Disaster recovery of the INFN Tier-1 data center: lesson learned

Luca dell'Agnello

INFN - CNAF

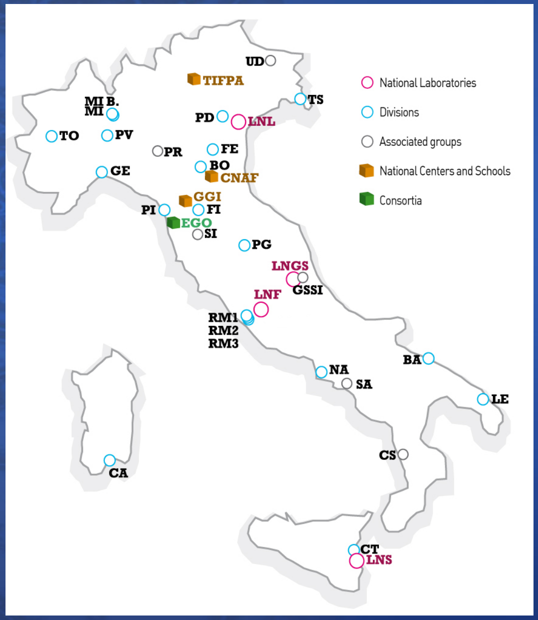
CHEP 2018 conference

Sofia, July 12 2018

INFN

INFN

- **National Institute for Nuclear Physics** (INFN) is funded by Italian government
- Main mission is the research and the study of elementary particles and physics laws of the Universe
- Composed by several units
  - ~ 20 units dislocated at the main Italian University Physics Departments
  - 4 Laboratories
  - 2 National Centers dedicated to specific tasks
- CNAF, located in Bologna, is the INFN National Center dedicated to computing applications



The INFN Tier-1

- First incarnation dates back to 2003 as computing center for BaBar, CDF, Virgo and prototypical for LHC experiments (Alice, ATLAS, CMS, LHCb)
- After a complete infrastructural refurbishing in 2008, it nowadays provides services and resources to more than 30 scientific collaborations
  - 70-80% resources for WLCG experiments....
  - (... but most effort due to support non-WLCG experiments!)
- Planning a new data center to cope with the high demanding computing requirements of HL-LHC and newly coming experiments

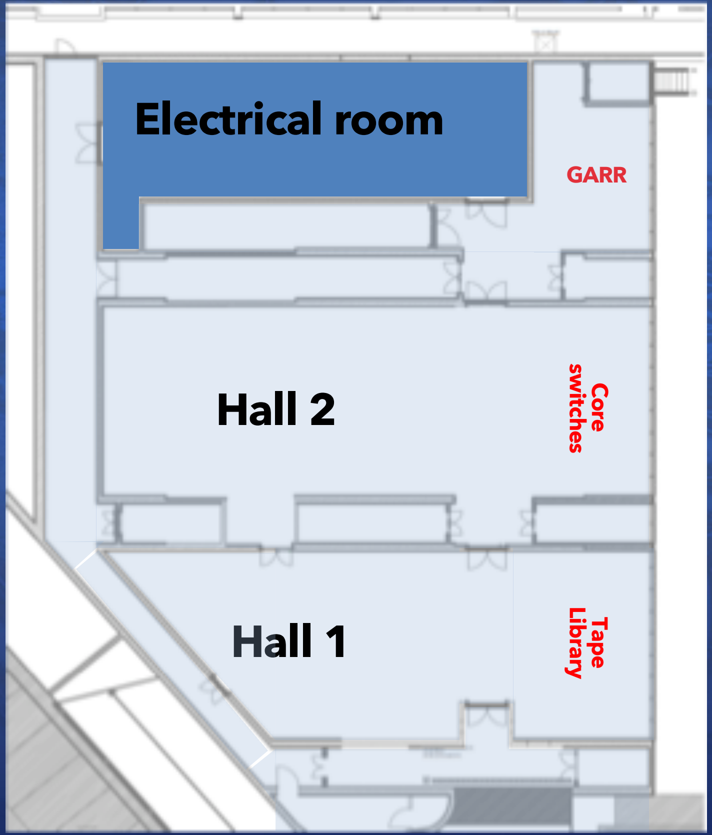
The INFN Tier-1: some figures

- The data center is located 2 levels under the street level
- It is divided in 4 main halls:
  - 2 halls for IT
  - 1 small hall for the GARR PoP
  - 1 electrical room
- Resources (before the flood)
  - ~1.000 WNs , ~20.000 computing slots, ~220 kHS06 (+ ~20 kHS06 in Bari-ReCaS)
    - Also small (~33 TFlops) HPC cluster available
  - ~23.4 PB of storage on disk
  - 1 tape library with 42 PB of data
  - Dedicated network channel (60 Gb/s) for LHC OPN + LHC ONE
    - 20 Gb/s reserved for LHC ONE
- 21 people working on Tier-1 (including facilities support staff)

The INFN Tier-1 location

Transformers

Electrical room



The flood

November 9: the flood

The flood occurred early in the morning due to the breaking of one of the main water pipelines in Bologna, located in a road next CNAF

The entrance of the data center at 7.37 CET

After I was alerted, my first thought was for the external doors (all Tier-1 doors are watertight)
Then, with a fast check, I realized the data center was no more on line.

