CERN Datacentre Network

WLCG pre-GDB workshop
data-centre network architectures

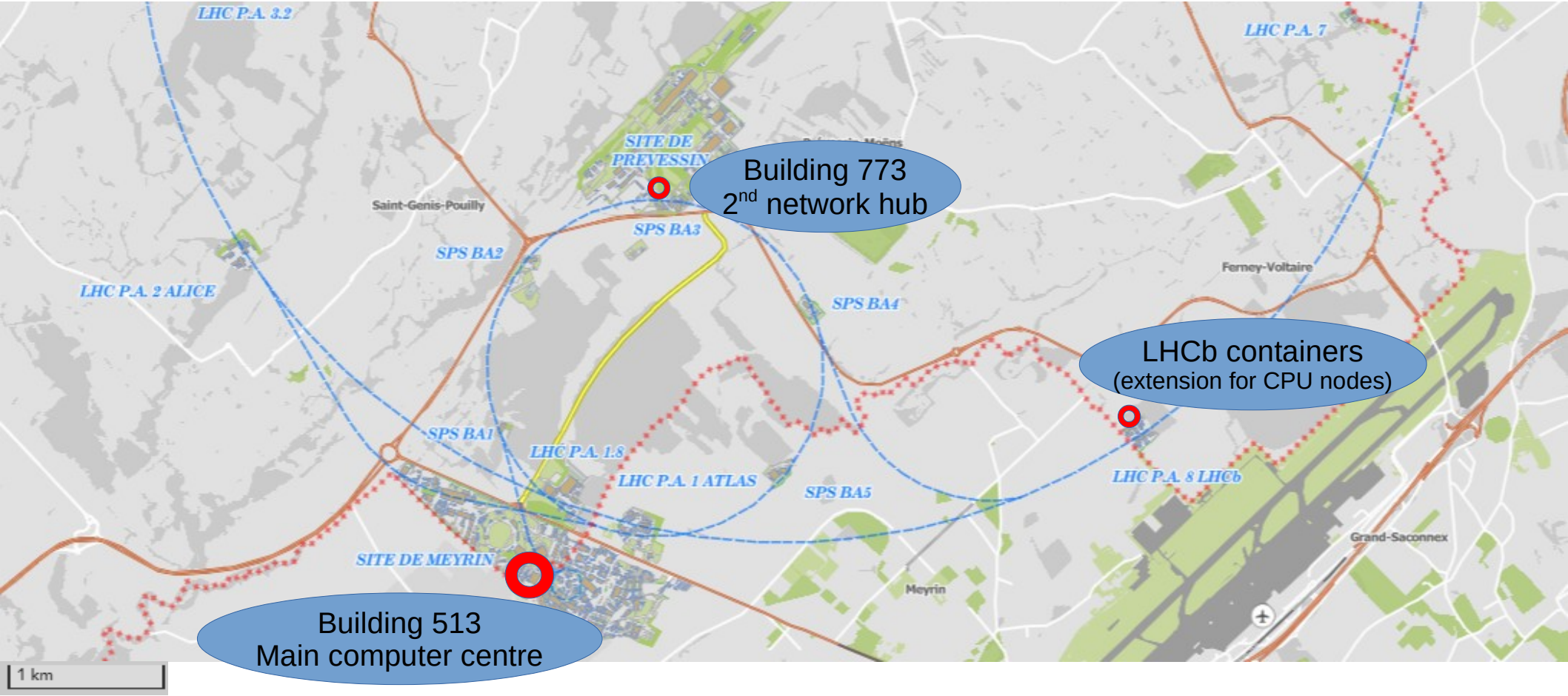
Vincent DUCRET – CERN IT/CS-CE

June 7th 2021

Summary

- **CERN Datacentre building(s)**
- **DC Network design:**
  - A few figures
  - General overview
  - Management network
  - VRFs
  - ToR switches
  - Routers
  - Routing protocols
- **Integration with other network services**
- **Plans for the future**

