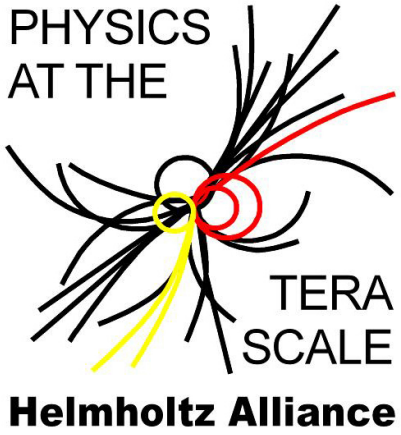
Networks for HEP data analysis - The Global Picture



Slides by: Volker Gülzow, Kars Ohrenberg - DESY
Presented by: Yves Kemp, Martin Gasthuber - DESY

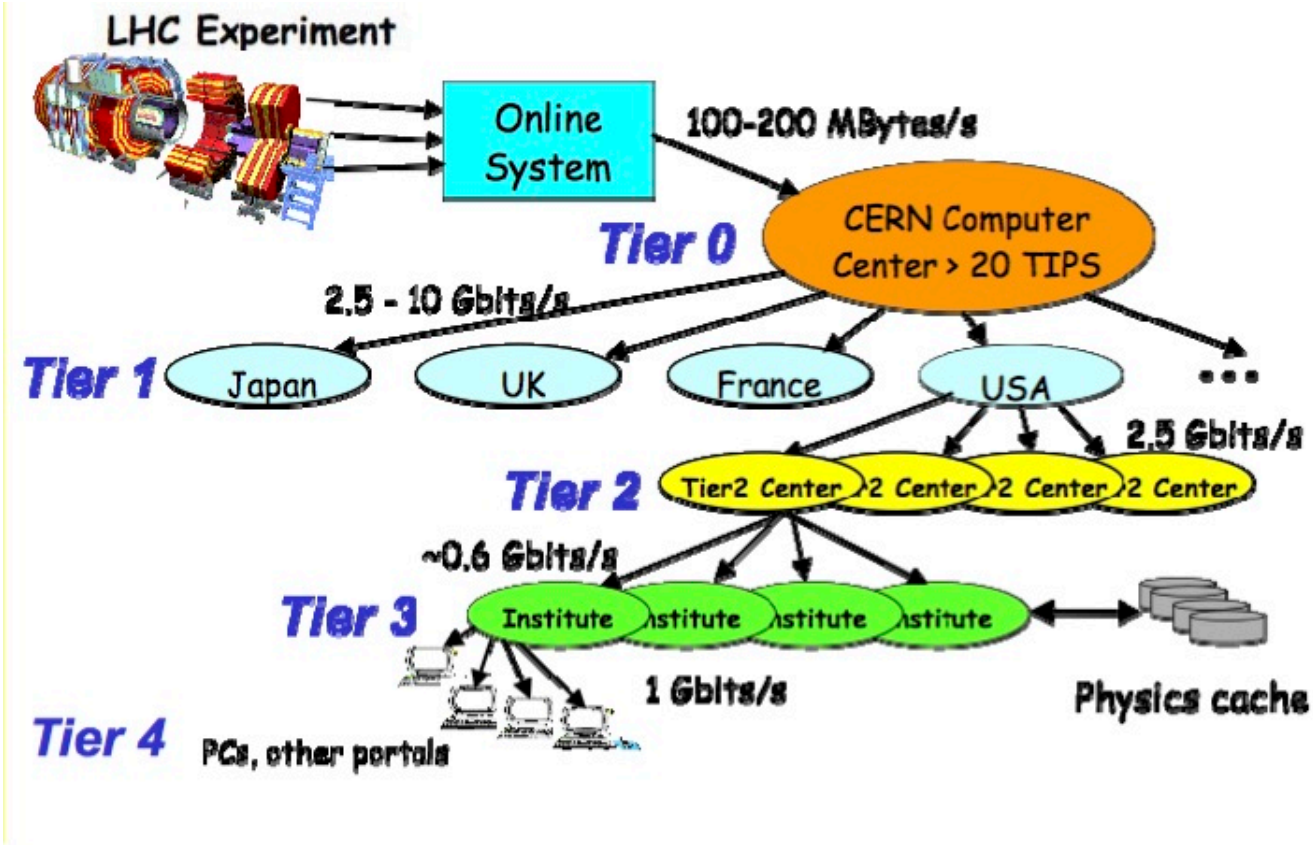
7th annual Terascale Alliance Workshop
KIT 3.12.2013