November 9: the flood

The flood occurred early in the morning due to the breaking of one of the main water pipelines in Bologna, located in a road next CNAF

The entrance of the data center at 7.37 CET

After I was alerted, my first thought was for the external doors (all Tier-1 doors are watertight) Then, with a fast check, I realized the data center was no more on line.

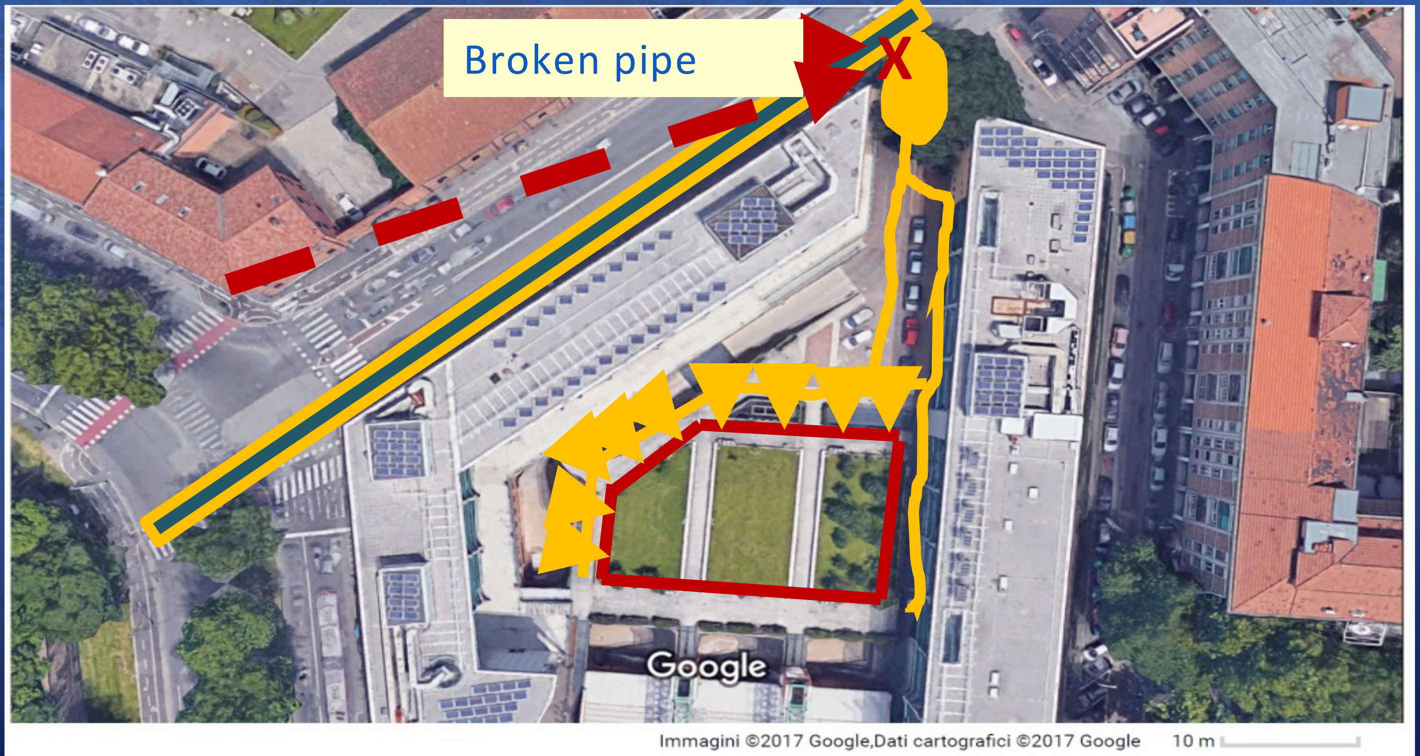
~~Our first Data lake~~ ☺

November 9 flood: some findings

- Height of water outside: 50 cm
- Height of water inside: 10 cm (on floating floor) for a total volume of ~500 m^3
- The water poured into the data center through walls and floor….

Ø = 20 cm

The broken pipe

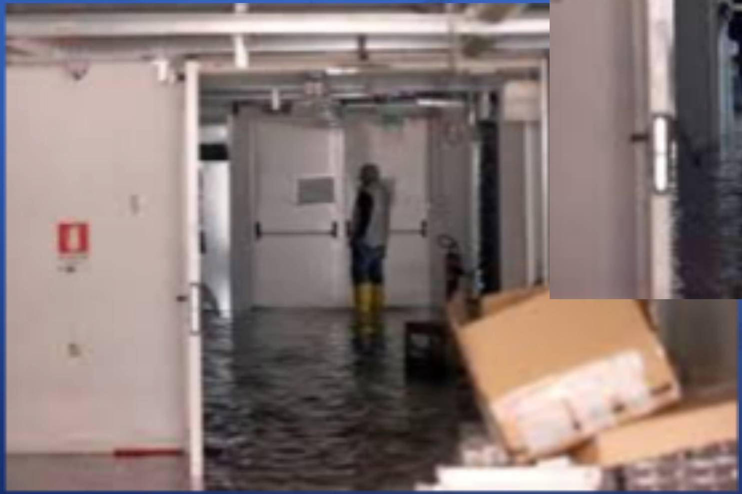


The situation outside.....