CERN Datacentre Buildings

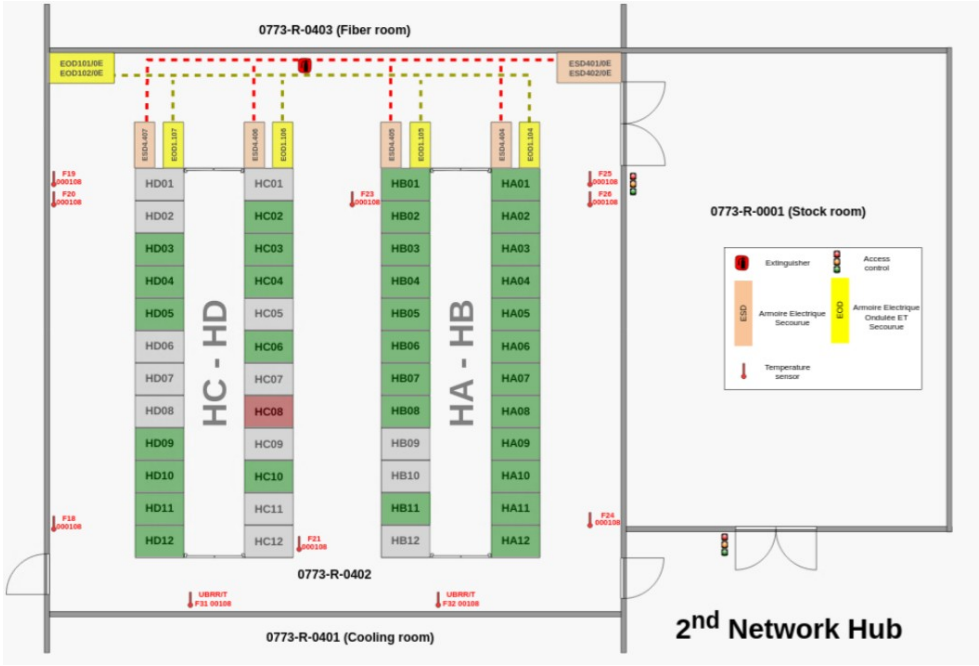


CERN Datacentre Buildings – LHCb container

- 2 containers temporarily lent by LHCb to CERN IT
- 24 racks per containers
- Used as an extension of the main computer centre
- Used for CPU nodes

CERN Datacentre Buildings – Blg 773 2nd Network Hub

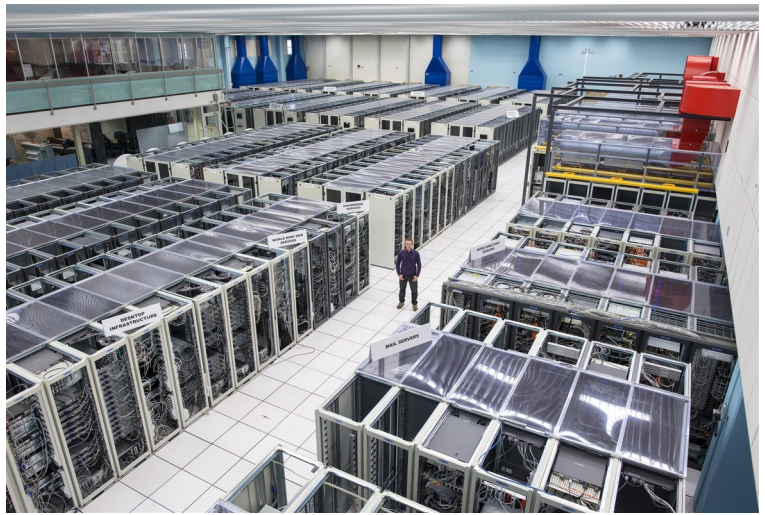
- **Built on Prévessin site to provide network redundancy**
- **Main purpose is to host network devices**
- **Provides redundancy for different part of the CERN network**
  - Backbone routers for Datacentre, TN, Campus, Wi-Fi, etc…
  - Public internet access
  - LHCOPN/ONE access (GEANT, NORDUnet, etc...)
- **Host a few servers for service redundancy (DHCP, DNS, Radius, one DB, few EOS servers…) but it is not intended to provide full business continuity for all services**



CERN Datacentre Buildings – Blg 513

- **Main computer centre**
- **3 separate rooms**
  - Main room (power partially protected by diesel)
  - “Vault”
  - “Barn” (power fully protected by diesel)