HELMHOLTZ GEMEINSCHAFT

DESY

LHC Computing Infrastructure

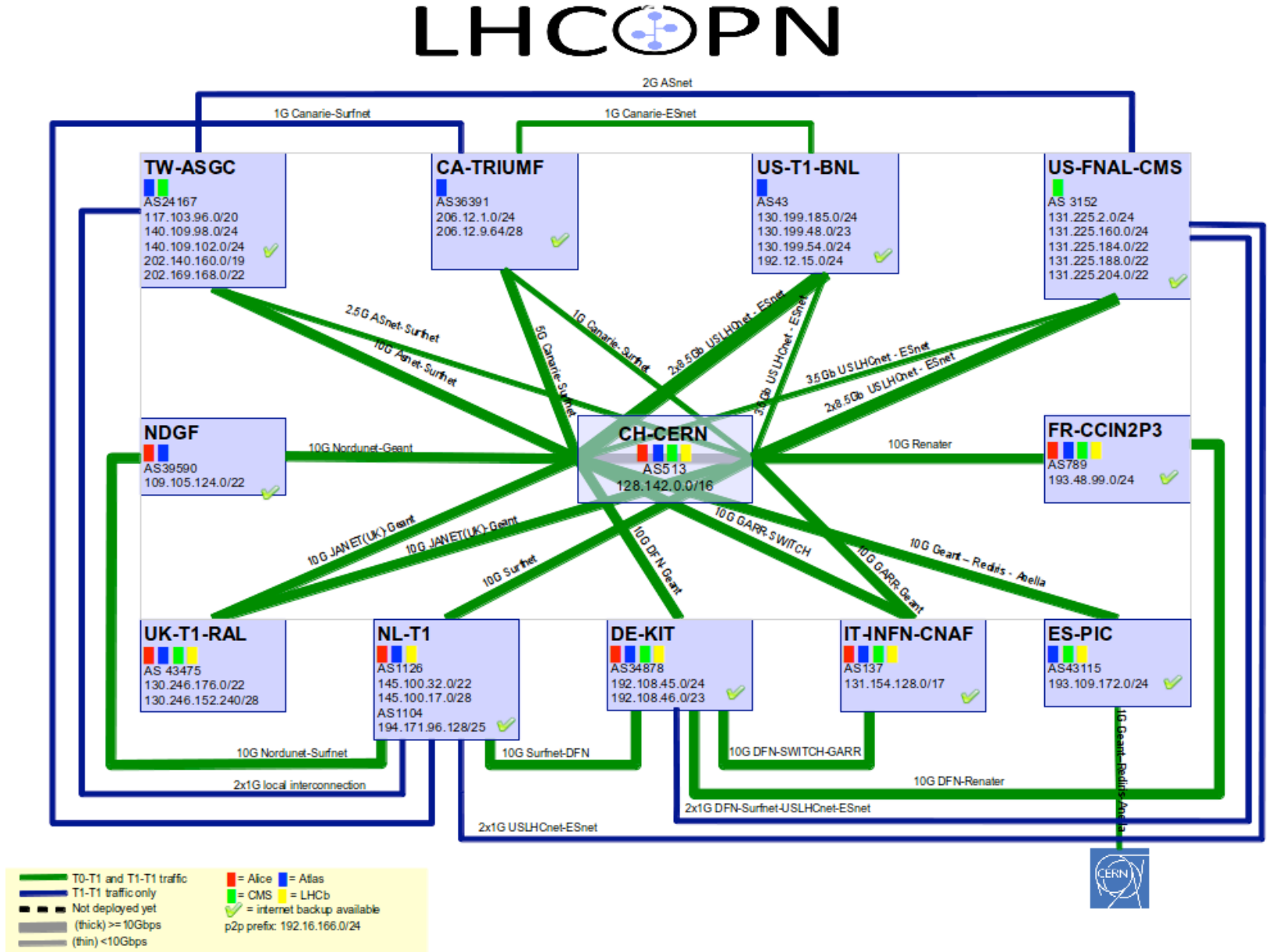
- WLCG in brief:
  - 1 Tier-0, 11 Tier-1s, ~ 140 Tier-2s, O(300) Tier-3s worldwide



The LHC Optical Private Network

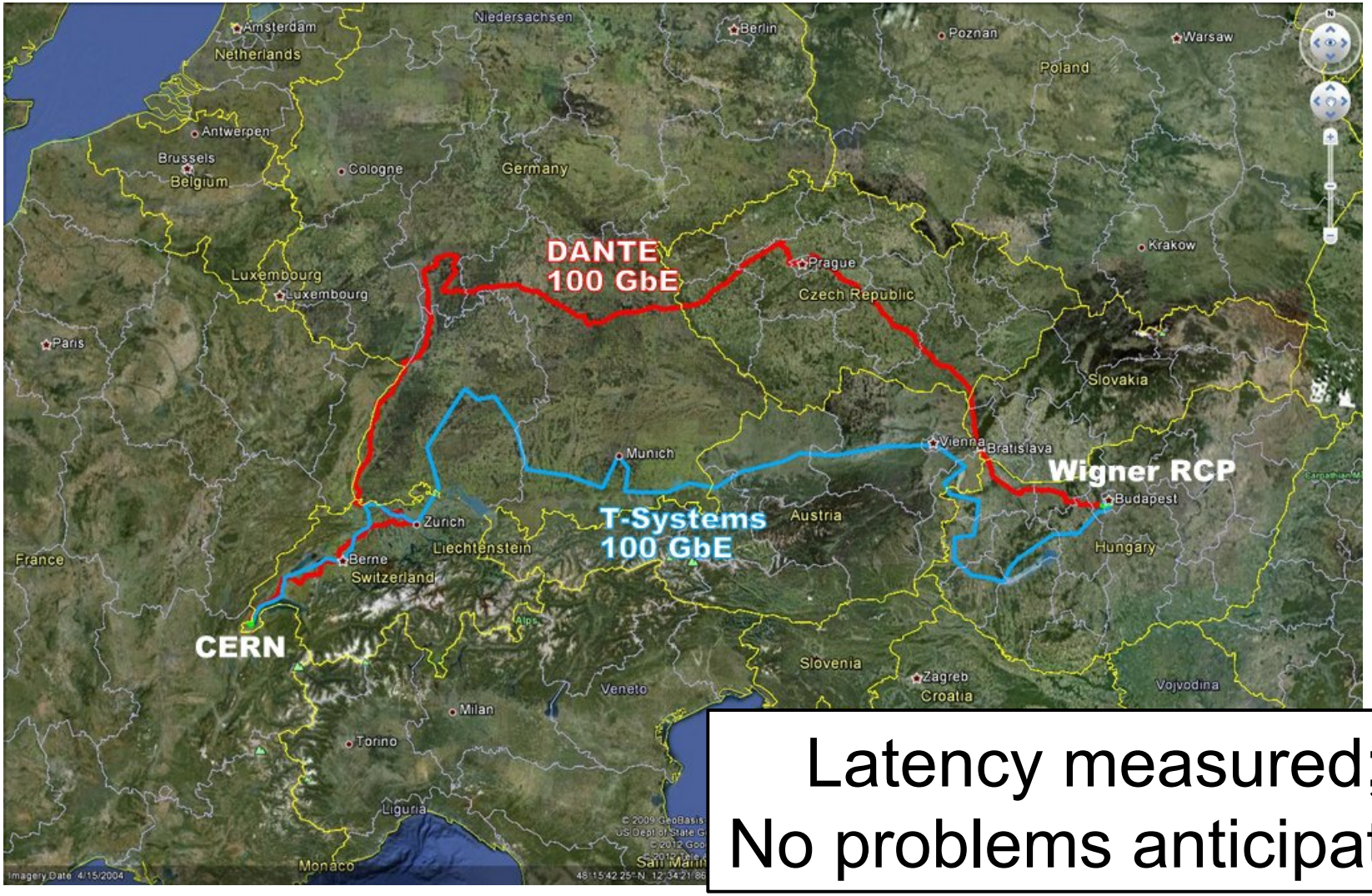
- The LHCOPN (from http://lhcopn.web.cern.ch)
  - The LHCOPN is the private IP network that connects the Tier0 and the Tier1 sites of the LCG.
  - The LHCOPN consists of any T0-T1 or T1-T1 link which is dedicated to the transport of WLCG traffic and whose utilization is restricted to the Tier0 and the Tier1s.
  - Any other T0-T1 or T1-T1 link not dedicated to WLCG traffic may be part of the LHCOPN, assuming the exception is communicated to and agreed by the LHCOPN community
- Very closed and restricted access policy
- No Gateways