The road collapsed under the weight of the fire truck at CNAF entrance

Men at work on broken pipe

…. and inside the data center



Steam produced by Joule effect (electrical room)

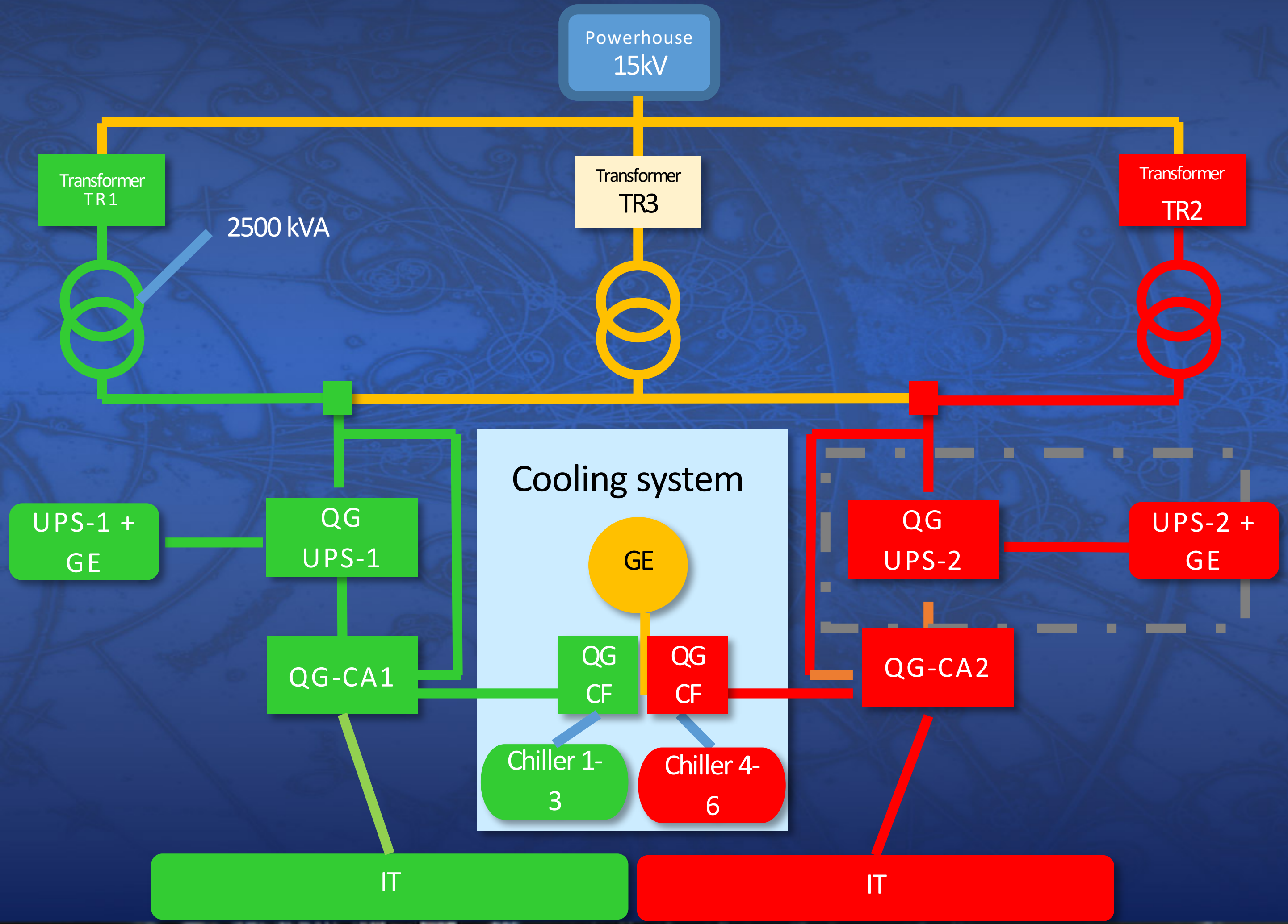
Photos taken in the afternoon after part of the water had been pumped out

Damage assessment and first intervention

- The power center was the most damaged part
  - Both 1.4 MW power lines compromised (including control for UPS's/diesel engines)
- The two lower units of all racks in the IT halls were submerged
  - Including the two lowest rows of tapes in the library
  - All storage systems involved
- The 3 Core Switch/Routers and the General IP Router were safe for few centimeters
- First operations: data center dried over the first week-end
- Cleaning from dust and mud started immediately after
  - Specialized company supported us
  - Operation completed during the first week of December

