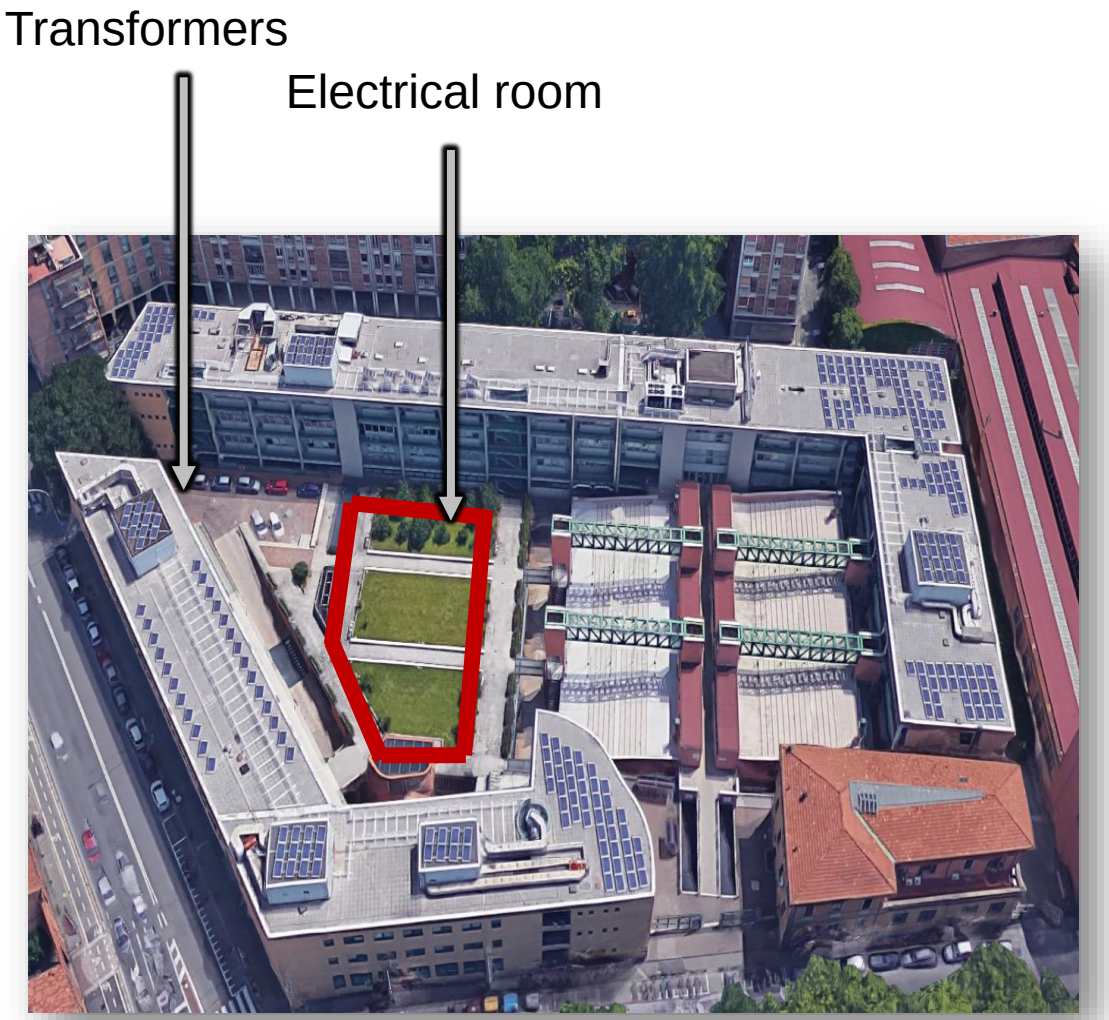
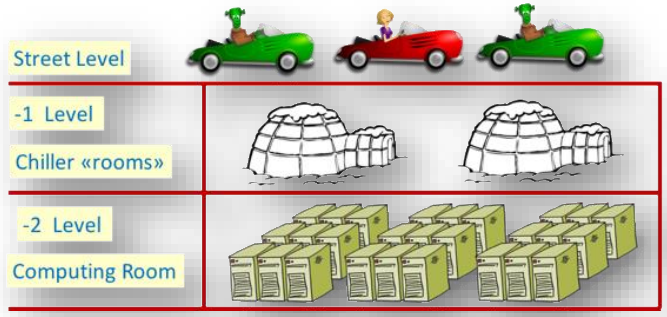
INFN Tier-1 status

