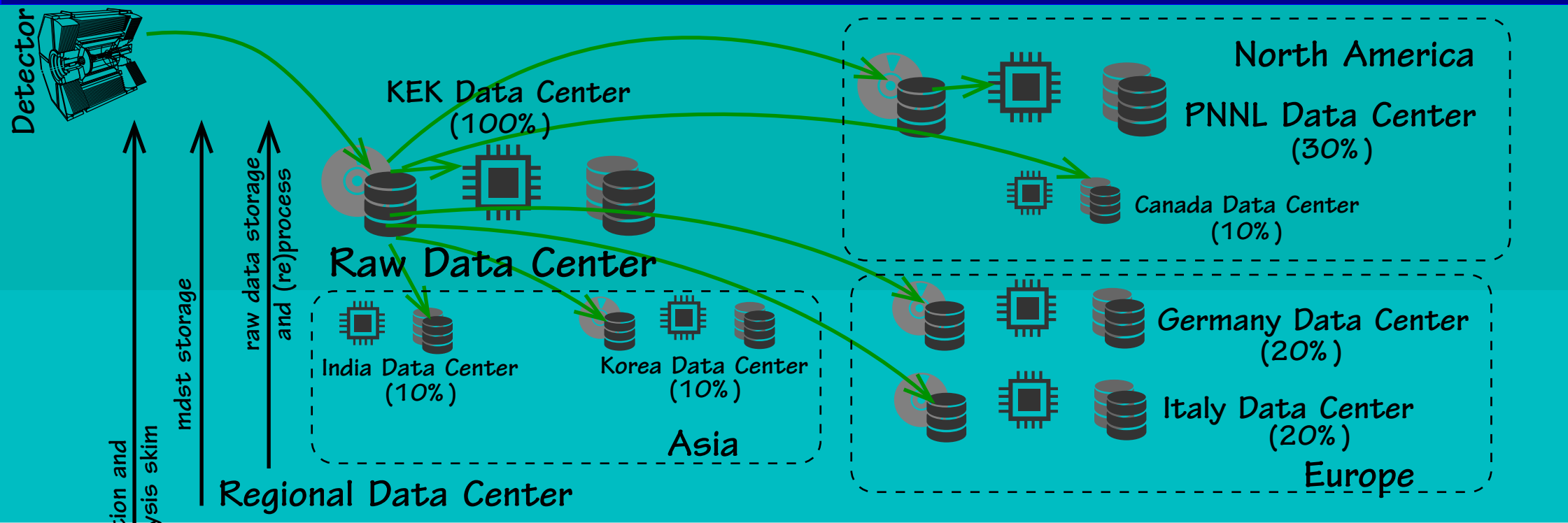
First official release of MC samples

BB generic decay/continuum tau pair

(corresponding to 100fb^{-1} w/ and w/o BG)

Trans-pacific / trans-atlantic network data transfer challenge



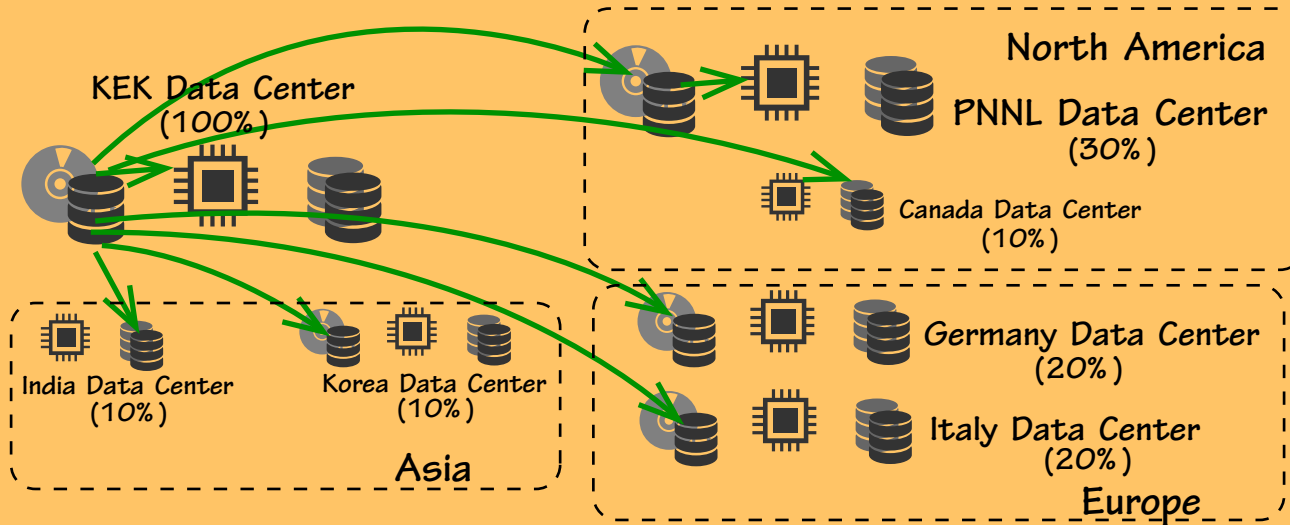


Raw Data Distribution

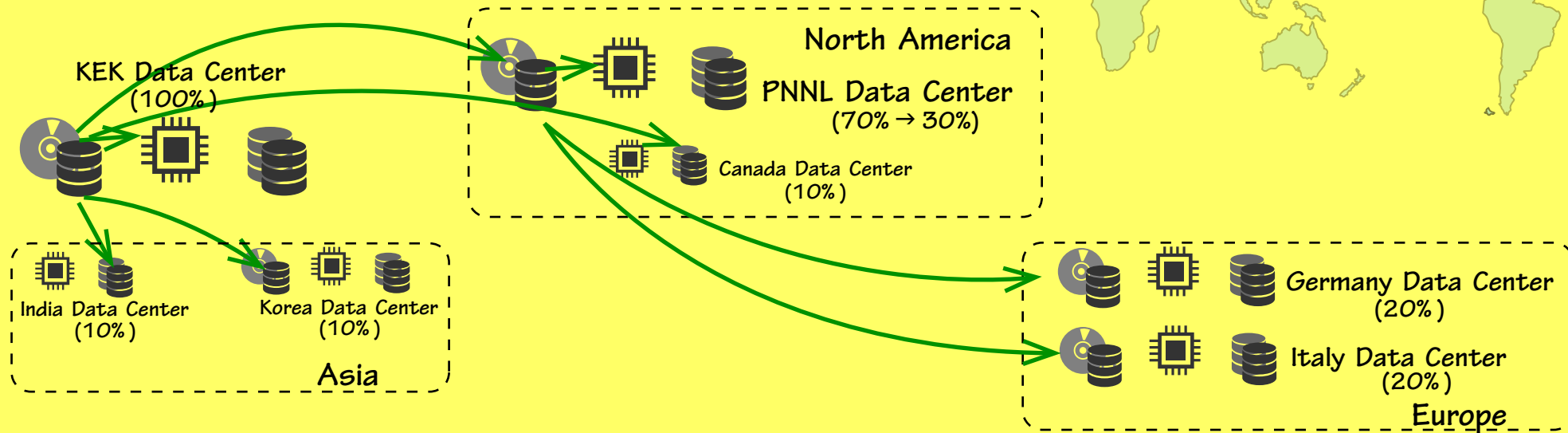
until Year 3



Scenario 1
(copy from KEK)



Scenario 2
(2step copy, KEK → PNNL → Europe)