Power Center configuration before the flood

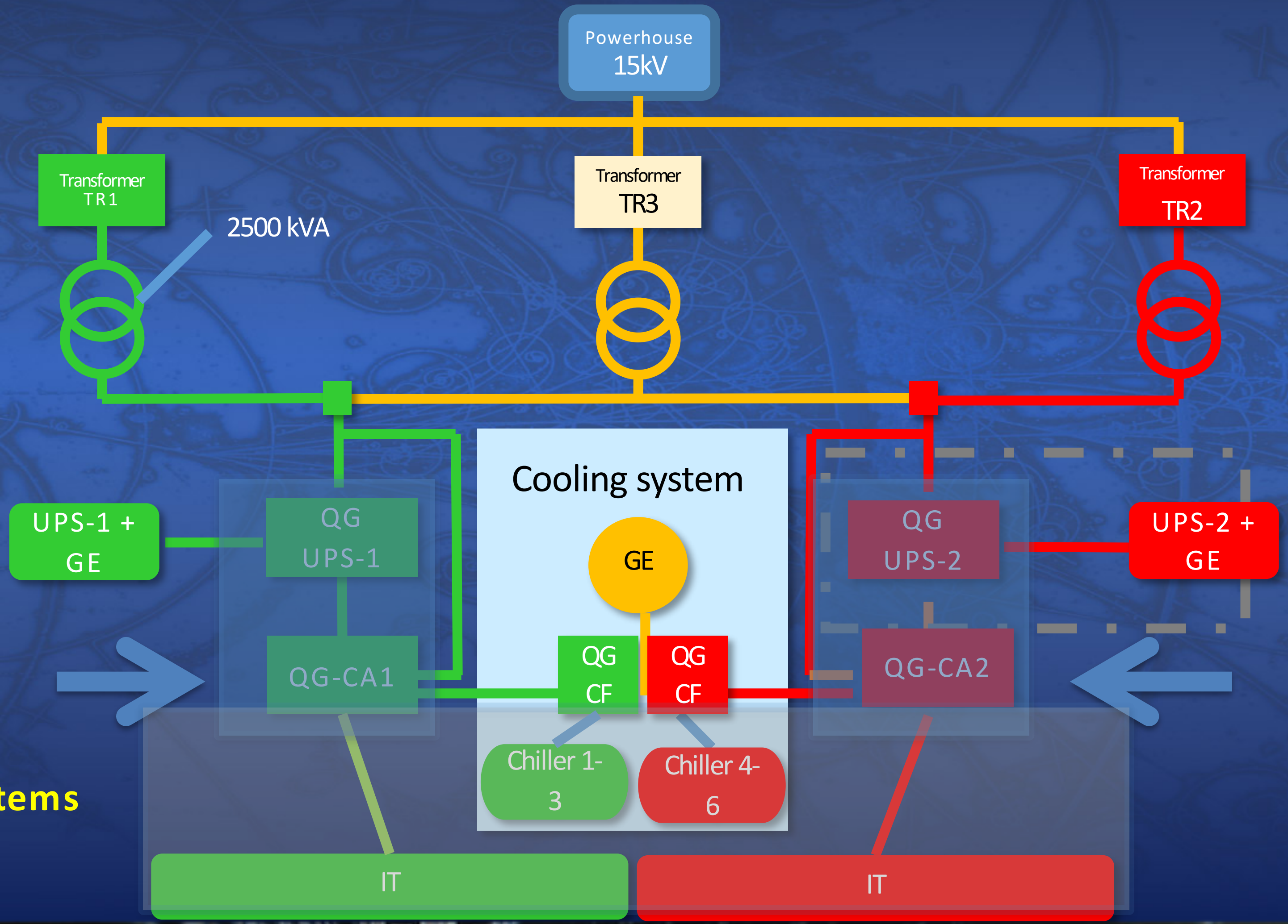
Electrical Transformers
(Energy Supplier)



QG=Electric Panel
GE=Generator

Power Center after the flood

Electrical Transformers (Energy Supplier)