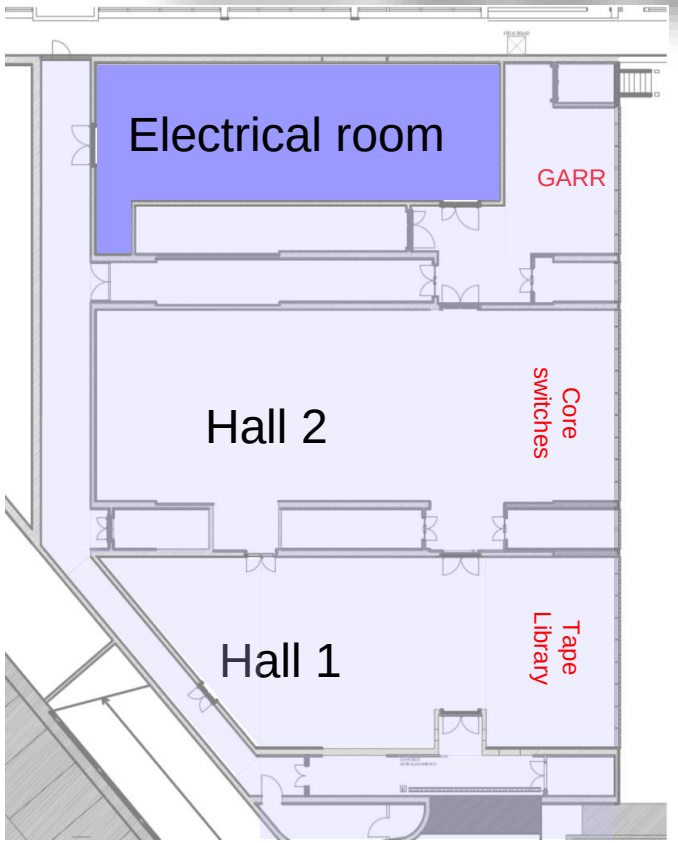
Luca dell'Agnello
Daniele Cesini

GDB@CERN - Feb 14 2018

The Tier-1 location

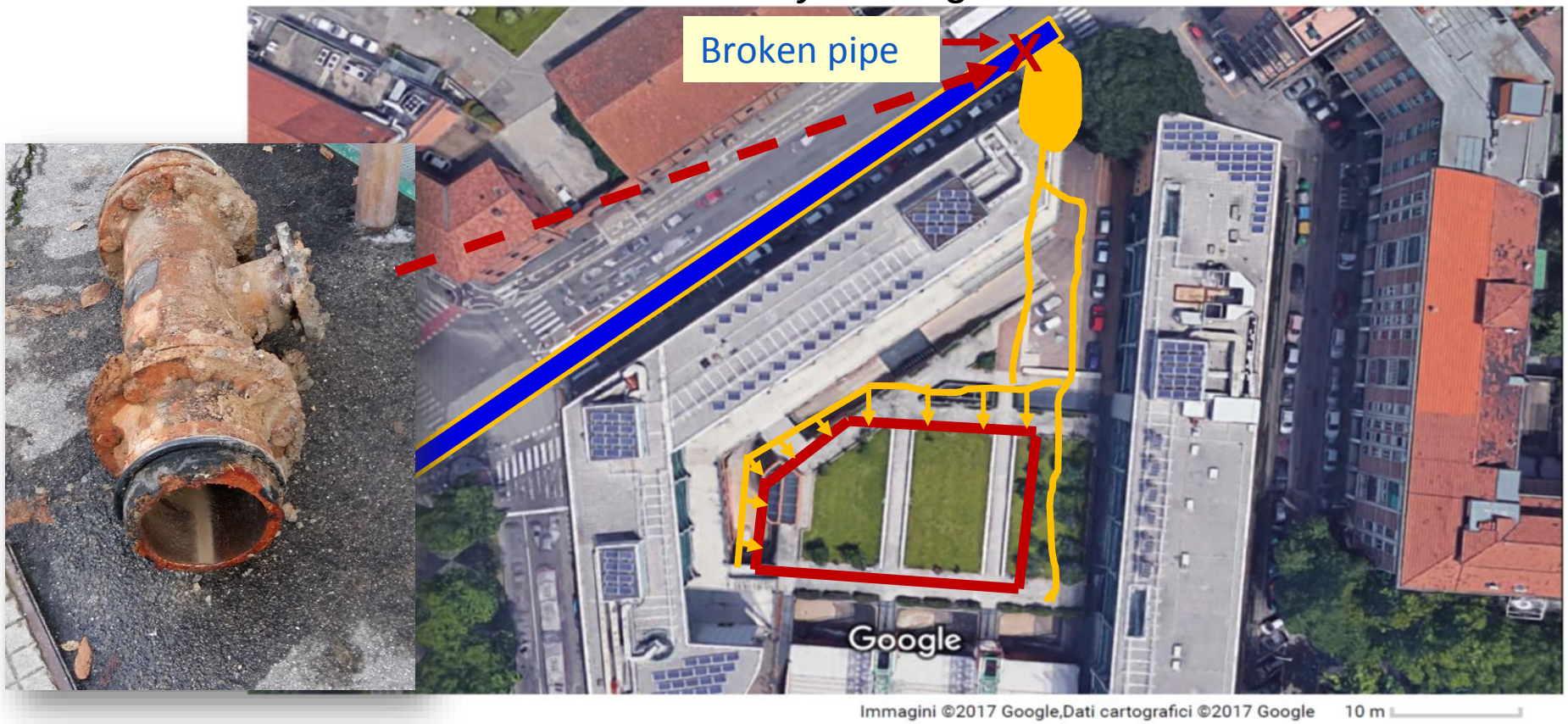






11/9: the flood

- The flood happened on November 9 early in the morning
  - Breaking of one of the main water pipelines in Bologna
  - Also the road near CNAF seriously damaged



The Tier-1 entrance that morning

All Tier-1 doors are watertight
Height of water outside: 50 cm
Height of water inside: 10 cm (on floating floor) for a total volume of ~500 m^3

First inspection

- Access to data center possible only in the afternoon