Host almost all the Computer Centre services



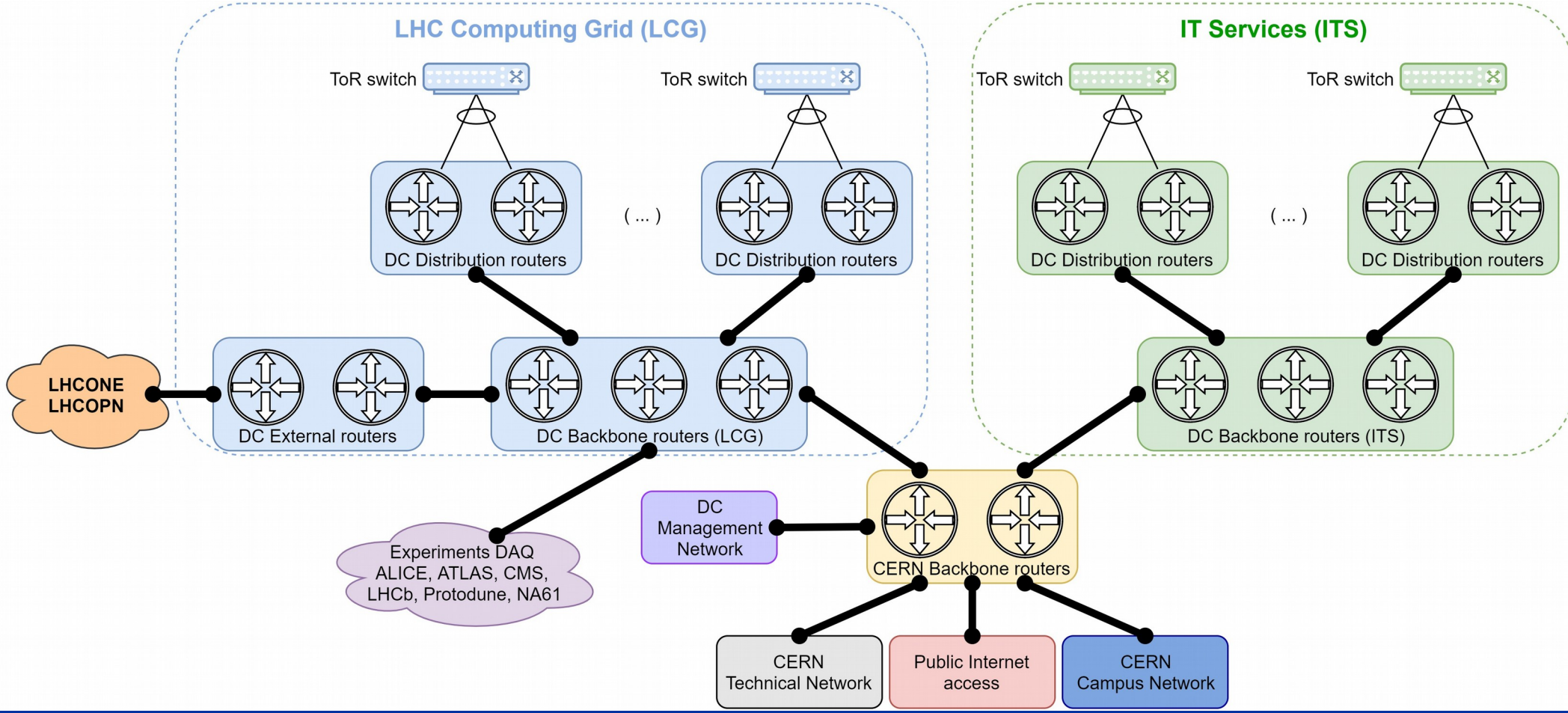
DC Network design – a few figures

- **500+ ToR switches**

- **17 routers**

- **Current “Backbone” total used capacity = 20.8 Tbps**

DC Network design – general overview



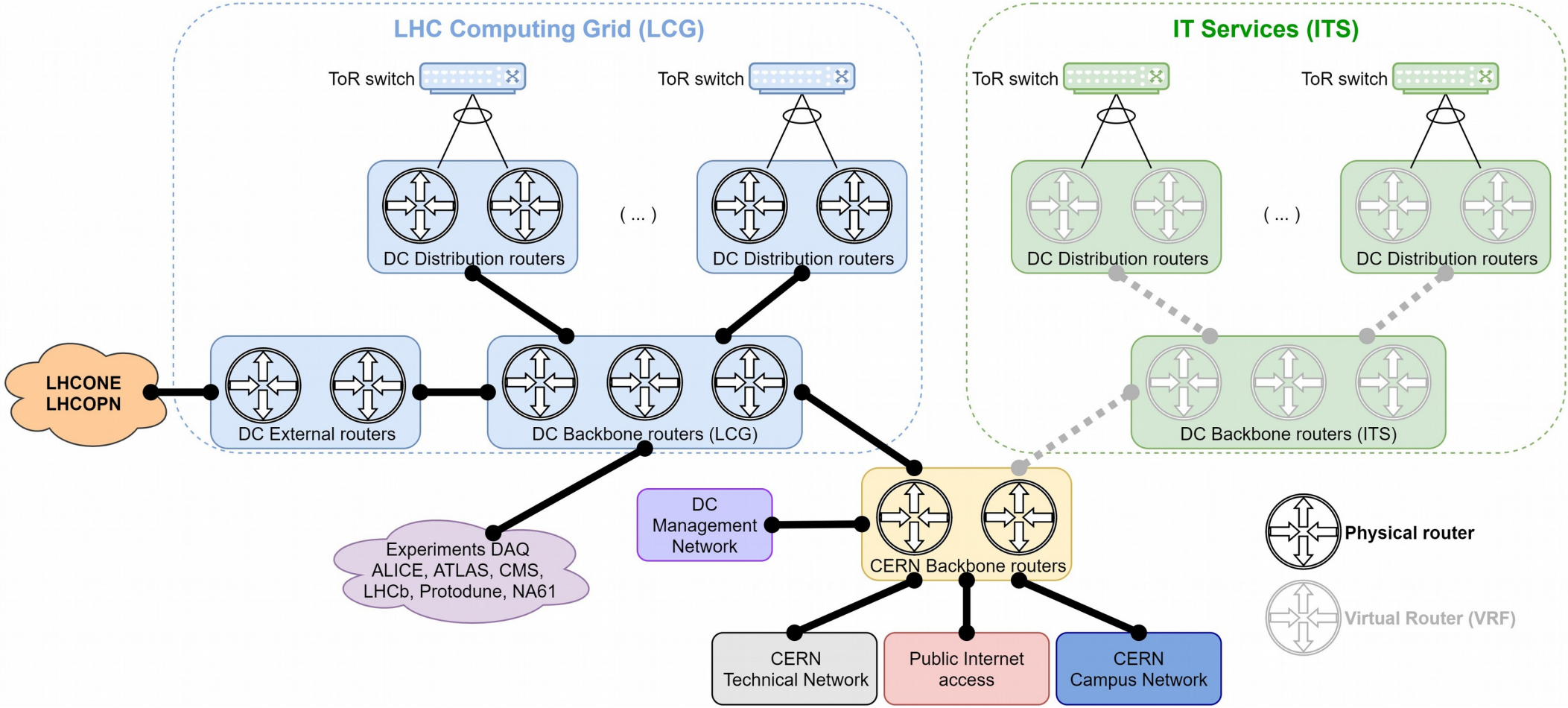
DC Network design – Management Network

- **Separate network (dedicated routers and switches)**
- **Basic setup (no redundancy)**
- **1G copper switches**
- **Used for servers IPMI interfaces, and manageable PDUs**
- **Private network**
- **Protected via ACLs**

DC Network design – VRFs

- **Same physical routers for LCG and ITS domains**
- **LCG = default VRF ITS = virtual routers**
- **Based on “VRF-lite” (no MPLS or any kind of VRF tagging)**
- **Communications between the two domains is done via the CERN Backbone routers**
- **Only LCG domain has access to LHCOPN/LHCONE**