LHCOPN Network Map



Connectivity (100 Gb/s)

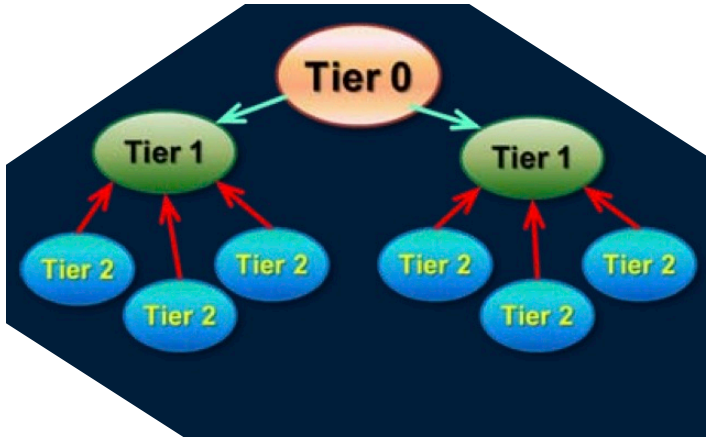
Slide by Ian Bird

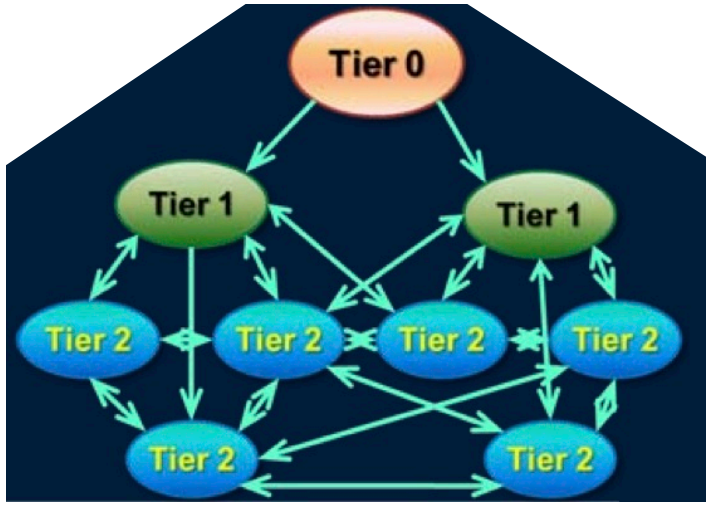


Latency measured;
No problems anticipated

Computing Models Evolution

- The original MONARC model was strictly hierarchical
- Changes introduced gradually since 2010
- Main evolutions:
  - Meshed data flows: Any site can use any other site as source of data
  - Dynamic data caching: Analysis sites pull datasets from other sites „on demand“, including from Tier-2s in other regions
  - Remote data access
- Variations by experiment
- LHCOPN only connects T0 and T1