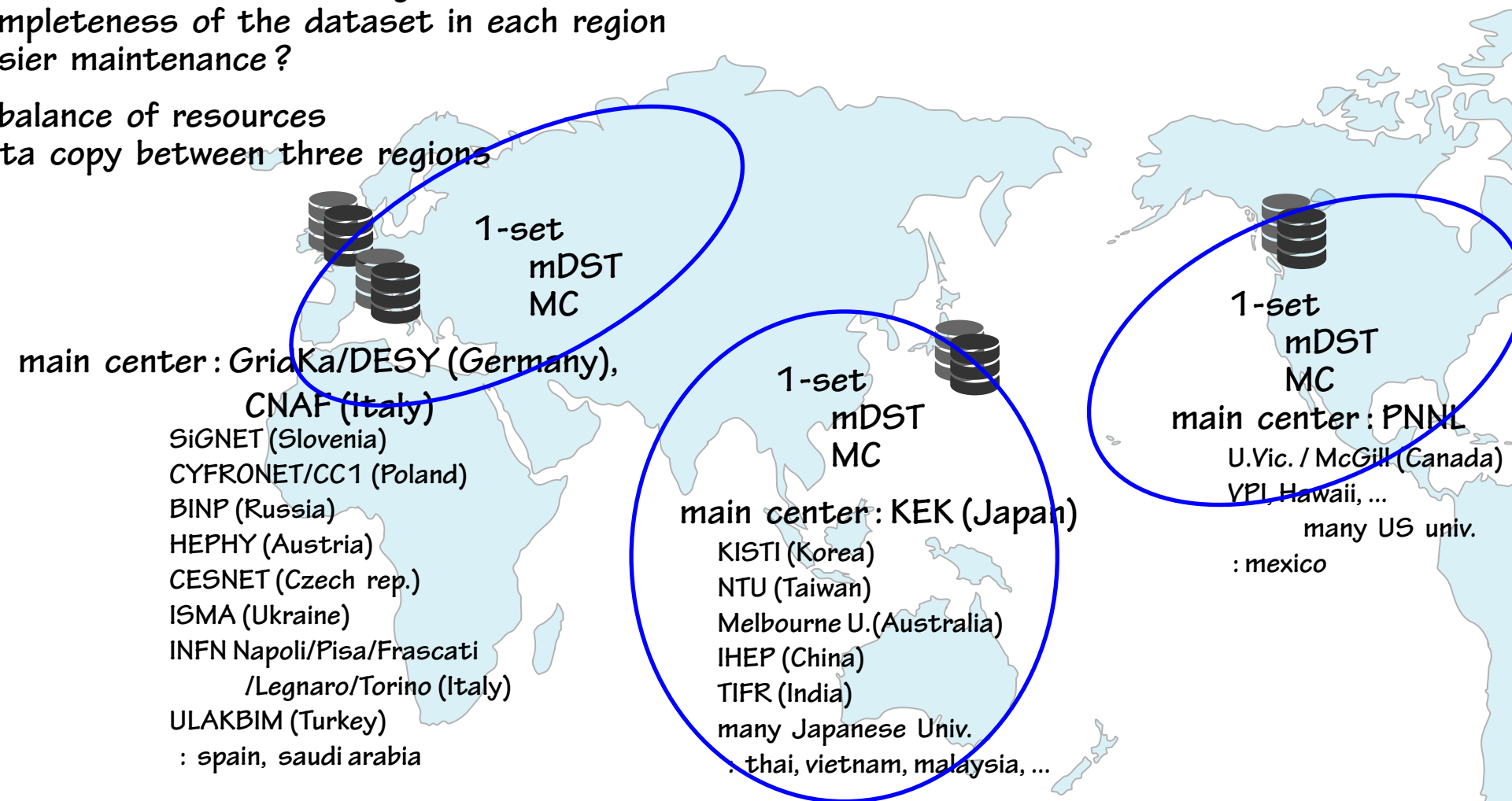
mDST/MC Data Distribution

mDST (data) is copied in Asia, Europe, and USA

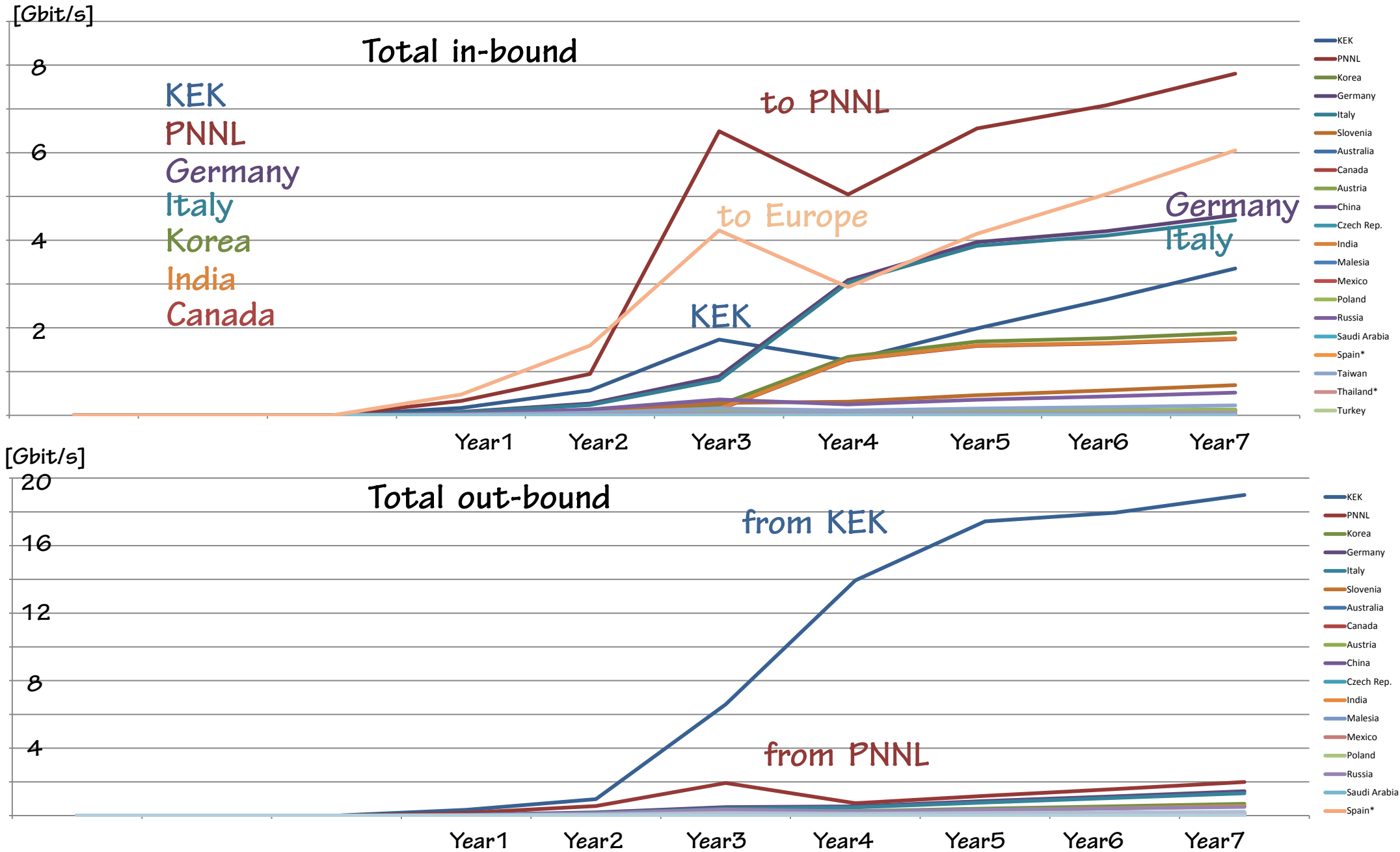
For the MC data seems to be natural to be the similar structure

better network? in each region
 completeness of the dataset in each region
 easier maintenance?

unbalance of resources
 data copy between three regions



Scenario 1



Network Connectivity

Current Connectivity

Trans-Pacific

10G : Tokyo - LA
10G : Tokyo - NY
10G : Osaka - Washington

Trans-Atlantic

3 x 10G : NY - Amsterdam
3 x 10G : Washington - Frankfurt
ANA-100G NY - Amsterdam