DC Network design – VRFs

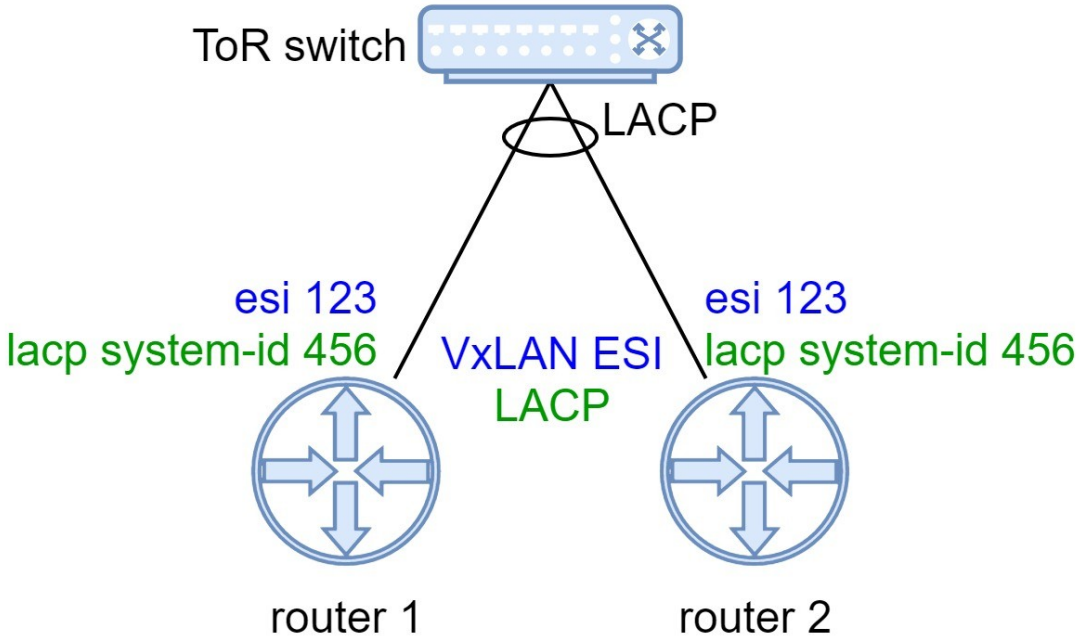


DC Network design – ToR switches

- **All switches are attached to two distribution routers, via LACP (active/active links)**
- **ToR switch are configured as Basic L2 switches**
- **All switch ports in the default VLAN (no VLAN/VxLAN on the TOR switch)**
- **“Keep it simple” rule**
- **Current usual setup:**
  - 40x 10G (copper) connections to servers + 2x40Gbps uplink (BF= 80:400 = 1:5)
  - 32x 25G (fibre) connections to servers + 4x100Gbps uplink (BF= 200:400 = 1:2)
  - Some 100G NIC servers are used in specific setup with a switch BF = 1:1