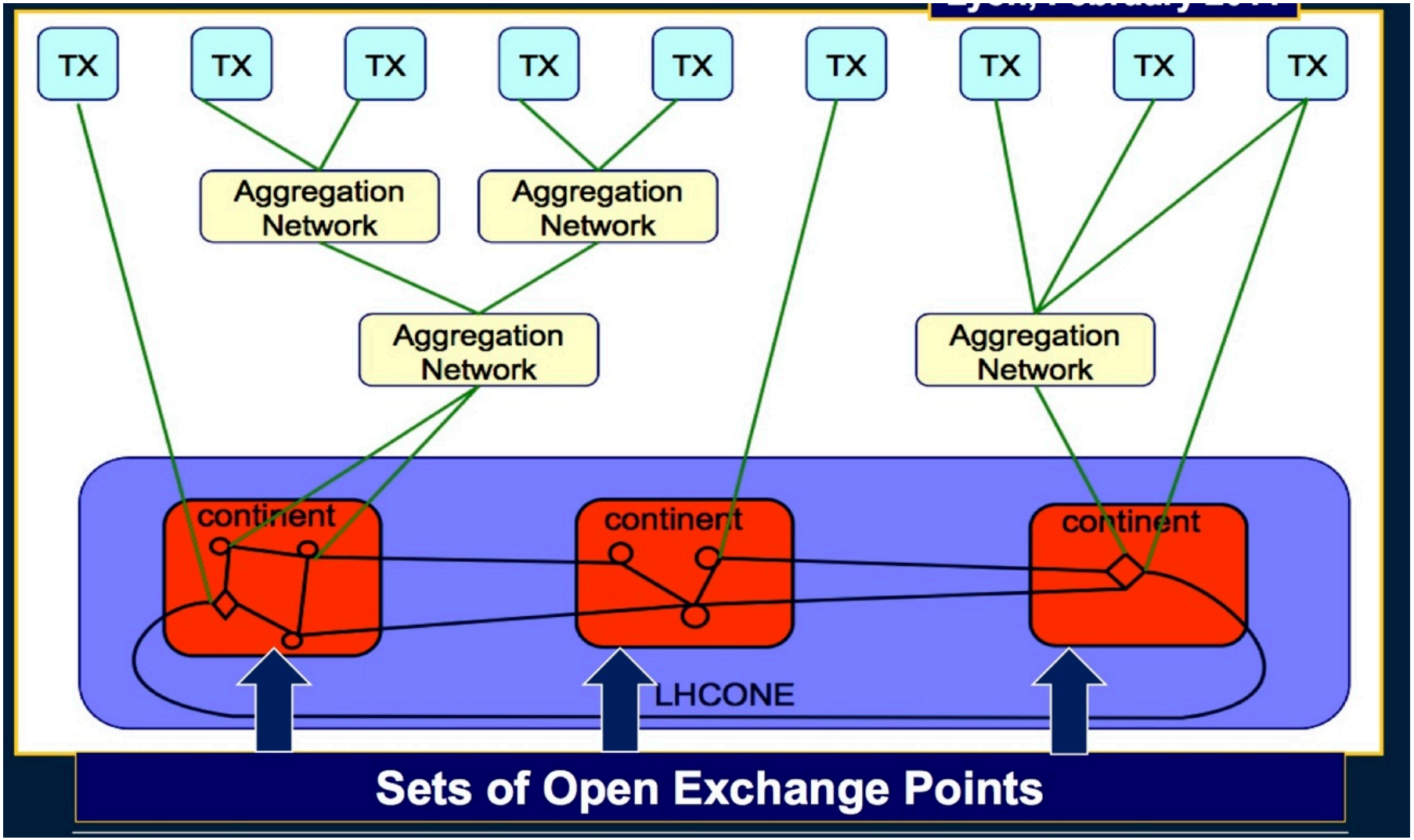




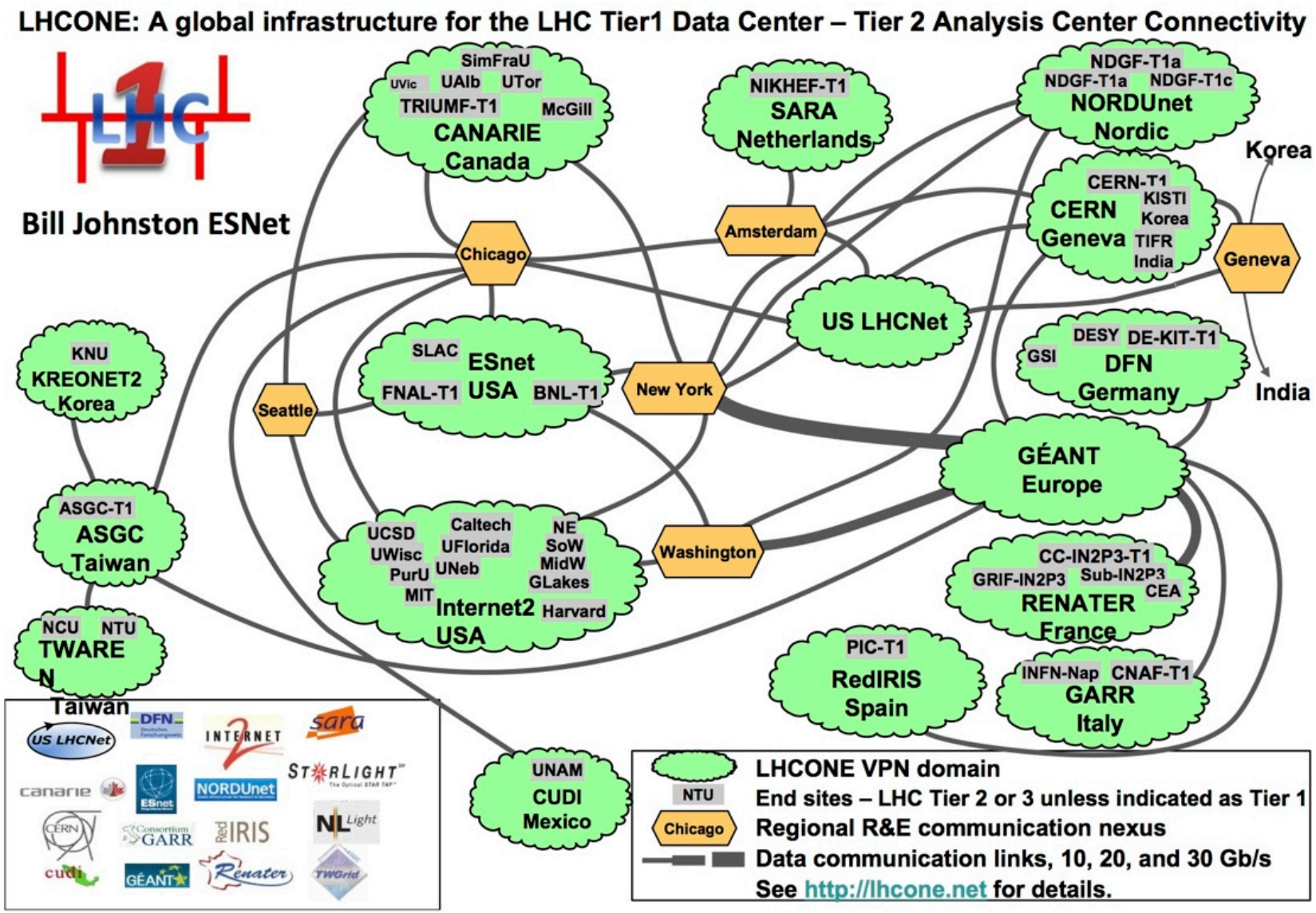
LHC Open Network Environment

- With the successful operation of the LHC accelerator and the start of the data analysis, there has come a re-evaluation of the computing and data models of the experiments
- The goal of LHCONE (LHC Open Network Environment) is to ensure better access to the most important datasets by the worldwide HEP community
- Traffic patterns have altered to the extent that substantial data transfers between major sites are regularly being observed on the General Purpose Networks (GPN)
- The main principle is to separate the LHC traffic from the GPN traffic, thus avoiding degraded performance
- The objective of LHCONE is to provide entry points into a network that is private to the LHC T1/2/3 sites.
- LHCONE is not intended to replace LHCOPN but rather to complement it