Trans-Asia

2.5G : Madrid-Mumbai
2.5G : Singapore-Mumbai
10G : Japan-Singapore

“Planned” Connectivity

Trans-Pacific

SINET5
100G link to US
in 2016

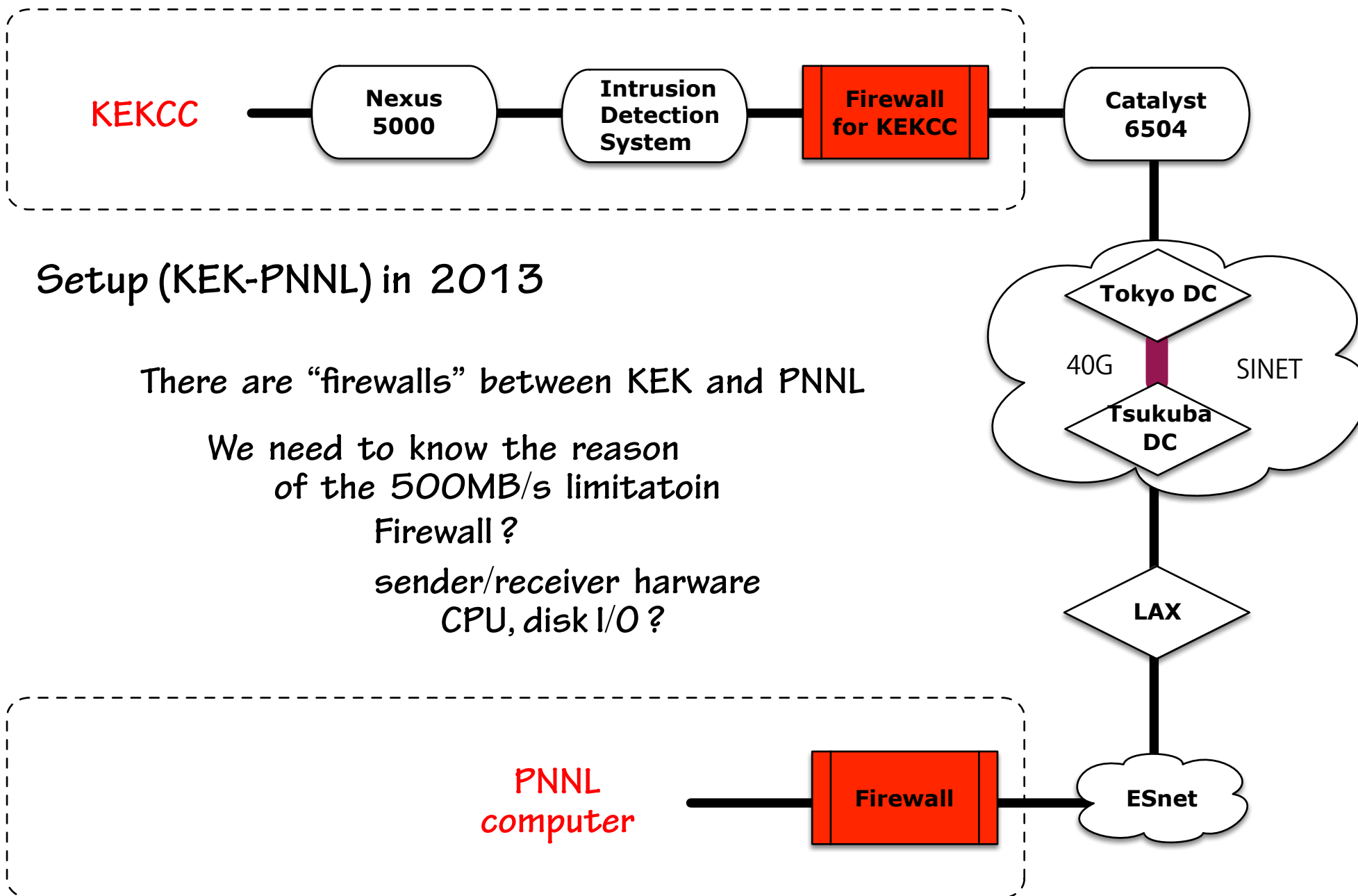
Trans-Atlantic

EEX (ESNet Extension to Europe)
2 x 100G : NY - London
100G : Washington - Geneva
40G : Boston - Amsterdam

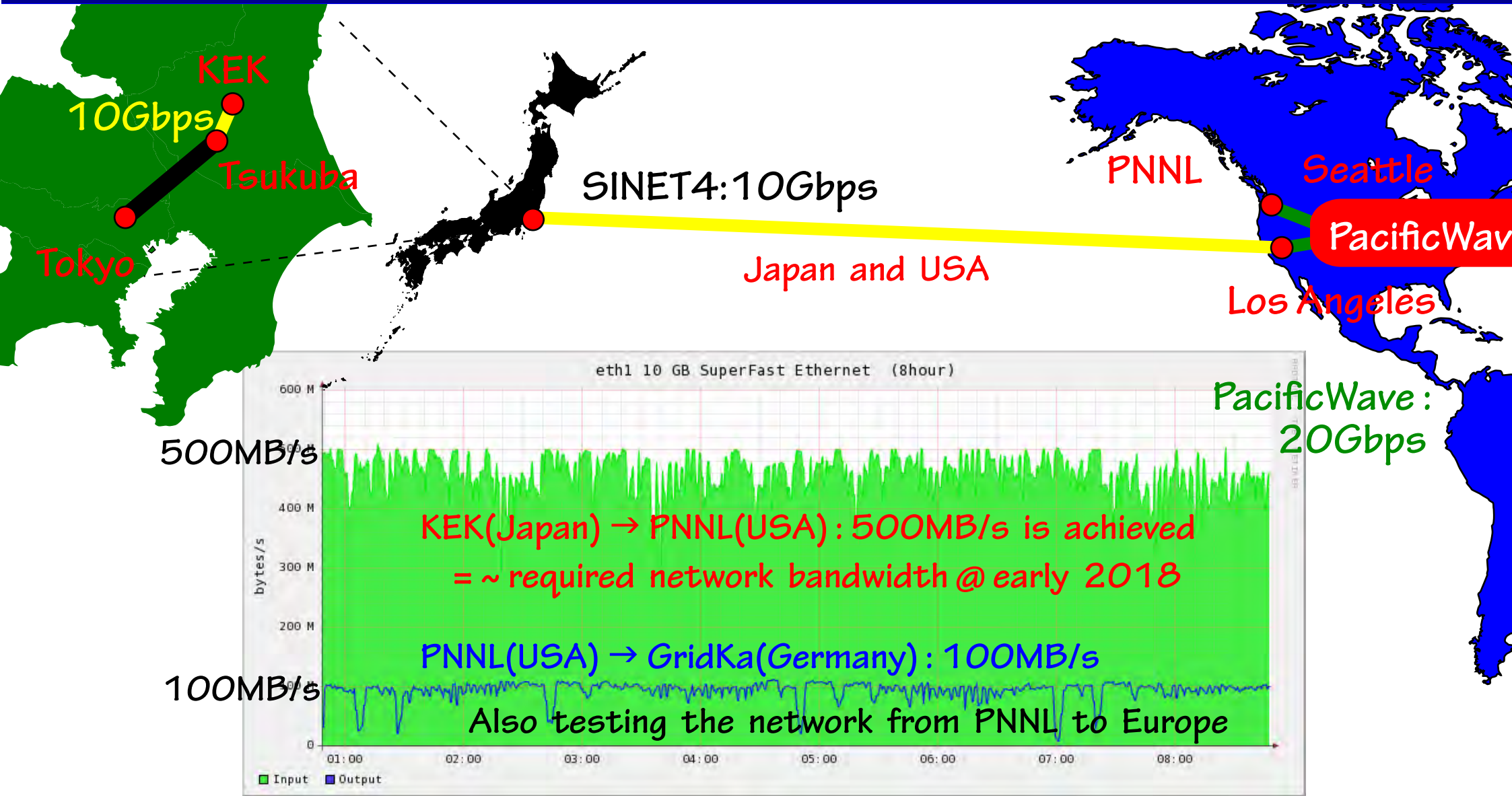
Trans-Asia

10G : Mumbai - GEANT
SINET ?

Trans-Pacific data challenge



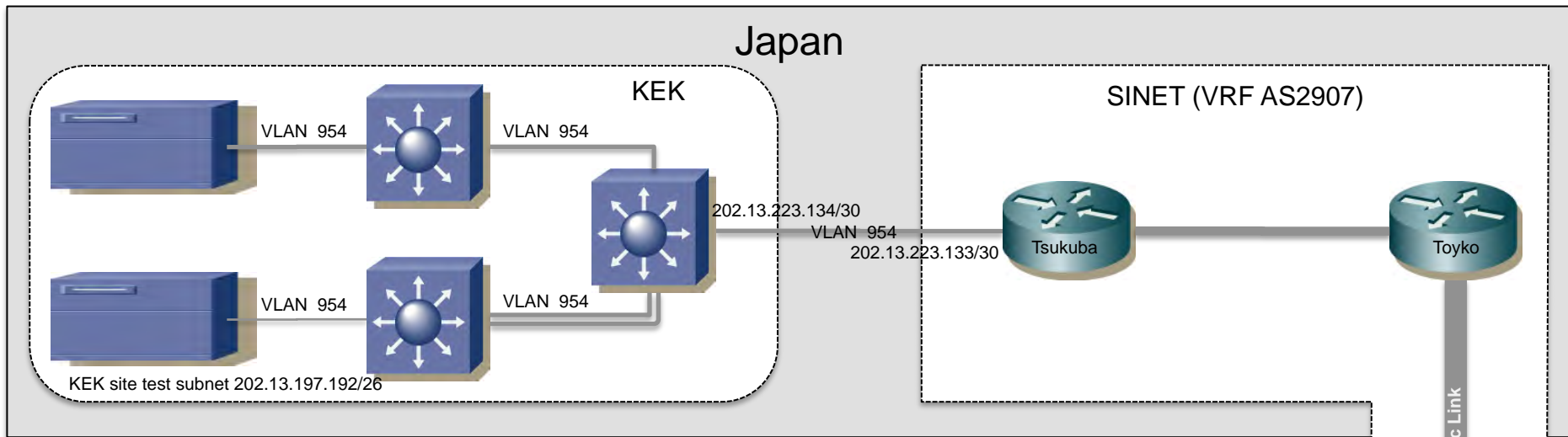
Trans-Pacific data challenge



But not enough for the network bandwidth @ middle of Year4 and later (~2GB/s)