- Nearly all the electrical equipment in the electrical room damaged by the water
  - Both power lines compromised
- The two lower units of all racks in the IT halls submerged
  - Including the two lowest rows of tapes in the library

Damage to IT equipment (1)

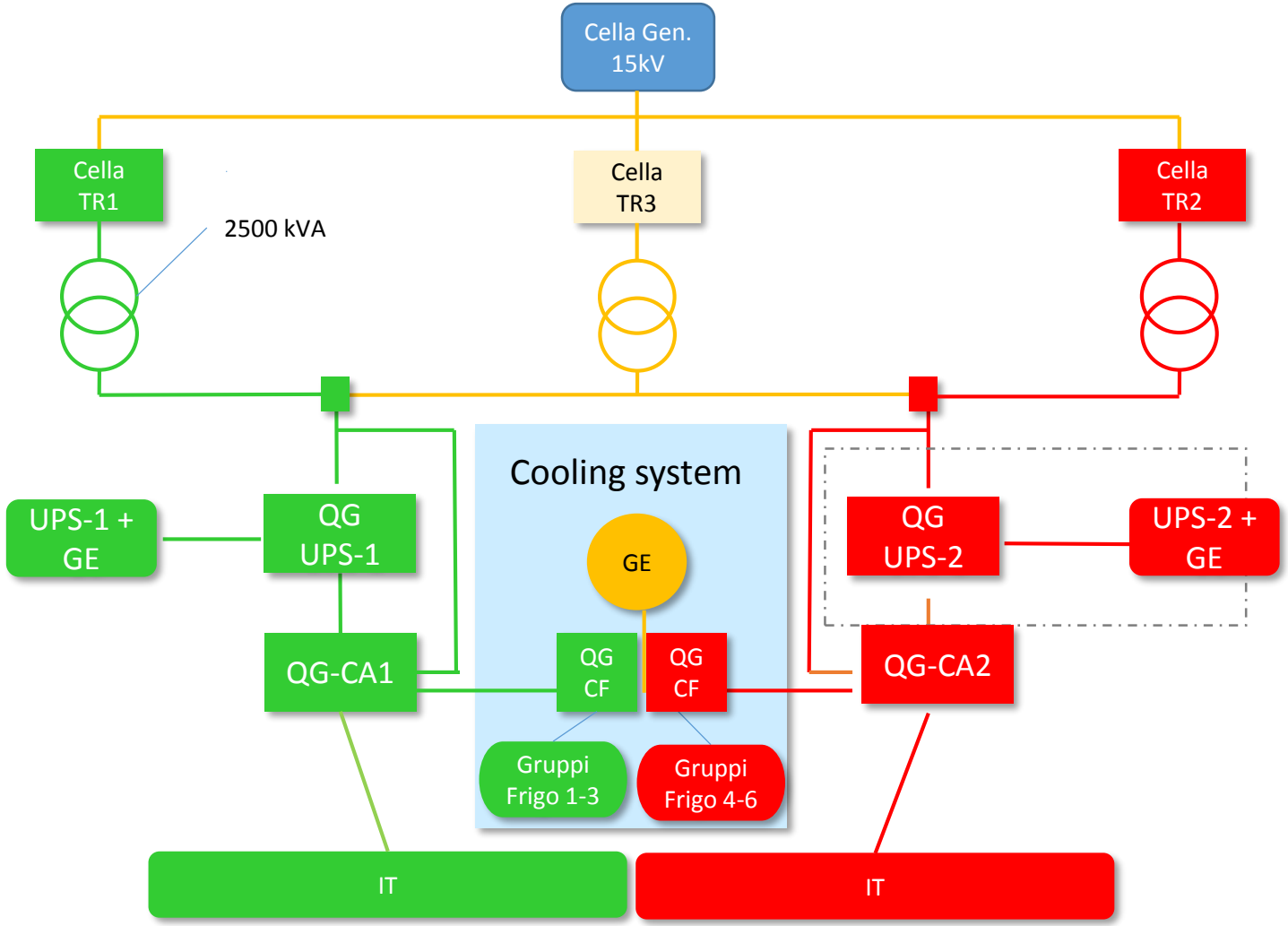
- Computing farm
  - ~34 kHS06 are now lost (~14% of the total 2018 capacity)
- Library and HSM system
  - 1 drive damaged
  - Cleaning needed (completed)
  - Recertification needed (completed)
- Tapes
  - 136 tapes damaged
  - “wet” tapes are being recovered in Oracle lab
    - Very slow process (ETA: end of March)
    - Prioritization based on requests from experiments