LHCONE Achitecture



LHCONE: A global Infrastructure



LHCONE Routing Policies

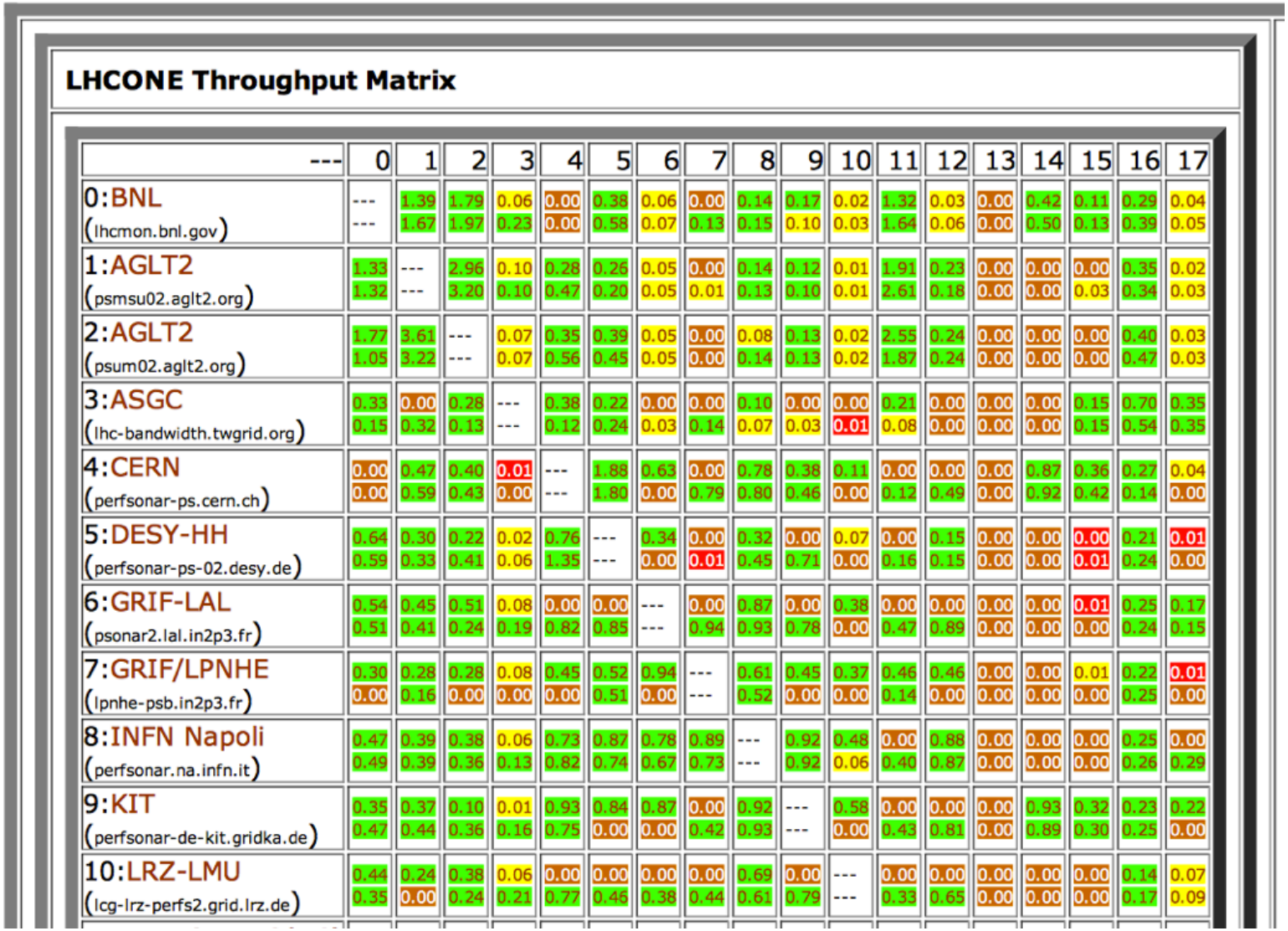
- Only the networks which are announced to LHCONE are allowed to reach the LHCONE
- Only these networks will be reachable via the LHCONE
- Other traffic uses the public, general purpose networks
- Asymmetric routing should be avoided as this will cause problems for traffic passing (public) firewalls

LHCONE - the current status

- Currently 126 network prefixes (29.11.2013)
  - German sites currently participating in LHCONE
    - DESY, KIT, GSI, RWTH Aachen, Uni Wuppertal
  - Europe
    - CERN, SARA , GRIF (LAL + LPNHE), INFN, FZU, PIC, ...
  - US:
    - AGLT2 (MSU + UM), MWT2 (UC), BNL, ...
  - Canada
    - TRIUMF, Toronto, ...
  - Asia
    - ASGC, ICEPP, ...
- Detailed monitoring via perfSONAR