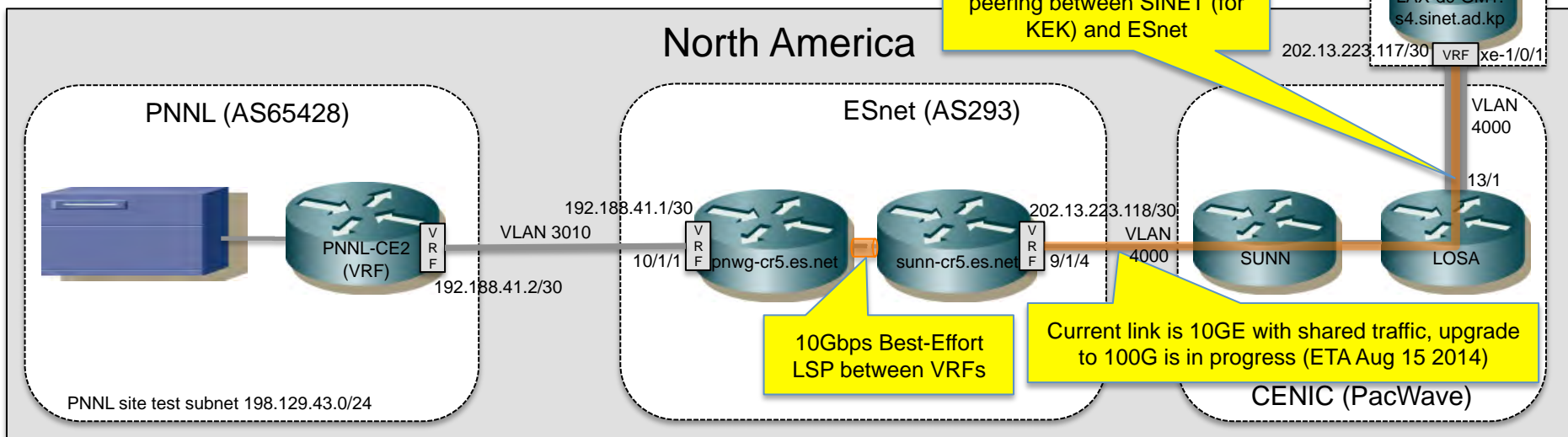
We need a 40Gbps - 100Gbps network between Japan and USA

New setup (KEK-PNNL)



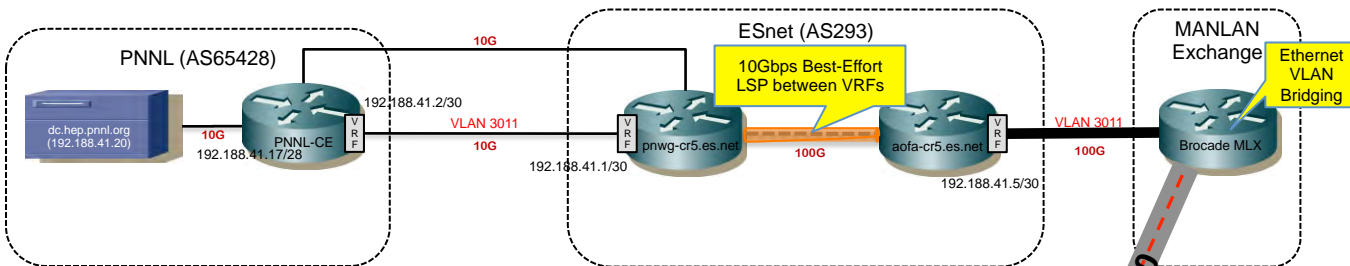
Belle-II Testing between PNNL and KEK

(Setup to stay in place thru 30 June 2016)



Trans-Atlantic data challenge

US side



Dedicated 10G link between PNNL DTN and ESNet
10G best-effort Label Switched Path in ESNet backbone

- “traceroute” was used to confirm the routing to each DTN
- “iperf” was used to do initial network transfer rate test
- “gridftp” and/or “srm-copy” was used to test site
- FTS3 server at GridKa was used to schedule data transfers

Test was done in May/June 2014

- Network providers setup the VLAN
- Local network providers and sites coordinated final configurations
- Sites must configure hardware interface to match destinations