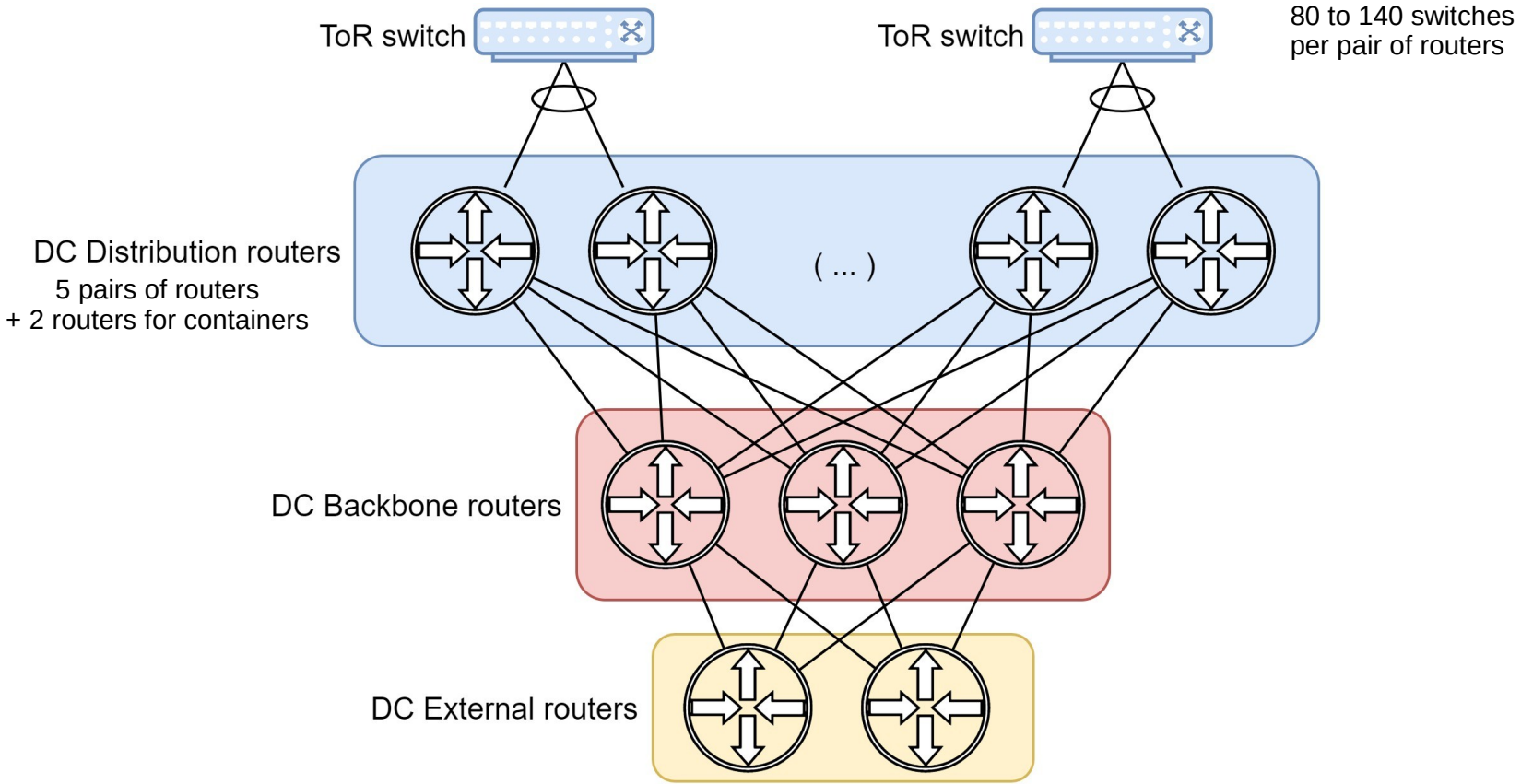
DC Network design – ToR switches

- **Attachment to two routers is done using EVPN/VxLAN ESI on the routers**
- **Active/active redundant links using standard protocols**