LHCONE Monitoring

d (old)
t)
d (old,test)
ation



LHCONE Throughput Matrix

---	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0:BNL (lhcmon.bnl.gov)	--- ---	1.39 1.67	1.79 1.97	0.06 0.23	0.00 0.00	0.38 0.58	0.06 0.07	0.00 0.13	0.14 0.15	0.17 0.10	0.02 0.03	1.32 1.64	0.03 0.06	0.00 0.00	0.42 0.50	0.11 0.13	0.29 0.39	0.04 0.05
1:AGLT2 (psmsu02.aglt2.org)	1.33 1.32	--- ---	2.96 3.20	0.10 0.10	0.28 0.47	0.26 0.20	0.05 0.05	0.00 0.01	0.14 0.13	0.12 0.10	0.01 0.01	1.91 2.61	0.23 0.18	0.00 0.00	0.00 0.00	0.00 0.03	0.35 0.34	0.02 0.03
2:AGLT2 (psum02.aglt2.org)	1.77 1.05	3.61 3.22	--- ---	0.07 0.07	0.35 0.56	0.39 0.45	0.05 0.05	0.00 0.00	0.08 0.14	0.13 0.13	0.02 0.02	2.55 1.87	0.24 0.24	0.00 0.00	0.00 0.00	0.00 0.00	0.40 0.47	0.03 0.03
3:ASGC (lhc-bandwidth.twgrid.org)	0.33 0.15	0.00 0.32	0.28 0.13	--- ---	0.38 0.12	0.22 0.24	0.00 0.03	0.00 0.14	0.10 0.07	0.00 0.03	0.00 0.01	0.21 0.08	0.00 0.00	0.00 0.00	0.00 0.00	0.15 0.15	0.70 0.54	0.35 0.35
4:CERN (perfsonar-ps.cern.ch)	0.00 0.00	0.47 0.59	0.40 0.43	0.01 0.00	--- ---	1.88 1.80	0.63 0.00	0.00 0.79	0.78 0.80	0.38 0.46	0.11 0.00	0.00 0.12	0.00 0.49	0.00 0.00	0.87 0.92	0.36 0.42	0.27 0.14	0.04 0.00
5:DESY-HH (perfsonar-ps-02.desy.de)	0.64 0.59	0.30 0.33	0.22 0.41	0.02 0.06	0.76 1.35	--- ---	0.34 0.00	0.00 0.01	0.32 0.45	0.00 0.71	0.07 0.00	0.00 0.16	0.15 0.15	0.00 0.00	0.00 0.00	0.00 0.01	0.21 0.24	0.01 0.00
6:GRIF-LAL (psonar2.lal.in2p3.fr)	0.54 0.51	0.45 0.41	0.51 0.24	0.08 0.19	0.00 0.82	0.00 0.85	--- ---	0.00 0.94	0.87 0.93	0.00 0.78	0.38 0.00	0.00 0.47	0.00 0.89	0.00 0.00	0.00 0.00	0.01 0.00	0.25 0.24	0.17 0.15
7:GRIF/LPNHE (lpnhe-psb.in2p3.fr)	0.30 0.00	0.28 0.16	0.28 0.00	0.08 0.00	0.45 0.00	0.52 0.51	0.94 0.00	--- ---	0.61 0.52	0.45 0.00	0.37 0.00	0.46 0.14	0.46 0.00	0.00 0.00	0.00 0.00	0.01 0.00	0.22 0.25	0.01 0.00
8:INFN Napoli (perfsonar.na.infn.it)	0.47 0.49	0.39 0.39	0.38 0.36	0.06 0.13	0.73 0.82	0.87 0.74	0.78 0.67	0.89 0.73	--- ---	0.92 0.92	0.48 0.06	0.00 0.40	0.88 0.87	0.00 0.00	0.00 0.00	0.00 0.00	0.25 0.26	0.00 0.29
9:KIT (perfsonar-de-kit.gridka.de)	0.35 0.47	0.37 0.44	0.10 0.36	0.01 0.16	0.93 0.75	0.84 0.00	0.87 0.00	0.00 0.42	0.92 0.93	--- ---	0.58 0.00	0.00 0.43	0.00 0.81	0.00 0.00	0.93 0.89	0.32 0.30	0.23 0.25	0.22 0.00
10:LRZ-LMU (lcg-lrz-perfs2.grid.lrz.de)	0.44 0.35	0.24 0.00	0.38 0.24	0.06 0.21	0.00 0.77	0.00 0.46	0.00 0.38	0.00 0.44	0.69 0.61	0.00 0.79	--- ---	0.00 0.33	0.00 0.65	0.00 0.00	0.00 0.00	0.00 0.00	0.14 0.17	0.07 0.09

DESY

Software-Defined Networking (SDN)

- … is a research activity within LHCone
- Is a form of network virtualization in which the control plane is separated from the data plane and implemented in a software application
- This architecture allows network administrators to have programmable central control of network traffic without requiring physical access to the network's hardware devices
- SDN requires some method for the control plane to communicate with the data plane. One such mechanism is OpenFlow which is a standard interface for controlling computer networking switches

Bandwidth Evolution @ DFN

- DFN has upgraded the optical platform of the X-WiN
  - Contract awarded to ECI Telecom (http://www.ecitele.com)
  - Migration work is finished
- High Bandwidth Capabilities
  - 88 wave length per fiber
  - Up to 100 Gbps per wave length
    - thus 8.8 Tbps per fiber!
  - 1 Tbps Switching Fabric (aggregation of 10 Gbps lines on single 100 Gbps line)