Vincenzo Capone,
Aleksandr Kurbatov, Mian Usman



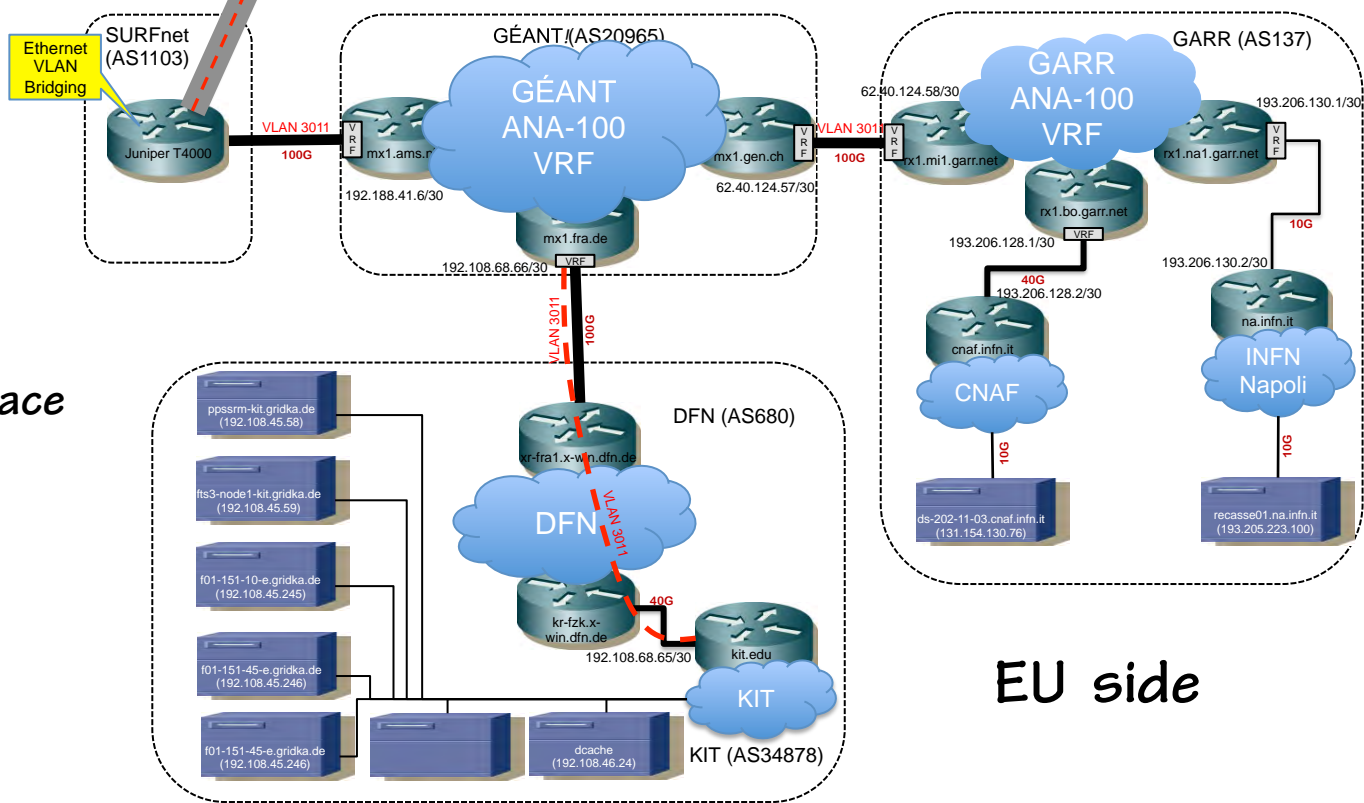
Chin Guok



Thomas Schmid, Hubert Weibel



Marco Marletta

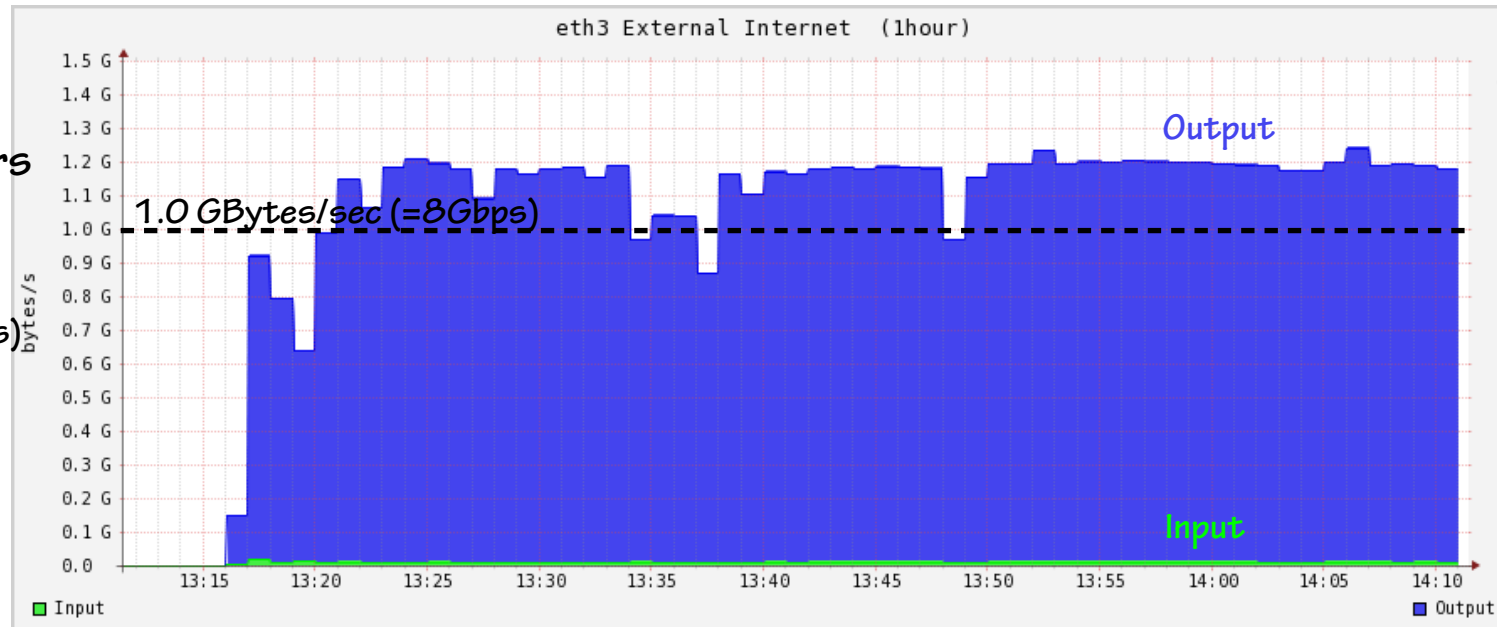


EU side

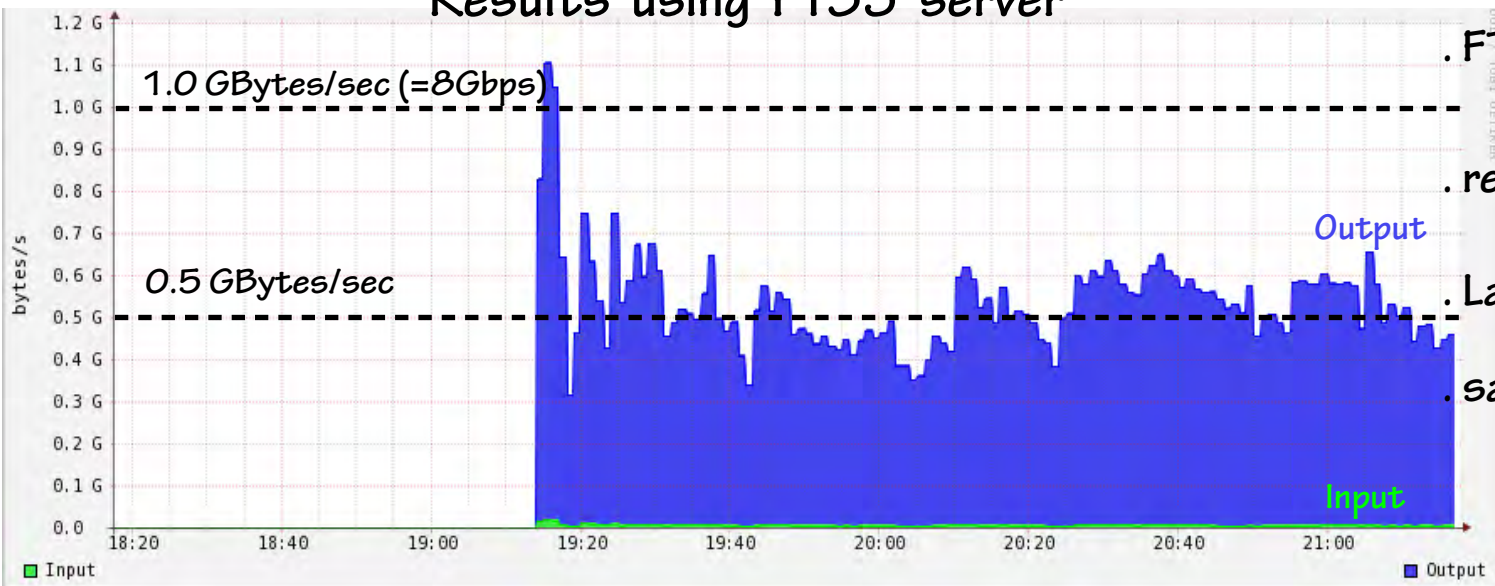
Trans-Atlantic data challenge

“iperf” results

- . Required several parallel transfers to reach network saturation
- . Reached ~9.6Gbps (>2x the Tier-1 EU site requirements)



Results using FTS3 server



. FTS3 optimization is not ideal:

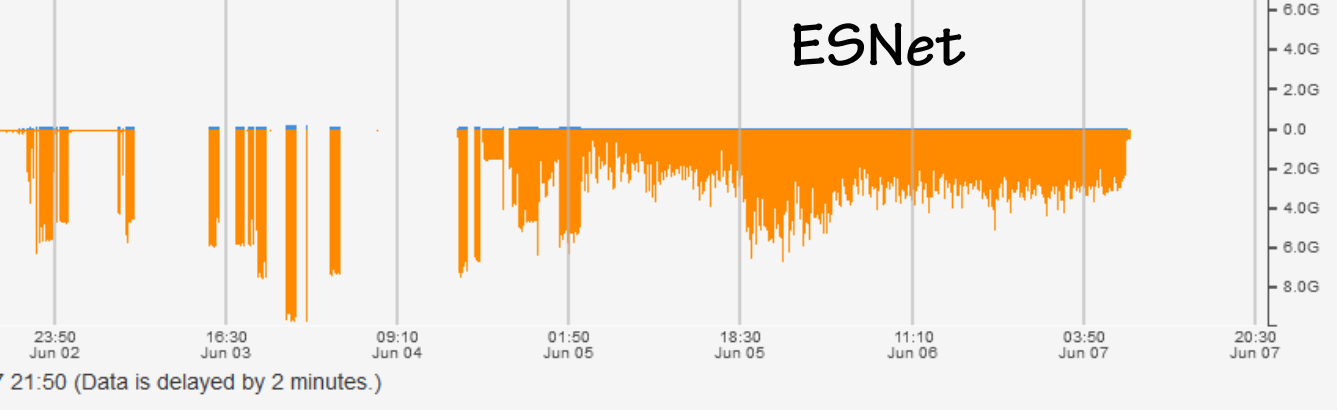
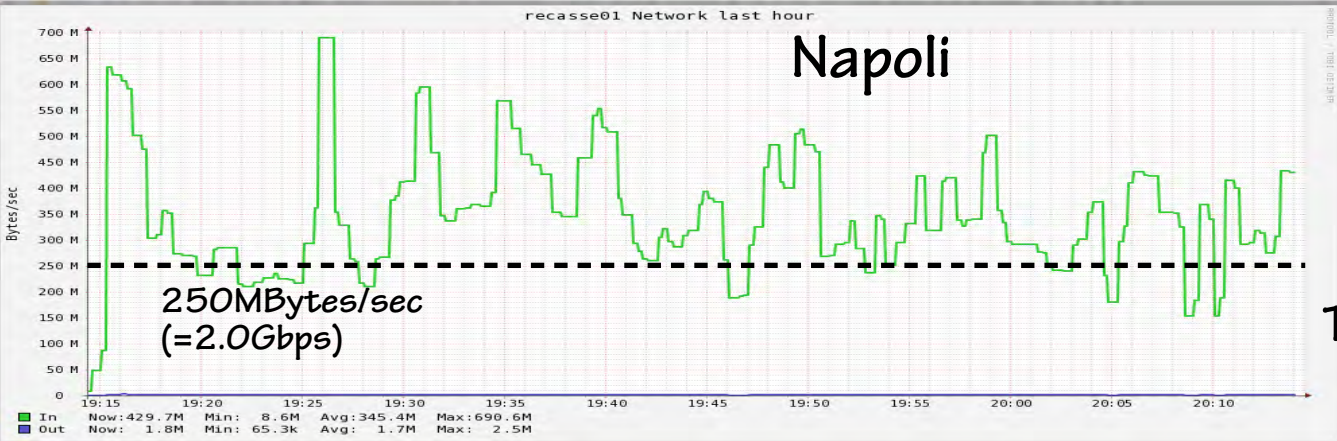
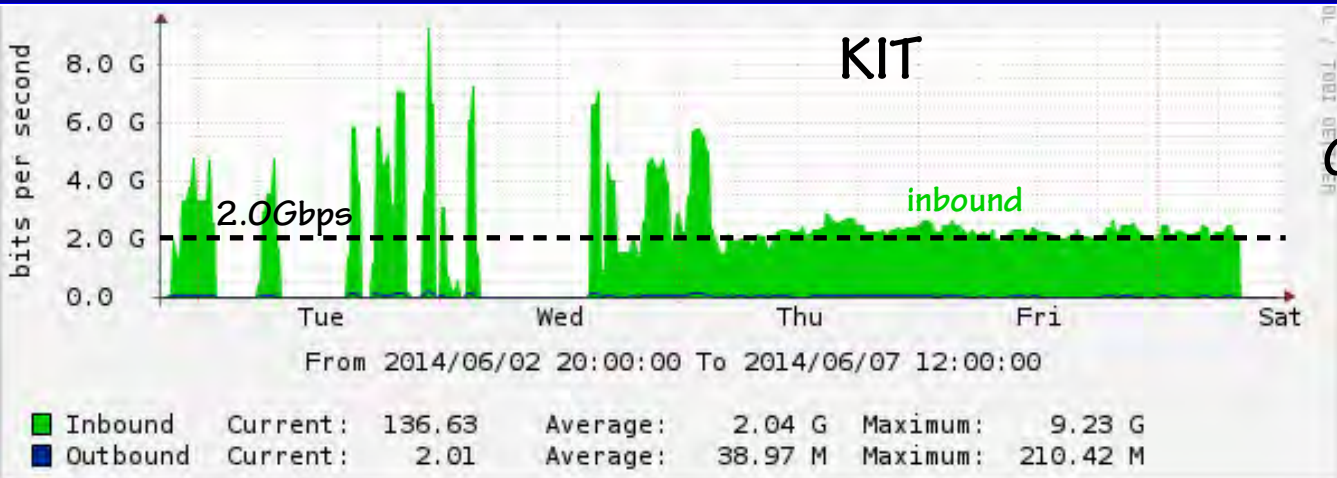
. reached network saturation but falls very quickly

. Large amount of drop packets

. satisfies the incoming network requirements for Tier1 EU sites up to calendar year “Year6”

(2021 or 2022)

Trans-Atlantic data challenge



Challenges encountered

- . The main issue was the configuration of the local network apparatus
- . Having all the servers at each site using/checking the proper network route
- . Hardware limitation (router, storage, etc)
- . Not having dedicated setups (shared with ATLAS, etc.)