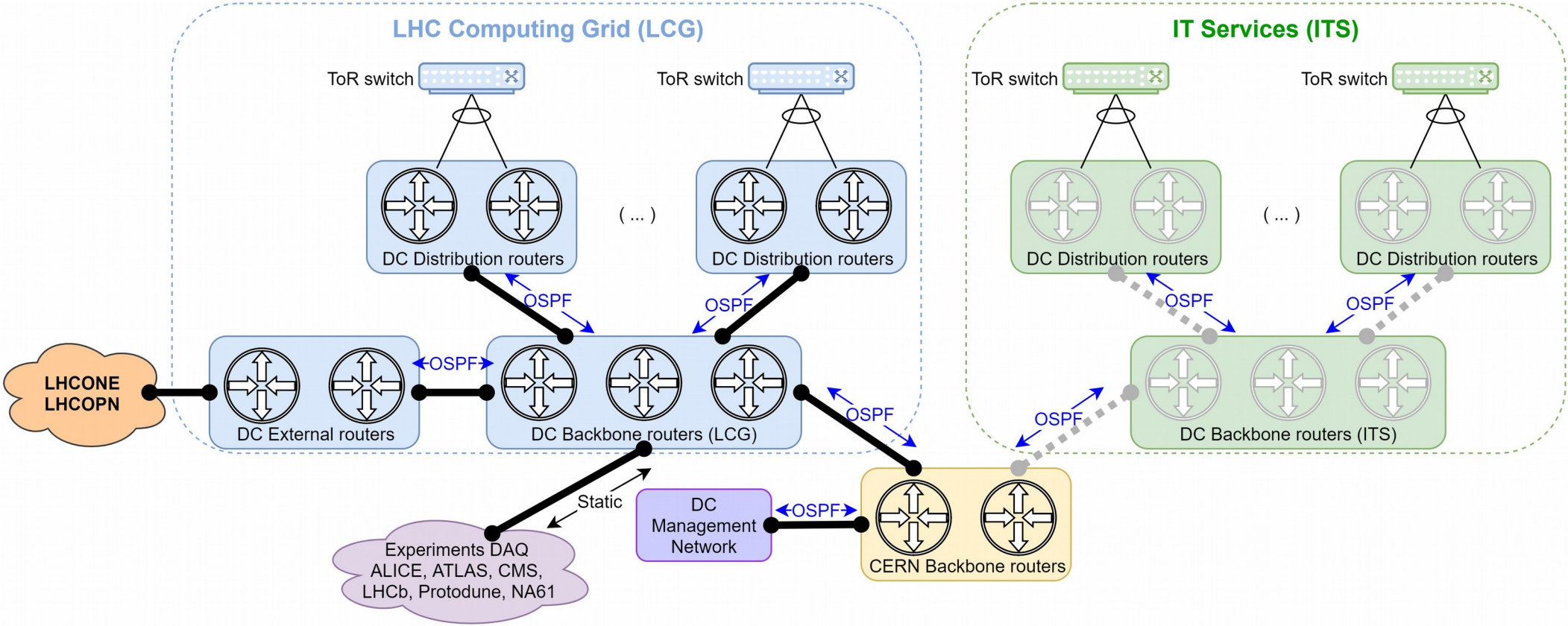
DC Network design – Routers

- **CLOS topology between routers**



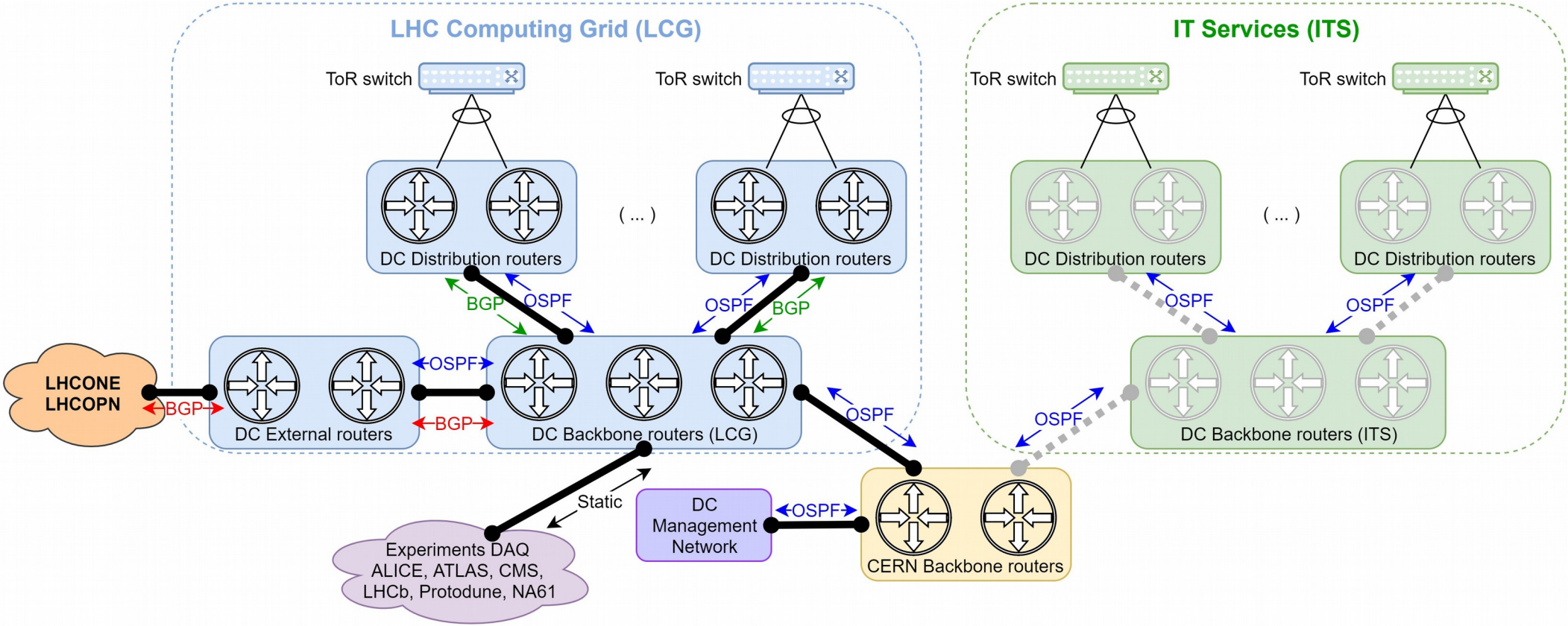
DC Network design – Routing protocols

- **OSPF "underlay"**



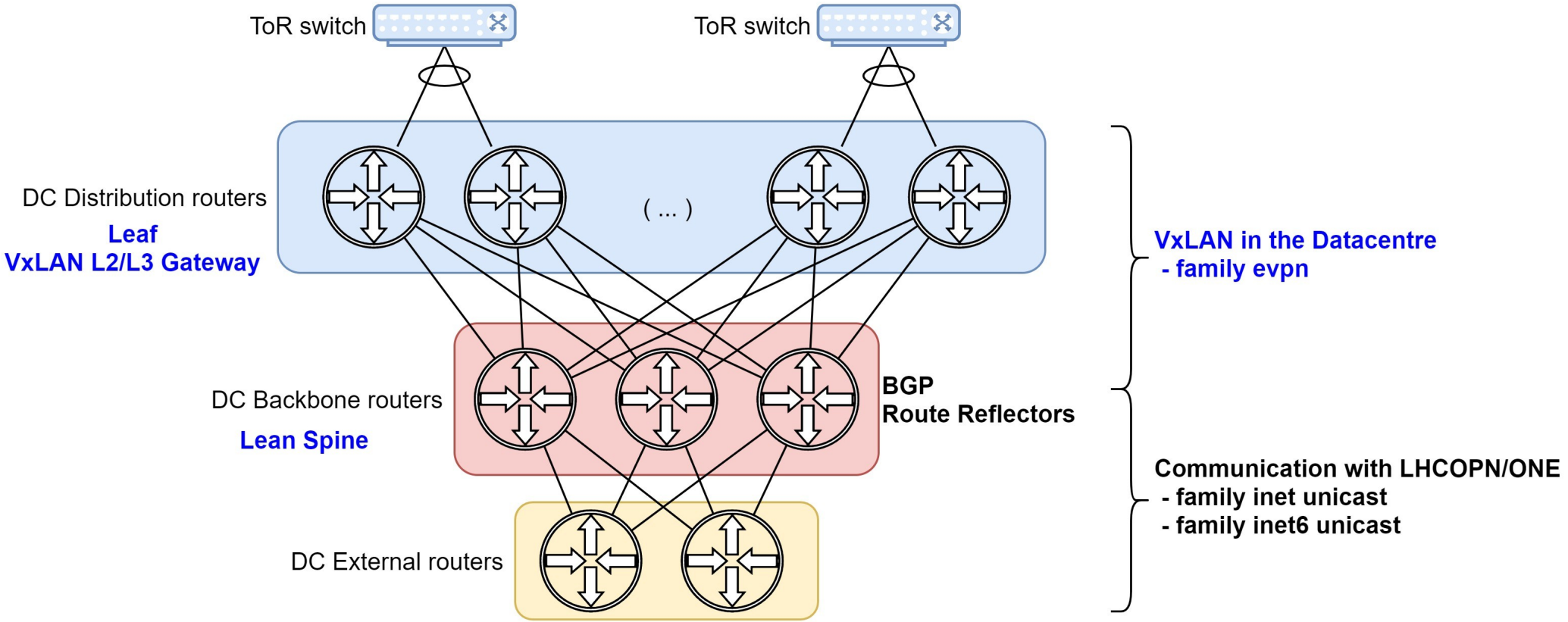
DC Network design – Routing protocols

- **BGP for LHCONE/OPN + BGP for EVPN/VxLAN "overlay"**



DC Network design – Routing protocols

- **BGP setup**



Integration with other network services

- **Internal Database used as “source of truth” (LANDB/CSDB)**
- **Switch configuration automated via internal tools (cfmgr)**
  - Provide a single tool for technicians, no matter the switch model
  - Ease installation and replacement
- **Server autoregistration process:**
  - Deep interaction with DHCP, DNS, LANDB/CSDB
  - Automatically declare the server in the database (MAC@, switch port, IP@, hostname, etc…)
- **Ongoing work to interconnect the physical network with a “SDN cloud” (Tungsten Fabric)**

Plans for the future

- **Automate Juniper device configuration (Based on “ZTP”, cfmgr, etc…)**
- **Add distribution routers dedicated for the 100G NIC servers**
- **Support Vlan and/or VxLAN up to the ToR switch**
  - support several broadcast domains per switch