Damage to IT equipment (2)

- Nearly all storage disk systems involved
  - 11 DDN JBODs (LHC, AMS)
    - RAID parity affected
  - 2 Huawei JBODs (all non-LHC experiments excepting AMS, Darkside, Virgo)
  - 2 Dell JBODs including controllers (Darkside and Virgo)
    - Most critical - 2 trays out of 5 went underwater.
  - 4 disk-servers (4 Alice) + 4 TSM-HSM servers

System	PB	JBODs	Disks	Involved experiments
Huawei	3.4	2	150 x 6 TB	Astro-particle and nuclear experiments excepting AMS, Darkside e Virgo
Dell	2.2	2	120 (48) x 4 TB	Darkside and Virgo
DDN 1,2	1.8	*4*		ATLAS, Alice and LHCb
DDN 8	2.7	*2*		LHCb
DDN 9	3.8	2		CMS
DDN 10, 11	10	3+2	252 x 8 TB	ATLAS, Alice and AMS
Total	**23.9**	**17**	**~4 PBytes**	

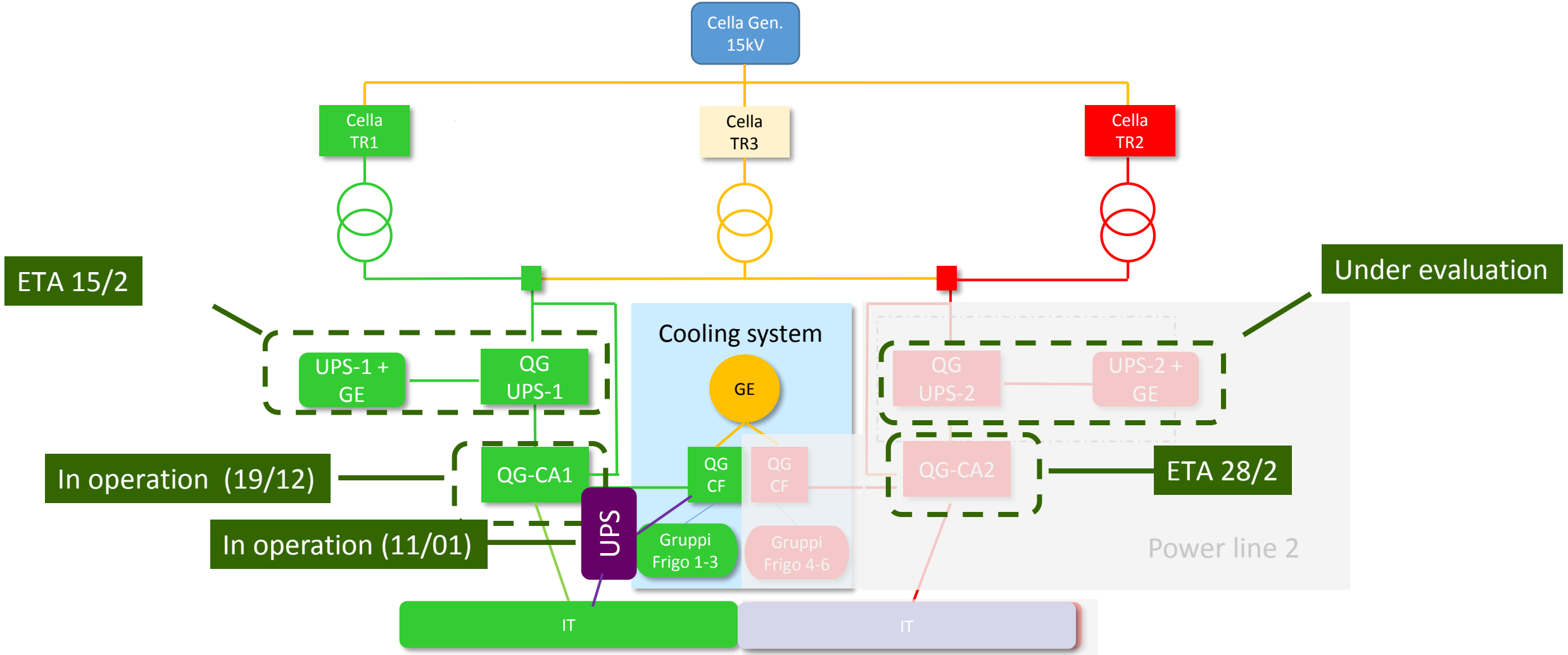
Power Center configuration before the flood



Restoring the Electrical Lines

- Activated a temporary power line (60 kW) after 1 week
  - Essential for GARR POP equipment
- Recovery of 1st electrical line (1.4 MW) w/o UPS completed before the end of December
- Temporary UPS (300 kW) + GE on 1st line in place since Jan 12
  - Enough to switch on all the storage systems and part of the farm
- Work in progress to restore full continuity on the 1st line (15/02)
- Work on 2nd line starting this week

Present Power Center status



IT services recovery and Halls cleaning

- Basic services
  - IT services (non scientific computing) immediately moved outside CNAF
  - The General IP connectivity restored few days after the flood
- Halls
  - Data center dried over the first week-end
  - Cleaning from dust and mud completed during the first week of December
  - Other miscellaneous activities (e.g. floating floor replacement, fire alarm system recertified etc…)