Damaged by water

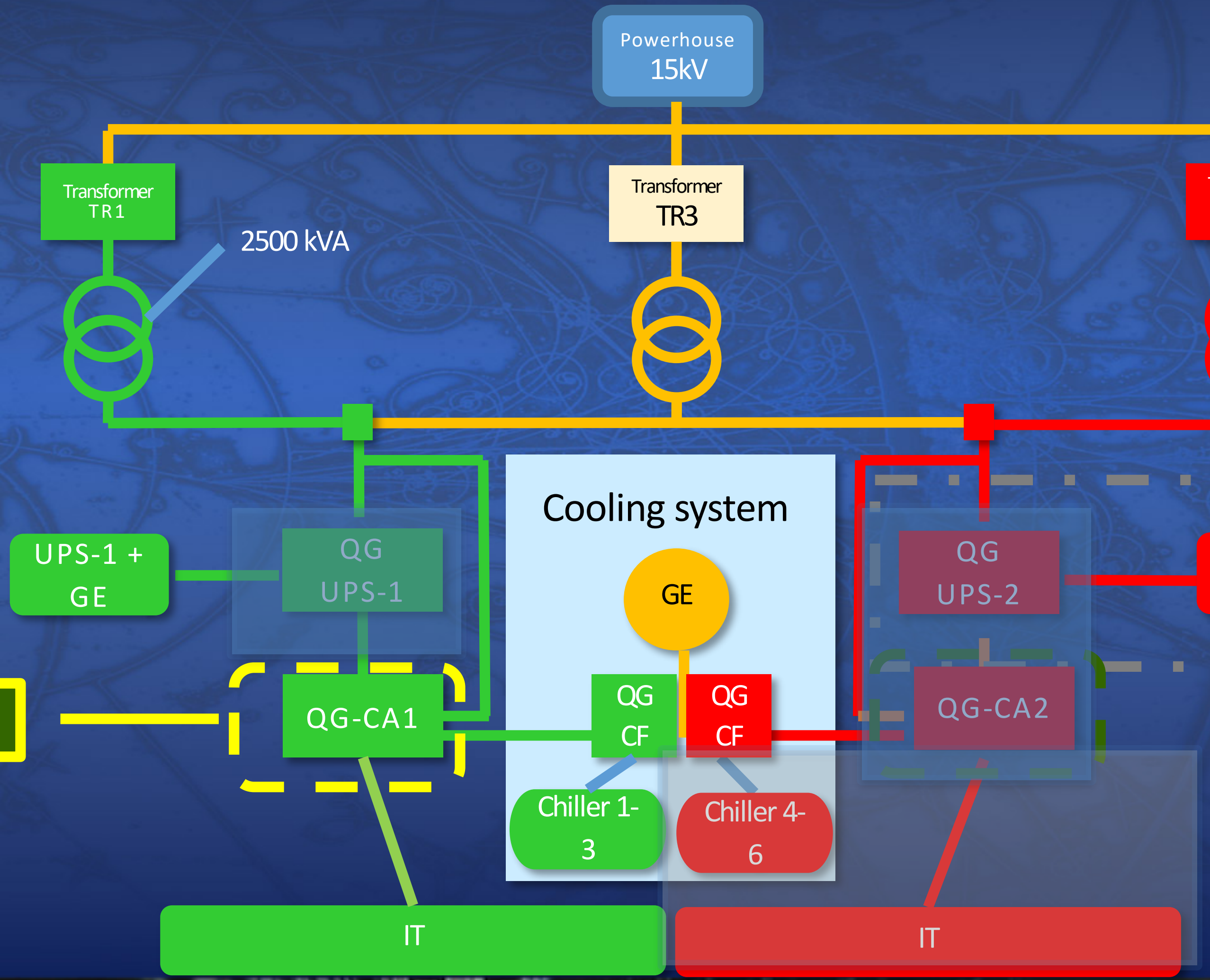
Damaged by water

All the electrical systems Affected !

QG=Electric Panel
GE=Generator

Power Center recovering….

Electrical Transformers
(Energy Supplier)