To accommodate the increased rates

- . Modification of TCP windows was performed at PNNL and Italy
- . Routing hardware interface
- . Configure/tune network interrupts for multicore
- . Modification of the FTS3 optimization & global-timeout

LHCONE for Belle II ??

LHCONE is for LHC experiments

In Belle II

- . European sites have already joined to LHCONE
- . while, KEK and PNNL does not belong to LHCONE now

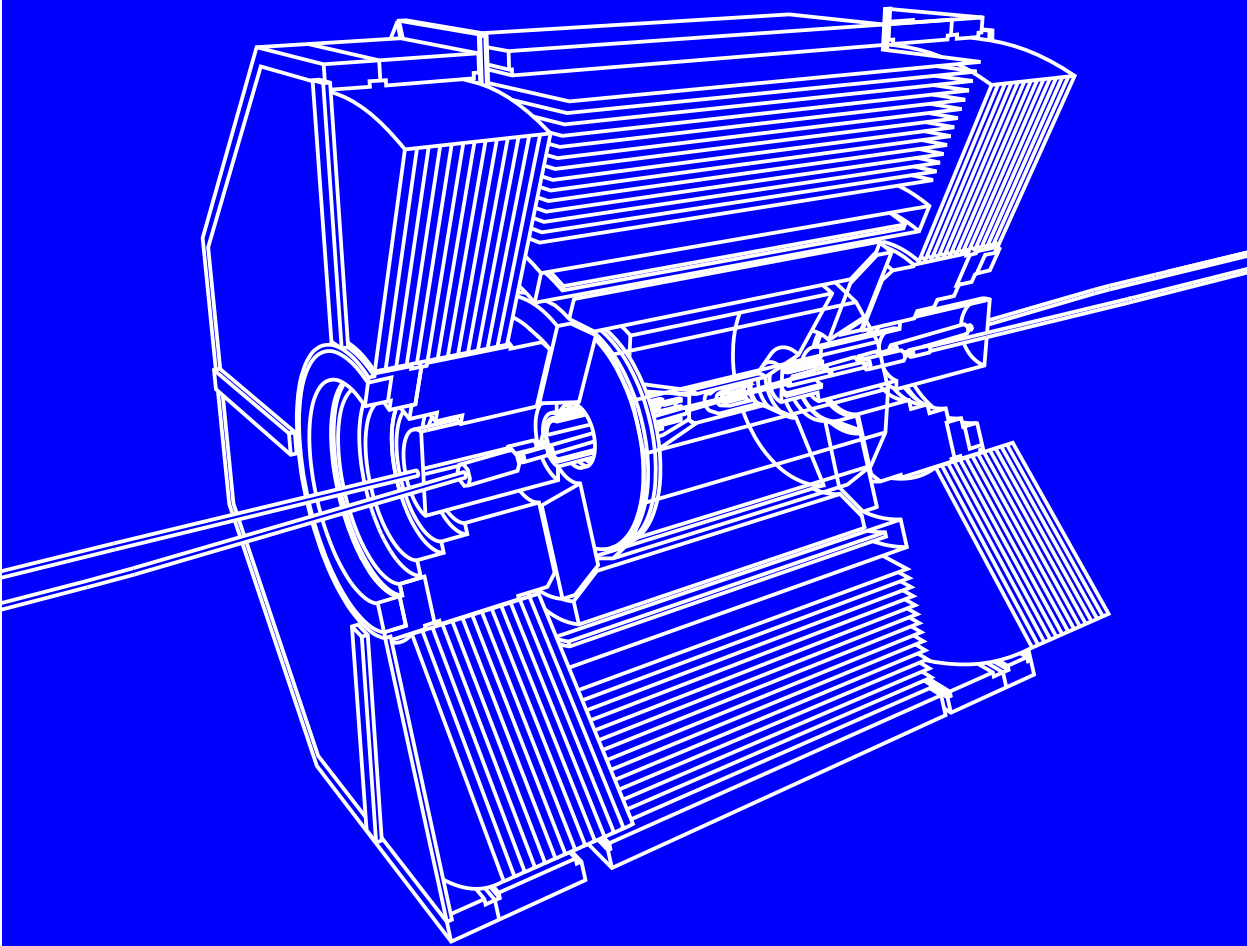
Our thoughts are

- . Belle II prefers to have a closed network like LHCONE
- . If configuring new VRFs for Belle II on each collaboration sites and related networks is difficult or makes any problem on operation, one possibility for Belle II is to join to LHCONE (if it is allowed.)