- In the meanwhile activity to recover wet IT equipment
  - Cleaned and dried (using oven when appropriate) disks, servers, switches,….
  - IT components to be replaced have been ordered
- Deep inspection of the data center to understand the flow of the water
  - Now understood, various sources inside the datacenter, including another broken water pipe

Cooling and Core switches

- Cooling
  - 3 chillers (out of 6) in operation since Jan 15
    - To switch on the other 3 chillers also the 2nd electrical line is needed
    - Limit the farm power
  - In-row cooling systems recertified
    - Possible to switch on also part of the farm

- Core switches tested and upgraded to 100 Gbit in Dec
  - Needed to install new storage and for connection to CINECA for farm extension

Library recovery

- Tape library cleaned and recertified
- SAN restored
- Replacing damaged TSM-HSM servers (next week)
- Currently performing audit on tapes

Storage recovery roadmap

- Dell systems recovered (Darkside and Virgo exp)
  - Replacement of damaged parts only (i.e. crates) and switched on Jan 15
  - “Compromised” disks replaced during normal operations
    - Wet disks must be replaced to have access to standard support
  - File-system checked and now in production
- DDN1, DDN2 (ATLAS, ALICE; LHCb): damaged components replaced by us with spare parts
- Replacement parts for DDN10 and DDN11 (ATLAS,ALICE, AMS) delivered on Feb 5
  - Installation on Feb 7-8 and now ready for production
- Huawei (Astro-particle and nuclear physics experiments) replacement parts delivered Jan 27
  - Still struggling to recover ….
- Data on DDN8 – LHCb (out of maintenance) will be moved onto new storage (after acceptance test)
- Disks of DDN8 will be used to replace wet disks of DDN9 (CMS)