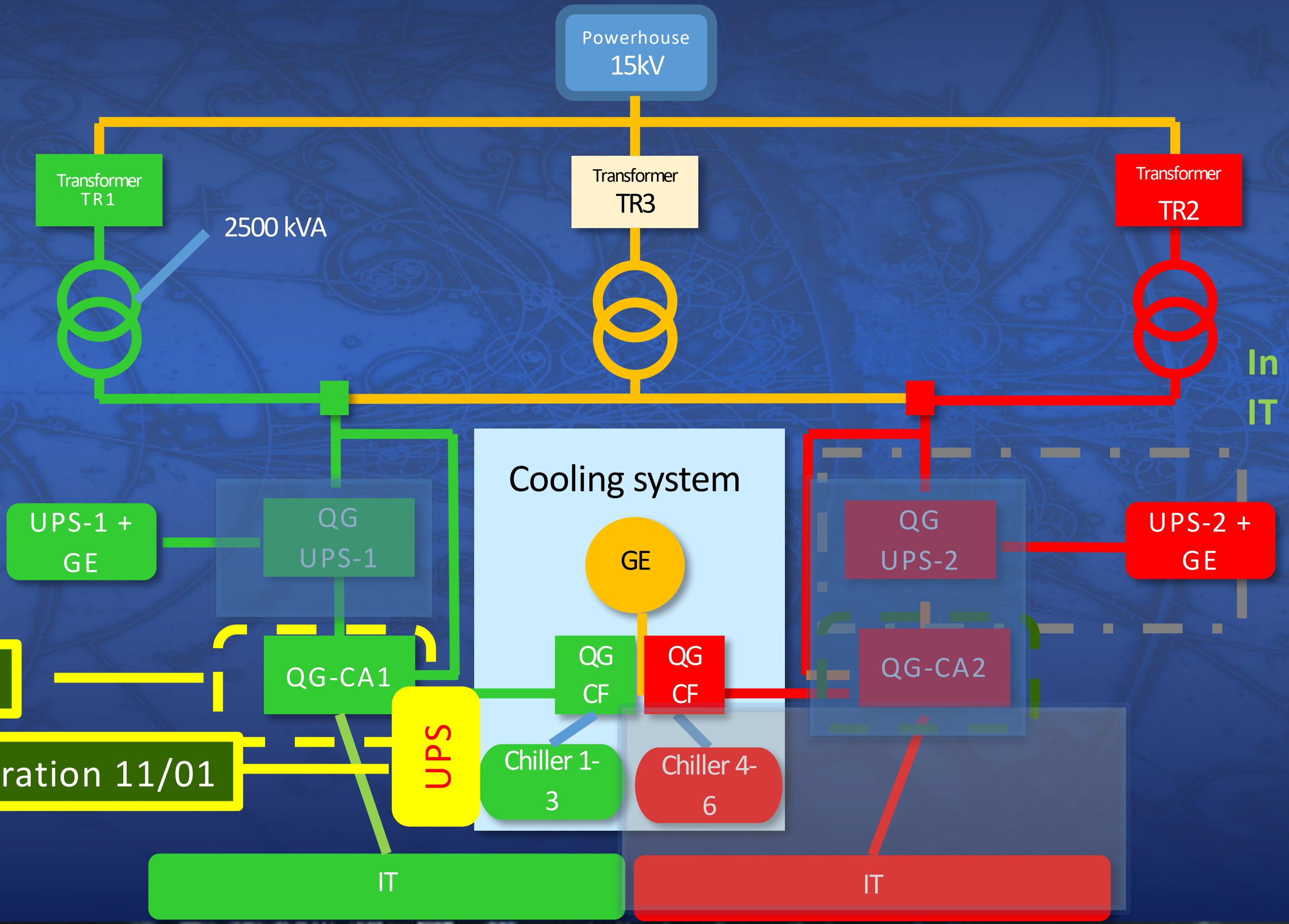
In operation 19/12

First power line active
(no UPS/diesel engine)

QG=Electric Panel
GE=Generator

Power Center recovering….

Electrical Transformers
(Energy Supplier)



In row air conditioning in the IT halls restored mid January

In operation 19/12

In operation 11/01

First power line active
(small UPS + diesel engine)