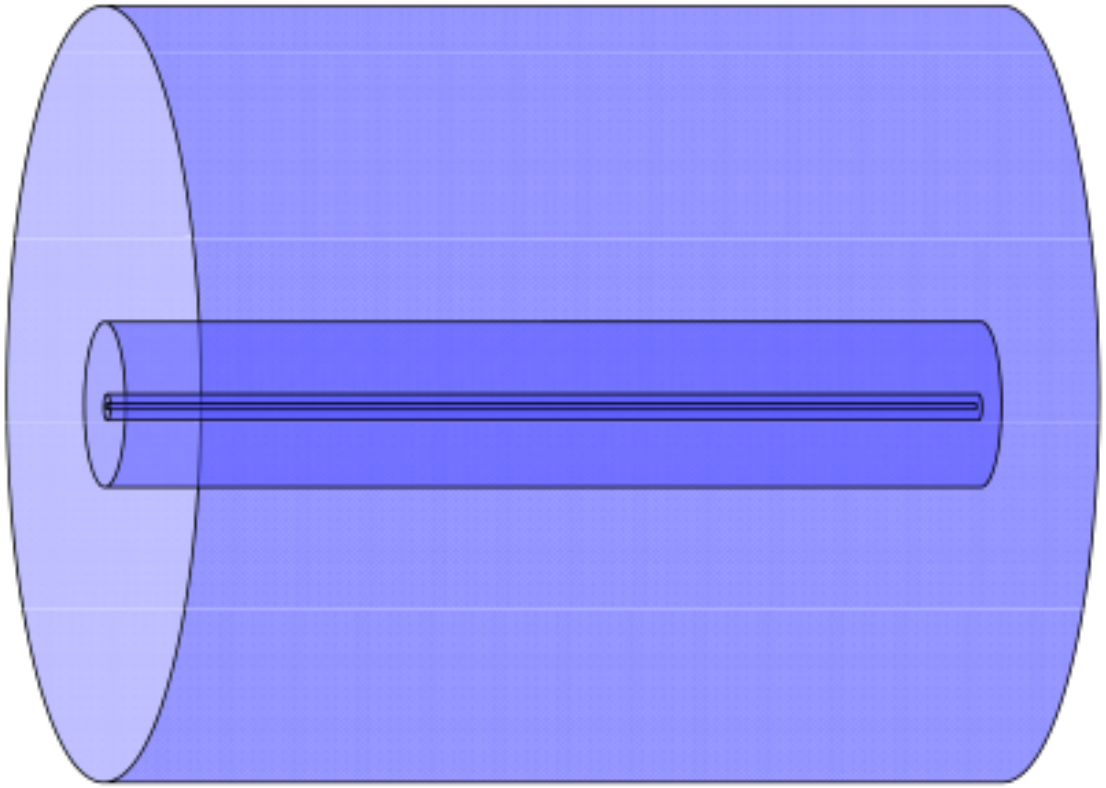
Growing WiN Capacities

X-WiN 2012: 8.800 Gbit/s

X-WiN 2006: 400 Gbit/s

G-WiN: 10 Gbit/s

B-WiN: 622 Mbit/s



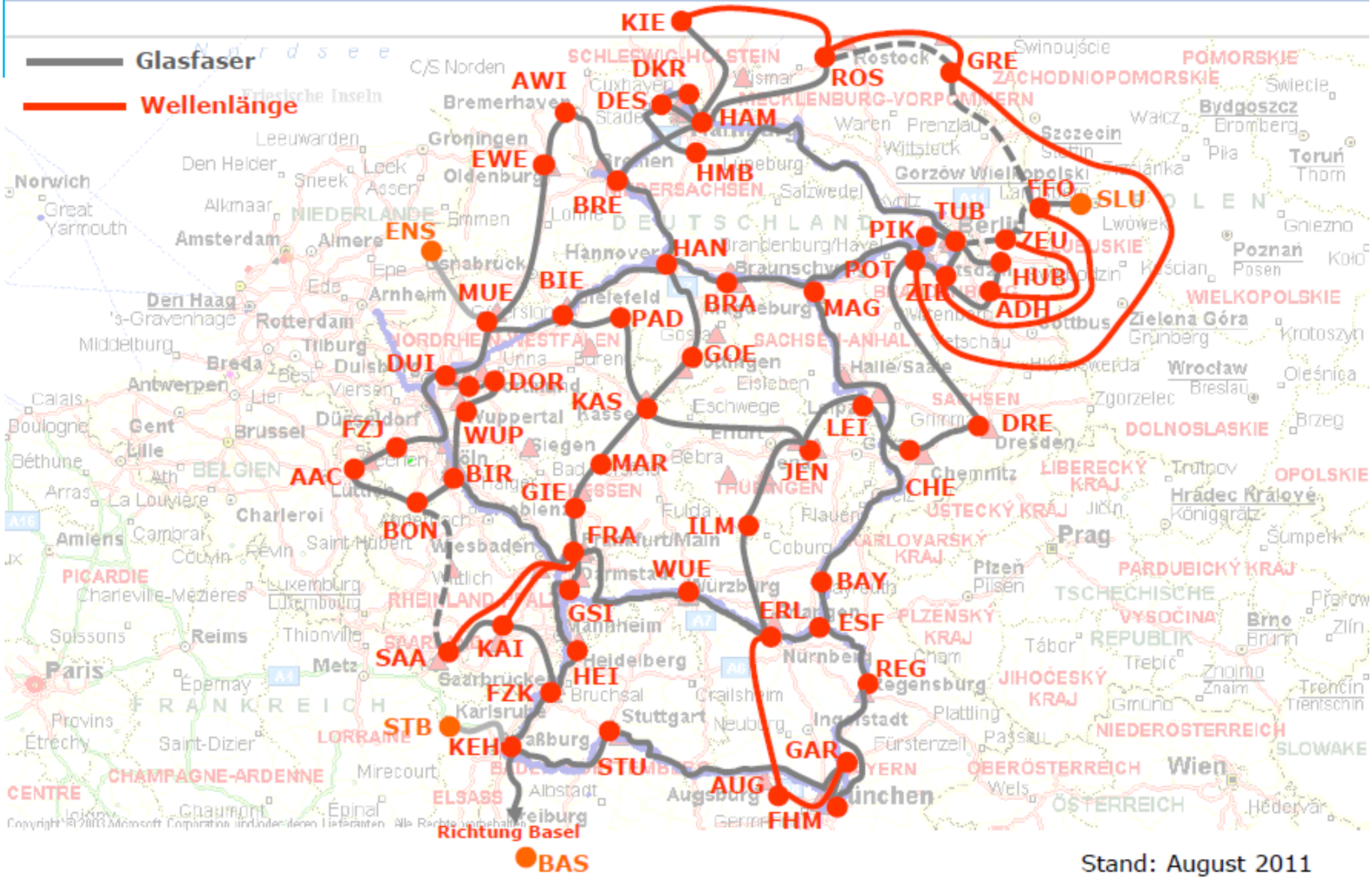
Summary

- The LHC computing and data models continue to evolve towards more dynamic, less structured, on-demand data movement thus requiring different network structures
- With the evolution of the new optical platforms bandwidth will get more affordable