Storage status

System	Type	Status	Readiness
Alice	Tape buffer	OK	PRODUCTION
	Disk	Parity ok	Ready (Next week in prod)
Atlas	Tape buffer	OK	PRODUCTION
	Disk	parity ok	Ready (Next week in prod)
CMS	Disk + tape buffer	Degraded parity: raid5 in few LUNs, raid 6 in the others Disks to be replaced	Ready (Next week in prod)
LHCb	Disk + tape buffer	Degraded parity: raid5 in all LUNs Data to be moved to the new system	Start moving during 2nd half of February Production in March
AMS	Disk	Parity ok	Ready (Next week in pord)
Virgo	Disk + tape buffer	OK	PRODUCTION
Darkside	Disk	OK	PRODUCTION
Astro-particle experiments	Disk	Maintenance intervention	UNKNOWN

Disk servers were mounted yesterday on the racks

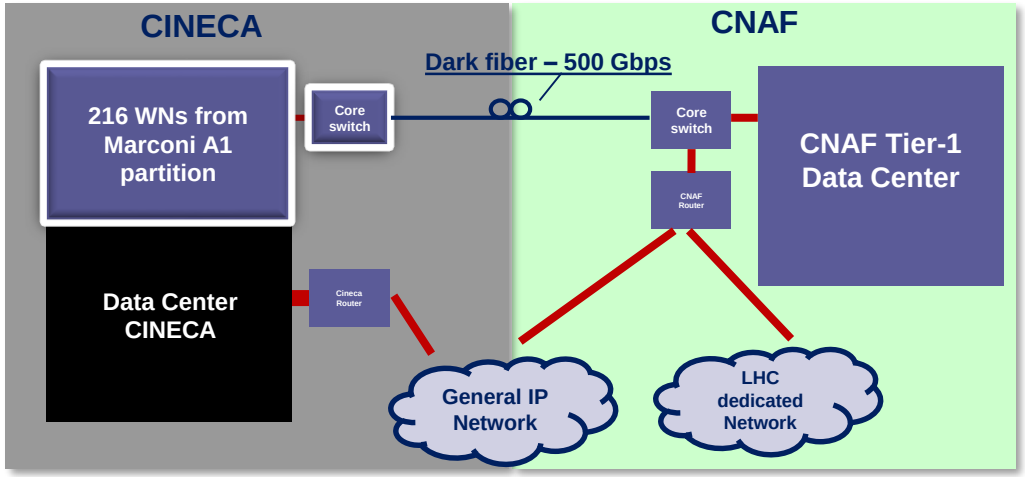
As soon as the new system will be available we will open ro the old fs and rw the data moved

Farm recovery

- Test and recovery of farm services completed
  - LSF masters, CEs, squids etc…
- Performed upgrade of WNs
  - Middleware, security patches (i.e. meltdown etc..)
- Part of the local farm powered on (only 3 chillers in production)
  - 50 kHS06 for the moment
  - Exploiting the CNAF farm elastic extension to provide more computing power
    - Remote farm partition in Bari-RECAS (~24 kHS06)
    - Install the CNAF-CINECA extension farm (~ 170 kHS06) next month

Farm remote extensions

- ~13% of CPU resources pledged to WLCG experiments are located in Bari-RECAS data center
  - Transparent access for WLCG experiments
  - Similar to CERN/Wigner extension
  - 20 Gbps VPN
  - Disk cache provided via GPFS-AFM
- In 2018 ~170 kHS06 will be provided by CINECA
  - Setup on going
  - 500 Gbps (→ 1.2 Tbps) VPN ready
  - No disk cache, direct access to CNAF storage
    - Quasi-LAN situation
- Participation to HNSciCloud project
- Tests of opportunistic computing on a commercial cloud providers (Aruba, Azure)