Considerations : to join LHCONE or to configure LHCONE-like VRF layer

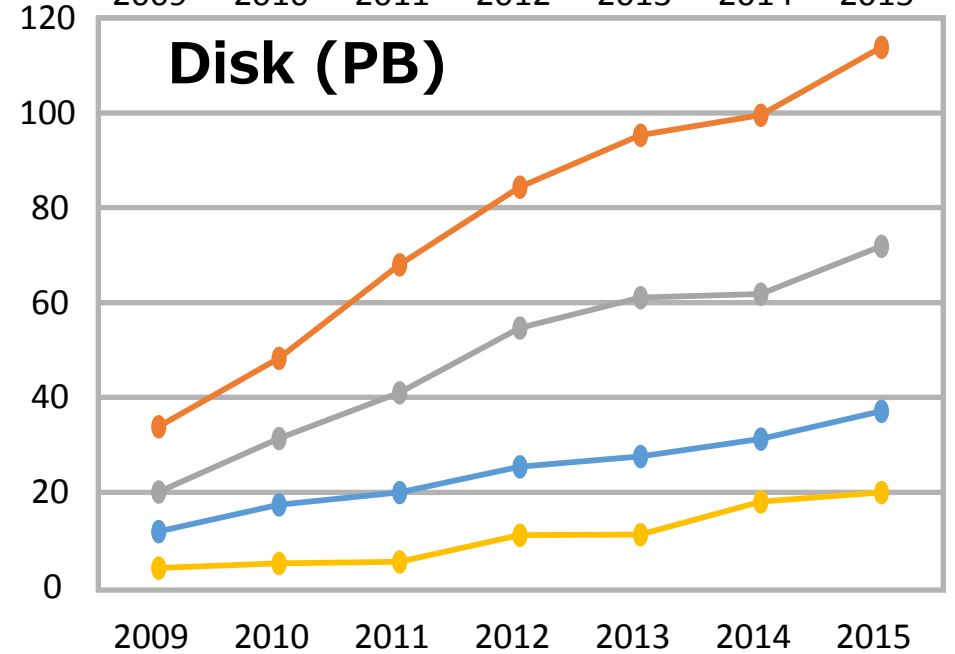
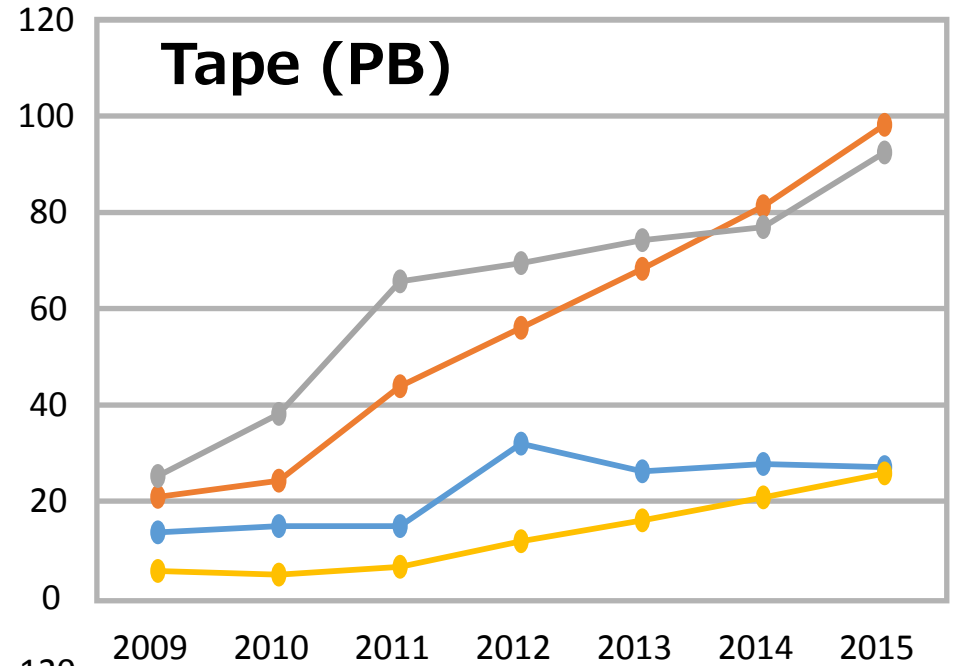
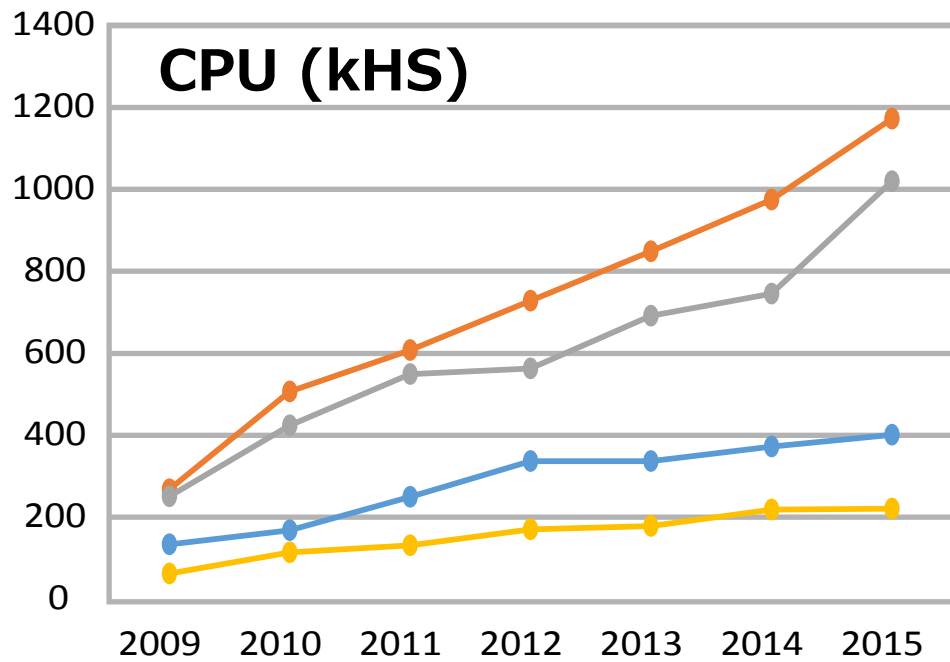
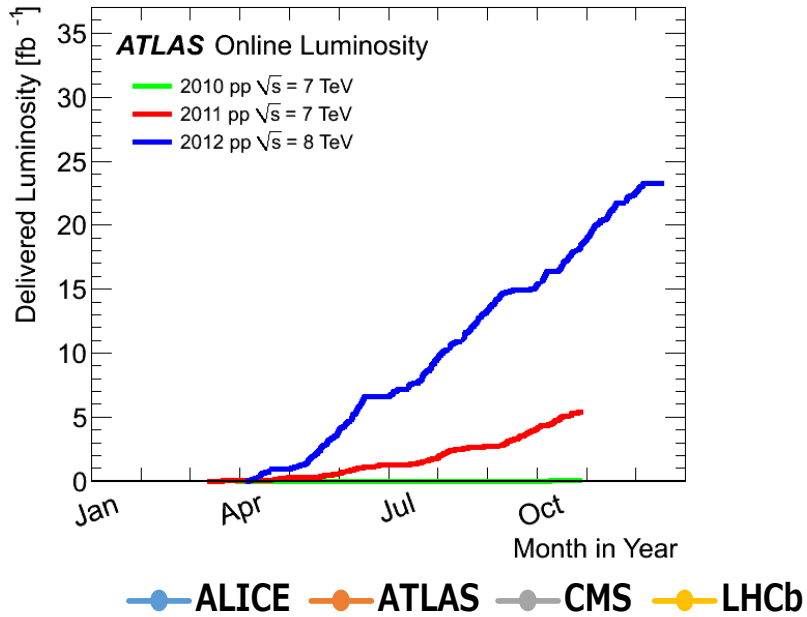
- . many Belle II computing sites overlap with computing sites in LHC experiments .
- . negotiation with each site could be easier under this umbrella ?
- . is it difficult to expand LHCONE to non-LHC experiments ?
- . Configuring another LHCONE-like VRF layer for Belle II could be difficult for some sites ??
- . Belle II traffic shares the same badnwidth with LHC experiments
 - . WAN traffic may be OK ?
 - . traffic pattern is different from LHC (Japan → US/Europe, US → Europe are main)
- . but we do not have any financial support in Belle II.

Under this condition, we want to find a better solution (your comments are highly appreciated)



Spare slides

Resources at LHC experiments

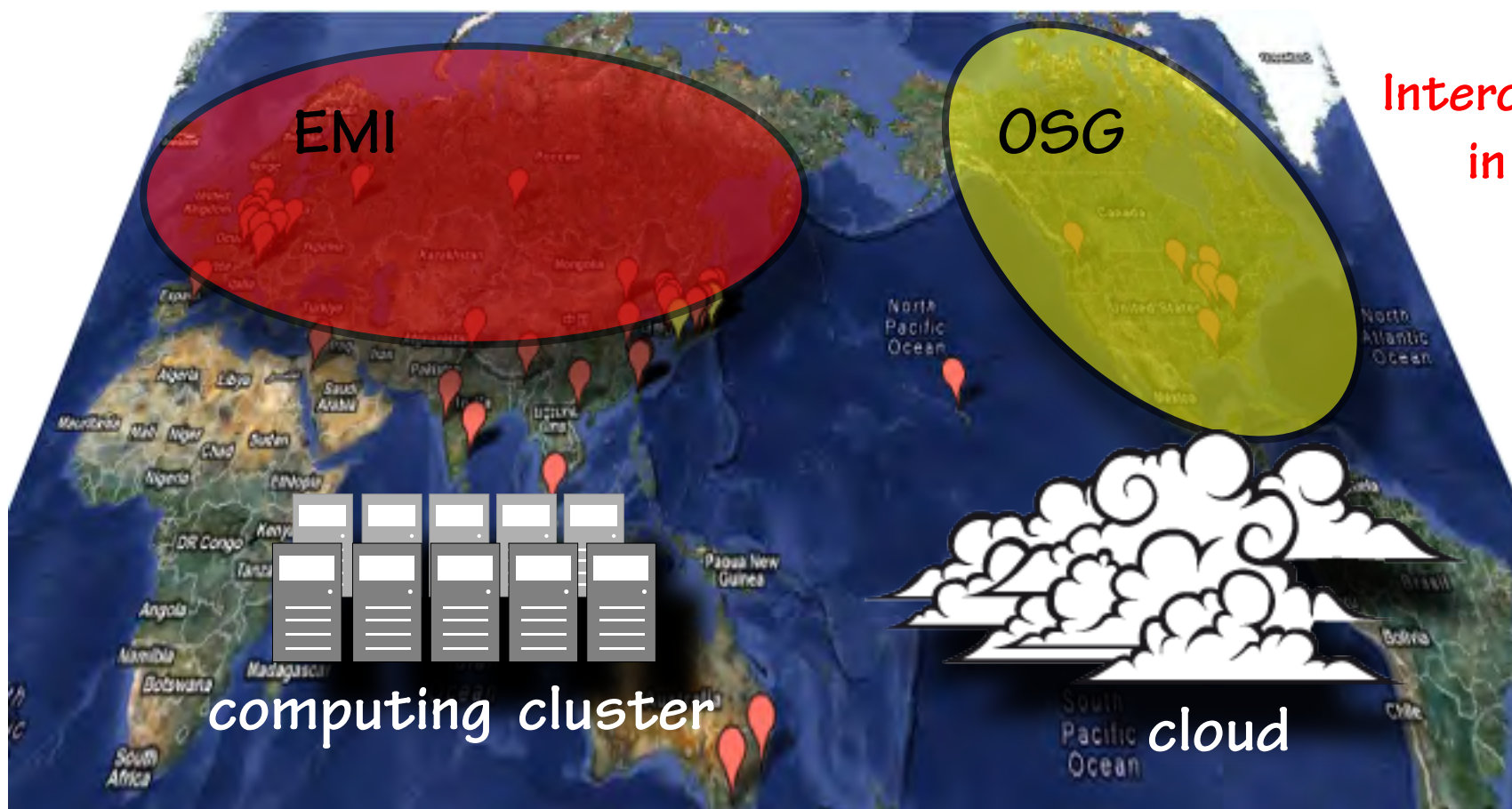


DIRAC

◆ Distributed Infrastructure with Remote Agent Control (developed by LHCb)

→ Pilot jobs

→ Modular structure that enabled it possible to submit jobs to different backends.