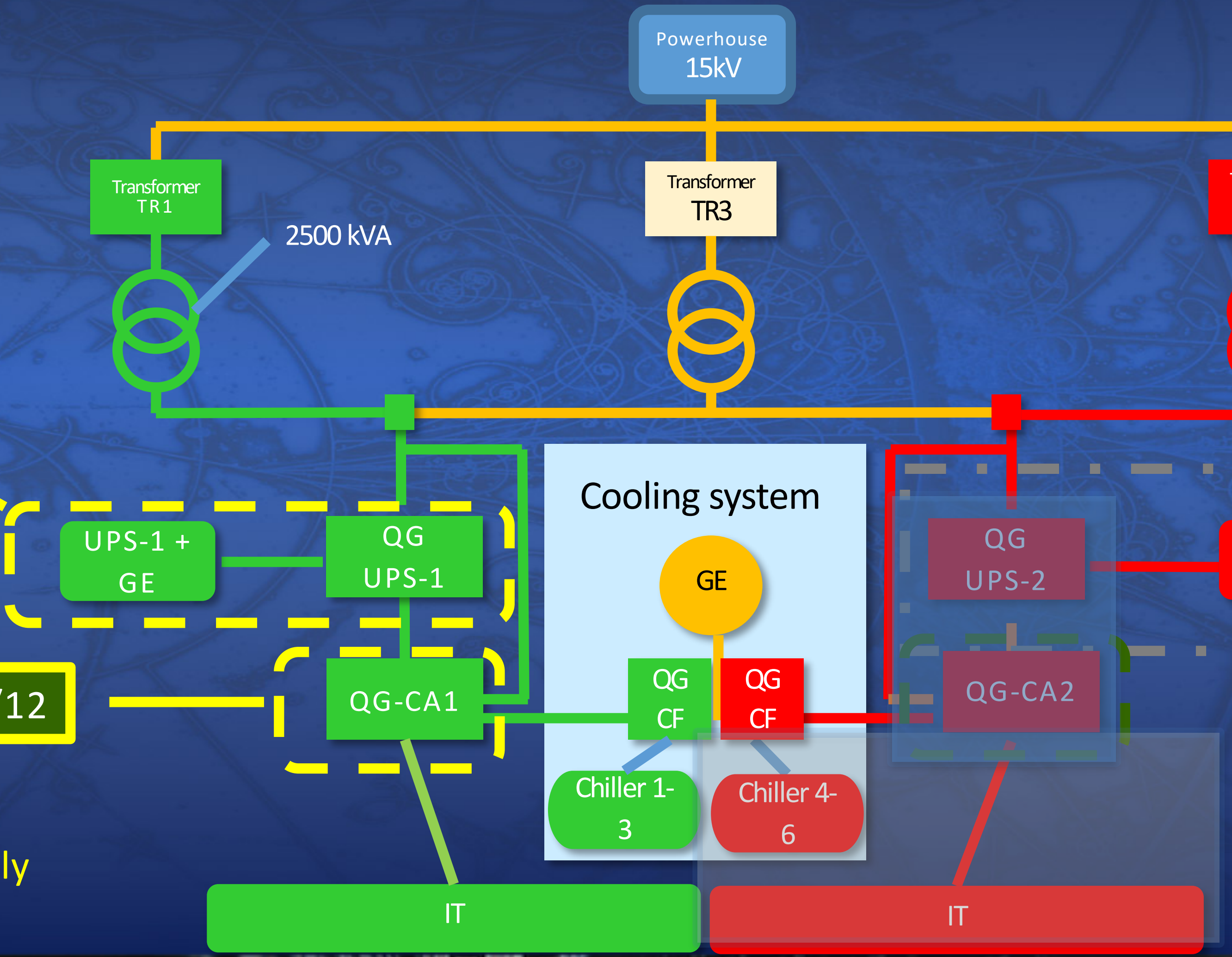
QG=Electric Panel
GE=Generator

Power Center recovering….

Electrical Transformers
(Energy Supplier)

In operation 15/2

In operation 19/12

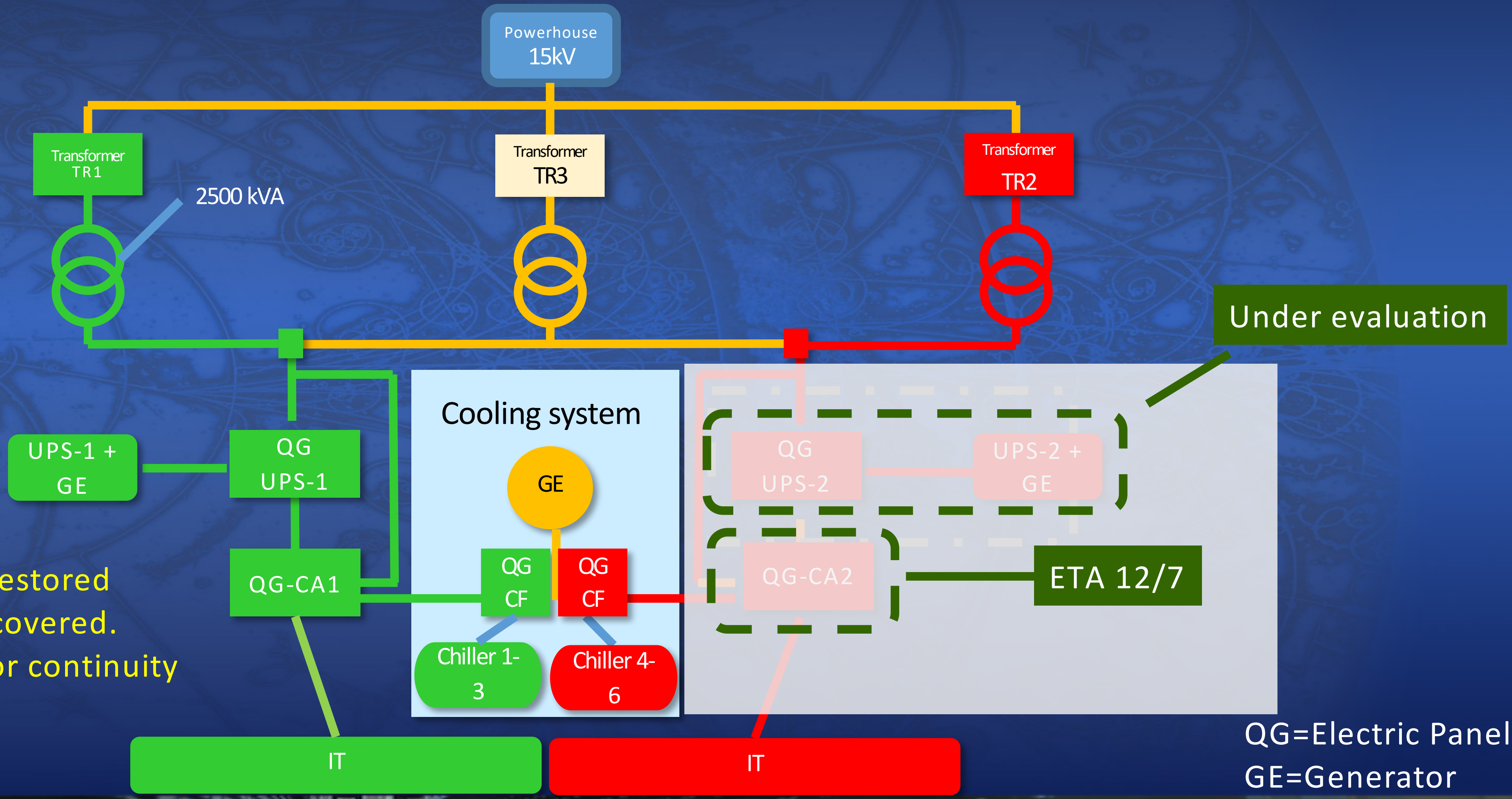


First power line completely restored

QG=Electric Panel
GE=Generator

Present Power Center status

Electrical Transformers
(Energy Supplier)