  - add complexity in the switch configuration and interactions with existing automated processes
- **Integration with “SDN” solution (Tungsten Fabric) –** *see next presentation from Ricardo Rocha*
  - Integration between the virtualized part and the physical routers
  - Add flexibility for some use cases
  - We need to ensure no conflict with existing tools/processes

Plans for the future

- **New Datacentre on Prévessin site (12MW, 576 server racks – construction in 3 phases)**
- **Support for 400G**
  - which standard will be the correct one?

	reference	Form Factor	Connector	Fiber type	distance	number of fibers per connection	Current use-case
100G	QSFP-100G-BXSR	QSPF28	Duplex LC	**MMF**	100m (OM4)	2	Switch-to-router Router-to-router short dist.
	QSFP-100G-SR4	QSPF28	MPO-12	**MMF**	100m (OM4)	12	Switch-to-Server (100G)
	QSFP-100G-LR4	QSFP28	Duplex LC	**SMF**	2m to 10km	2	Router-to-router long dist.
400G	QDD-400G-DR4	QSFP-DD	MPO-12	**SMF**	2m to 500m	12	Not used yet
	QDD-400G-FR4	QSFP-DD	Duplex LC	**SMF**	2m to 2km	2	Not used yet
	QDD-400G-LR4-10	QSFP-DD	Duplex LC	**SMF**	2m to 10km	2	Not used yet
	QDD-400G-LR8	QSFP-DD	Duplex LC	**SMF**	2m to 10km	2	Not used yet

home.cern