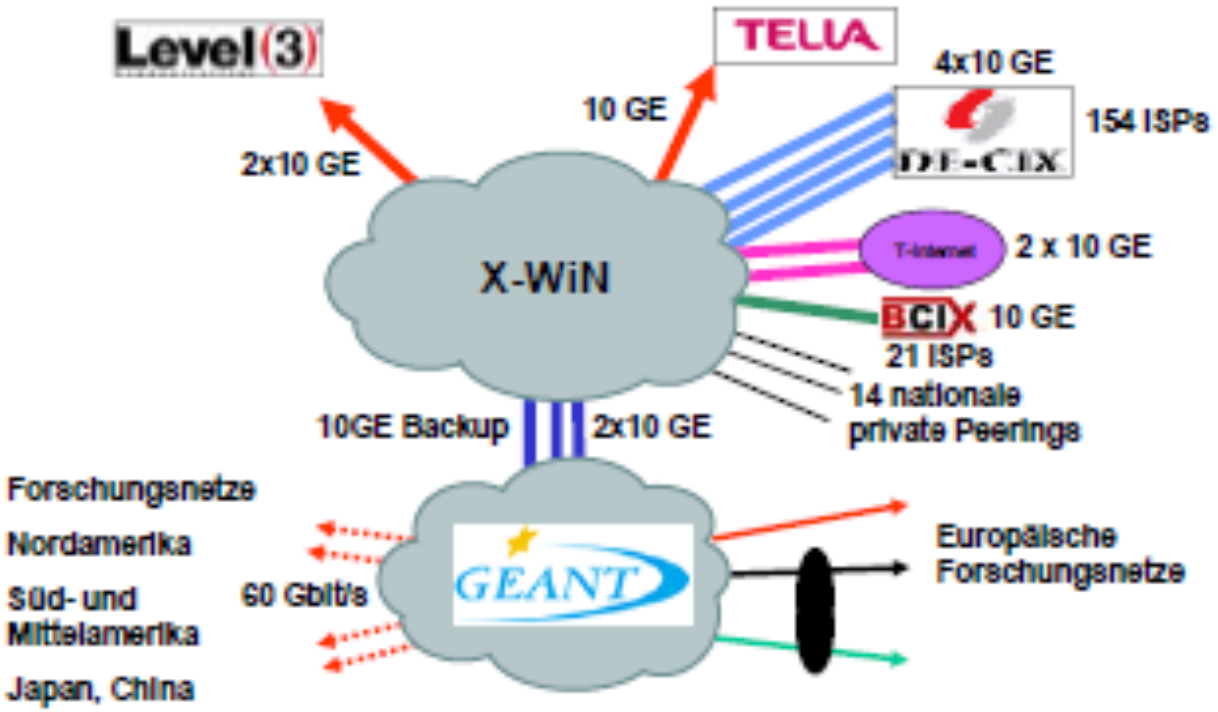
> Backup slides

Topologie X-WiN

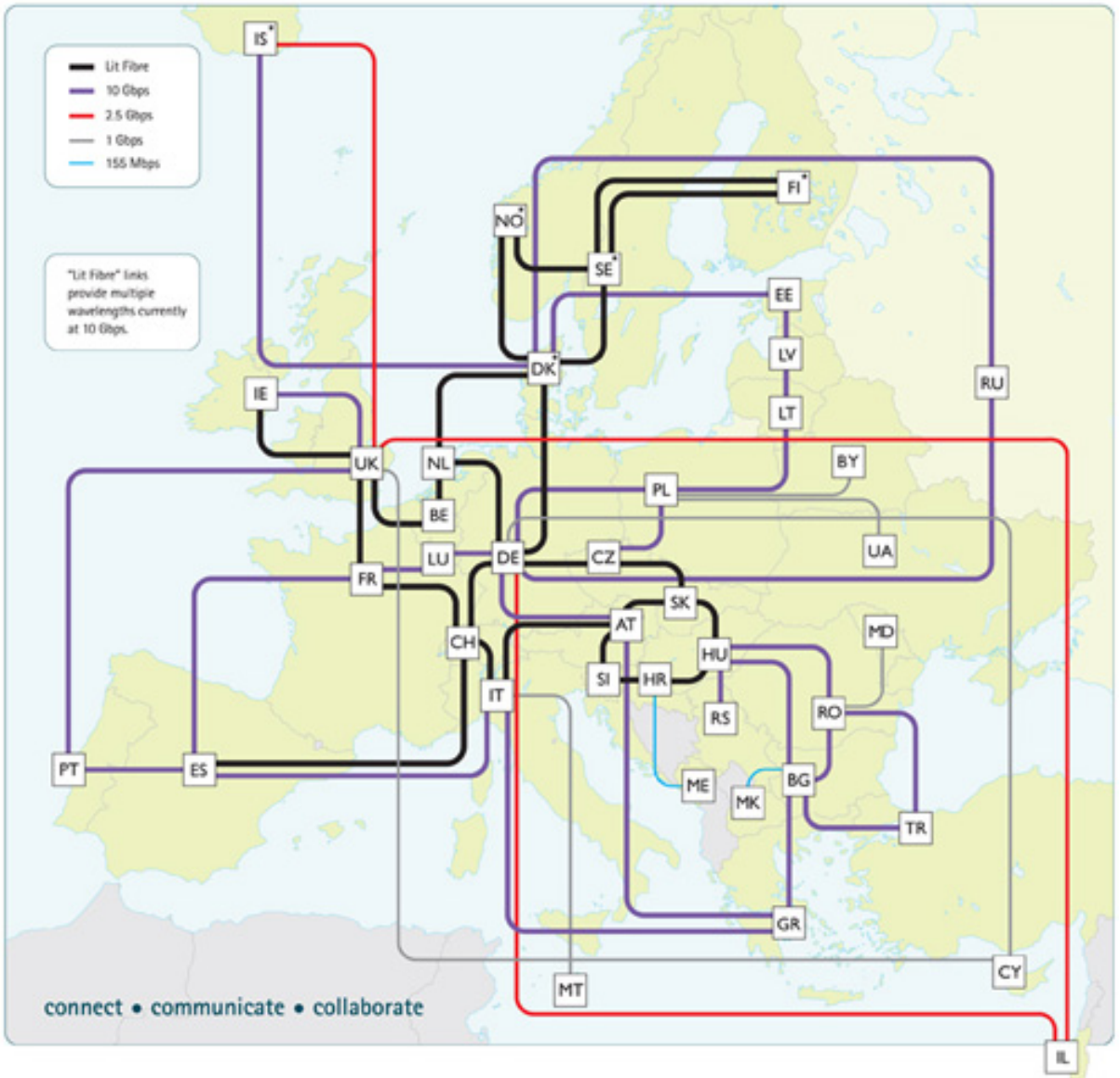
DFN Deutsches Forschungsnetz



Außenanbindungen X-WiN 2012



Geant3 topology



WAN + LHCONE Infrastructure at DESY