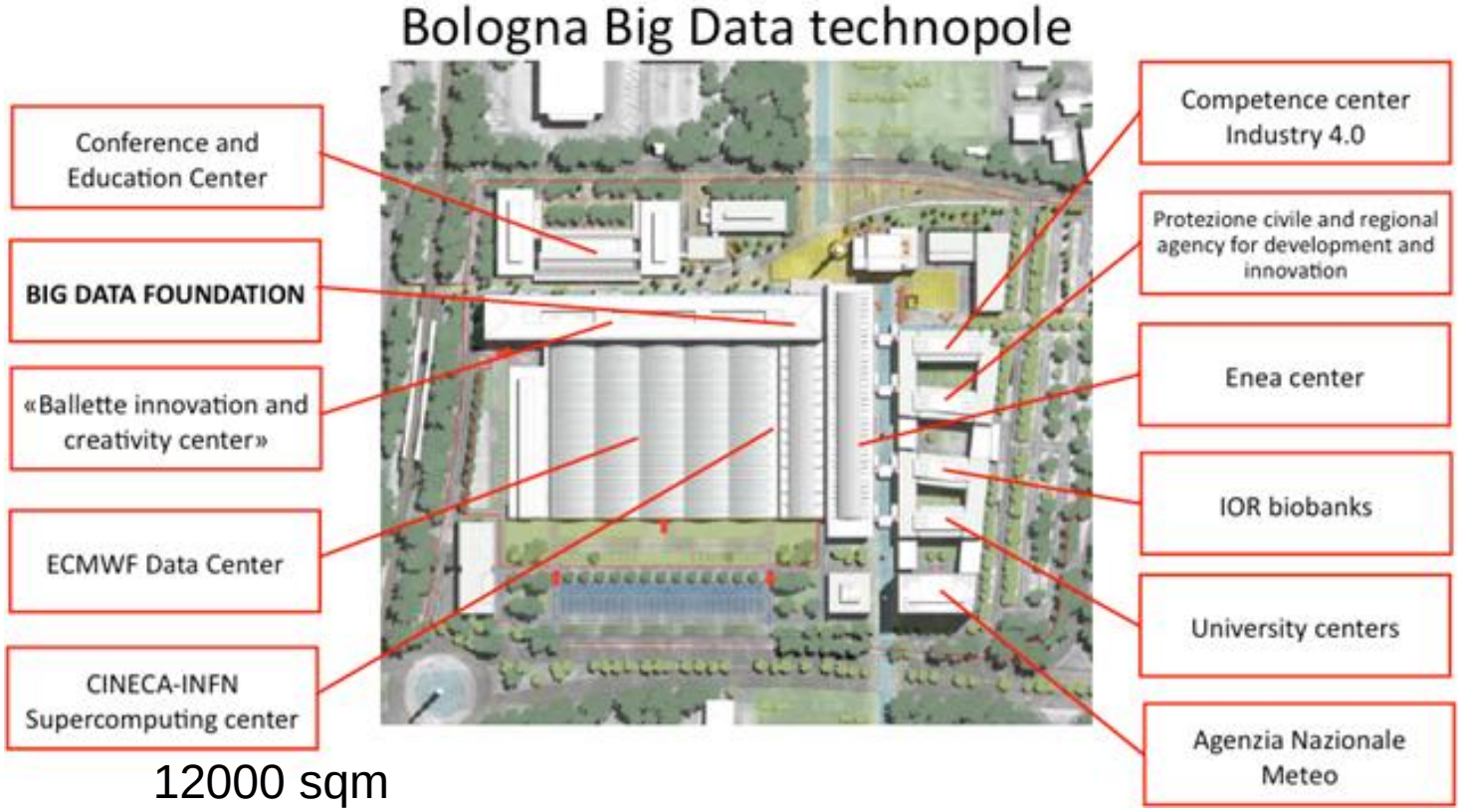
Search for a new location for the Tier-1

The goal: provide a new location for the INFN Tier-1 to take into account future expansion.

ECMWF center will be hosted in Bologna from 2019 in the Tecnopolo area.

Possibility to host in the area also:
- INFN Tier-1
- Cineca computing center

Already allocated 40 M€ for the Italian government to refurbish the area.
Looking for extra budget for INFN & CINECA



http://www.urbancenterbologna.it/en/bologna/regeneration-projects/tecnopolo

Summary

- One electrical line recovered
  - UPS on this line available end of this week
  - It allows to switch on the storage and part of the farm (due to cooling constraints)
- Farm services recovered
  - Missing resources (34 kHS06) provided by remote sites
    - INFN-Bari already in production (24kHS06), CINECA during 2018
- Storage for CMS, ATLAS, ALICE, AMS is ready and can be moved to production next week
- We need to find a temporary solution for LHCb
  - All data has to be moved to a new Storage system not yet ready
- Serious issues with astro-particle experiments storage
  - Not for VIRGO - in Production