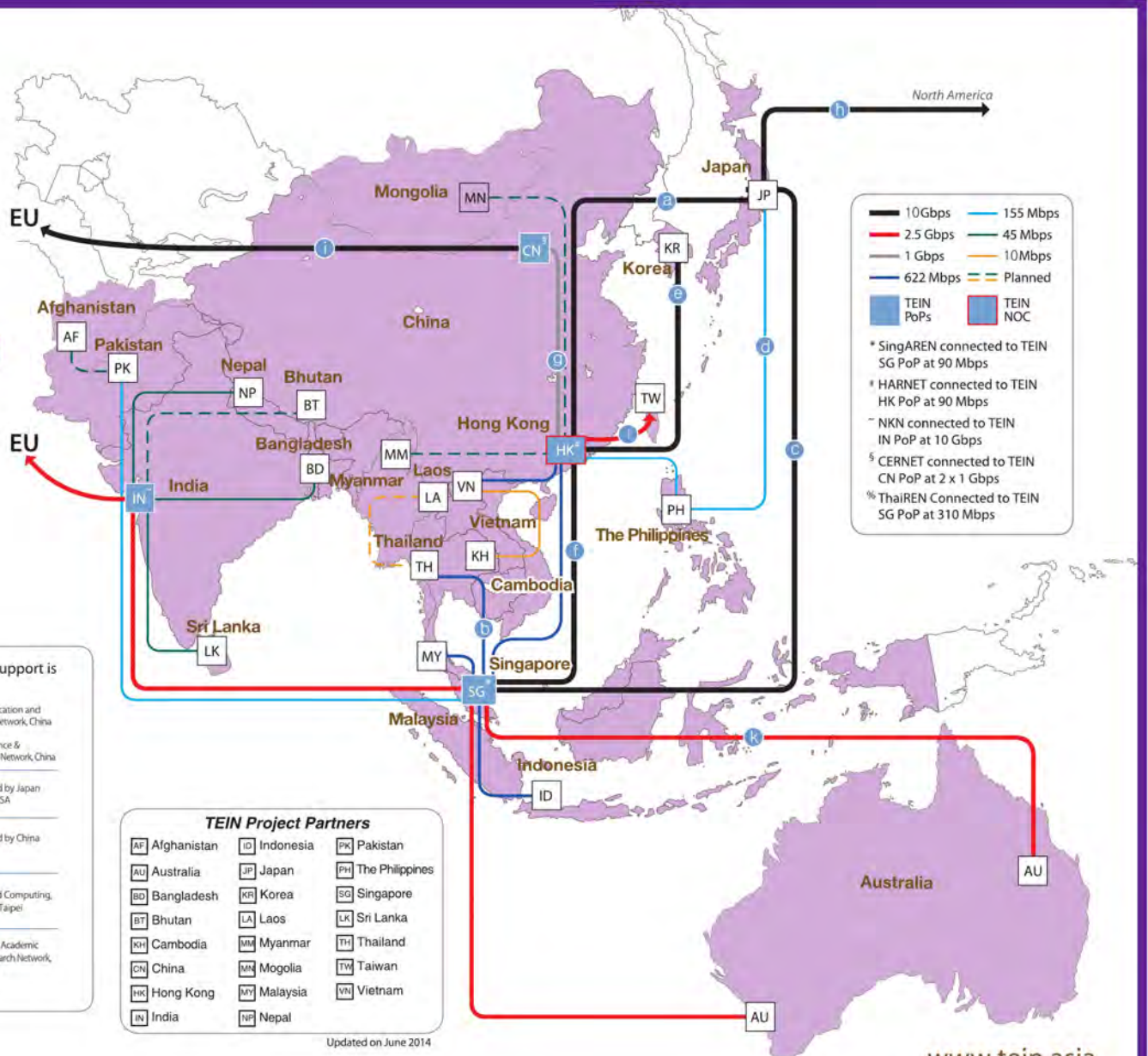


Interoperability
in heterogeneous
computings

Network Connectivity in Asia



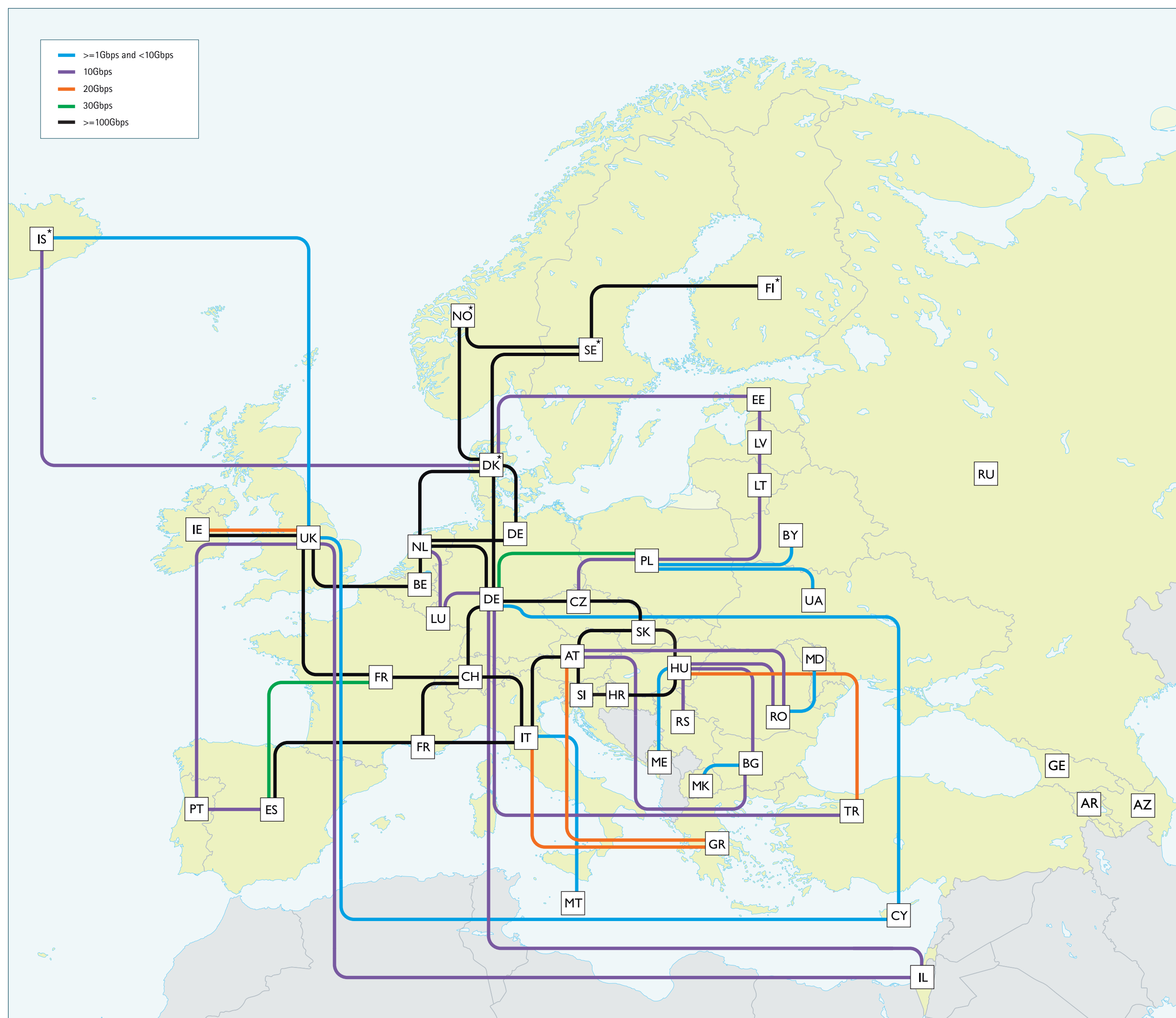
Connecting Asia and Europe's Research and Education Communities



- The following links are fully financed by the link owners whose support is gratefully acknowledged.
- a National Institute of Information and Communications, Japan
 - b National Institute of Information and Communications, Japan
 - c National Institute of Informatics, Japan
 - d Ministry of Agriculture, Forestry and Fisheries Research Network, Japan
 - e National Information Society Agency, South Korea
 - f National Information Society Agency, South Korea
 - g China Education and Research Network, China
 - h co-funded by Japan and the USA
 - i co-funded by China and EU
 - j Academia Sinica Grid Computing, Republic of Chinese Taipei
 - k Australia, Academic and Research Network, Australia

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.



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

AR Armenia	BG Bulgaria	DE Germany	FI Finland*	HR Croatia	IS Iceland*	LV Latvia	NL Netherlands	RO Romania	SE Sweden*	TR Turkey
AT Austria	CH Switzerland	DK Denmark*	FR France	HU Hungary	IT Italy	ME Montenegro	NO Norway*	RS Serbia	SI Slovenia	UK United Kingdom
AZ Azerbaijan	CY Cyprus	EE Estonia	GE Georgia	IE Ireland	LT Lithuania	MK F.Y.R. Macedonia	PL Poland	BY Belarus	MD Moldova	UA Ukraine
BE Belgium	CZ Czech Republic	ES Spain	GR Greece	IL Israel	LU Luxembourg	MT Malta	PT Portugal	RU Russia	SK Slovakia	

*Connections between these countries are part of NORDUnet (the Nordic regional network)