First power line completely restored
Second power line nearly recovered.
Still to be decided strategy for continuity for the second line

QG=Electric Panel
GE=Generator

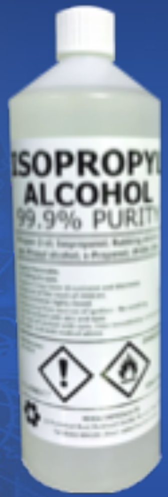
Damage to IT equipment: the list

- Computing farm
  - ~34 kHS06 are now lost (~14% of the total capacity)
  - No special action taken (replaced)
- Library and HSM system
  - 1 drive and several non critical components damaged
  - 4 TSM-HSM servers (replaced)
  - Library recertified in January
- 136 tapes damaged (75 tapes sent for recovery to lab)
  - 63 tapes fully recovered
  - 6 tapes partially recovered
  - 6 tapes still to be recovered
- Nearly all storage disk systems (and experiments) involved
  - 11 DDN JBODs
    - *RAID parity lost*
  - 2 Huawei JBODs (non-LHC experiments)
  - 2 Dell JBODs including controllers
  - 4 disk-servers

System	PB	JBODs	Involved experiments
Huawei	3.4	2	All astroparticle and nuclear experiments excepting AMS, Darkside e Virgo
Dell	2.2	2	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	ATLAS, Alice and AMS
Total	**23.9**	**9**	

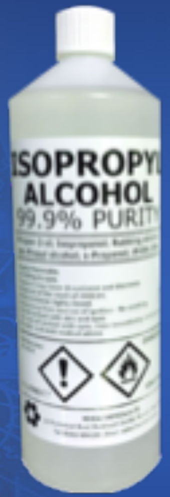
Storage recovery

- In parallel with the recovery of the power system, various activities performed to recover wet IT equipment
  - Cleaning and drying disks, servers, switches (using oven when appropriate)
- Apparently wet disks work after they are cleaned and dried (at least for a while….)
- Replacement components ordered only for systems still under support in 2018
  - DDN8 (LHCb) to be phased out in Q1 2018
- Moreover, some components not available for bulk replacement for old systems
  - i.e. disks for DDN8 (LHCb) and DDN9 (CMS) out of production
  - Other older systems (DDN1, DDN2 used in mirror as buffer) repaired with spare parts we had in house



Storage recovery

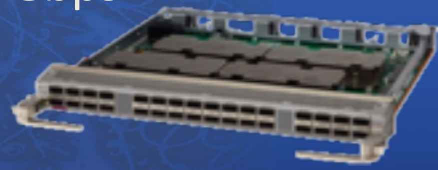
- In parallel with the recovery of the power system, various activities performed to recover wet IT equipment
  - Cleaning and drying disks, servers, switches (using oven when appropriate)
- Apparently wet disks work after they are cleaned and dried (at least for a while….)
- Replacement components ordered only for systems still under support in 2018
  - DDN8 (LHCb) to be phased out in Q1 2018
- Moreover, some components not available for bulk replacement for old systems
  - i.e. disks for DDN8 (LHCb) and DDN9 (CMS) out of production
  - Other older systems (DDN1, DDN2 used in mirror as buffer) repaired with spare parts we had in house



My wife’s comment on my concern for data recovery: “Why don’t you store your data into the cloud?” ☺

Storage recovery: an interlocking game

- Decided to move LHCb data to another storage systems and planned to use “good” disks from DDN8 to replace wet disks of DDN9 (CMS)
- To do this we needed to install 2017 tender storage to move there LHCb data
- But preliminarily we had to upgrade our network infrastructure (mid December) to support disk-servers for new storage (2x100 Gbps Ethernet connections)
  - Needed also for DCI to CINECA (remote farm extension) and OPN/ONE upgrade to 2x100 Gbps
- New 2017 storage delivered immediately before Xmas break
  - New storage installation completed in January
- Even if not validated we moved onto it all LHCb data from the damaged storage system ☺
- Later on, “good” disks from DDN8 used to replace wet disks of DDN9 (CMS) ☺



Storage recovery: not only good news

- Dell storage (Darkside and Virgo) easily recovered with the help of the support
  - After substitution of damaged controllers, wet disks switched on and replaced one by one to allow RAID rebuild
- Unfortunately we could recover only 1/3 of data on Huawei system (astroparticle experiments)
  - ~2.2 PB of data lost (mainly retransferred or regenerated)
- We suspect an erroneous strategy from the support
  - Damaged disks stayed switched on for days before the support decided what to do.....

Storage recovery: not only good news

- Dell storage (Darkside and Virgo) easily recovered with the help of the support
  - After substitution of damaged controllers, wet disks switched on and replaced one by one to allow RAID rebuild
- Unfortunately we could recover only 1/3 of data on Huawei system (astroparticle experiments)
  - ~2.2 PB of data lost (mainly retransferred or regenerated)
- We suspect an erroneous strategy from the support
  - Damaged disks stayed switched on for days before the support decided what to do.....

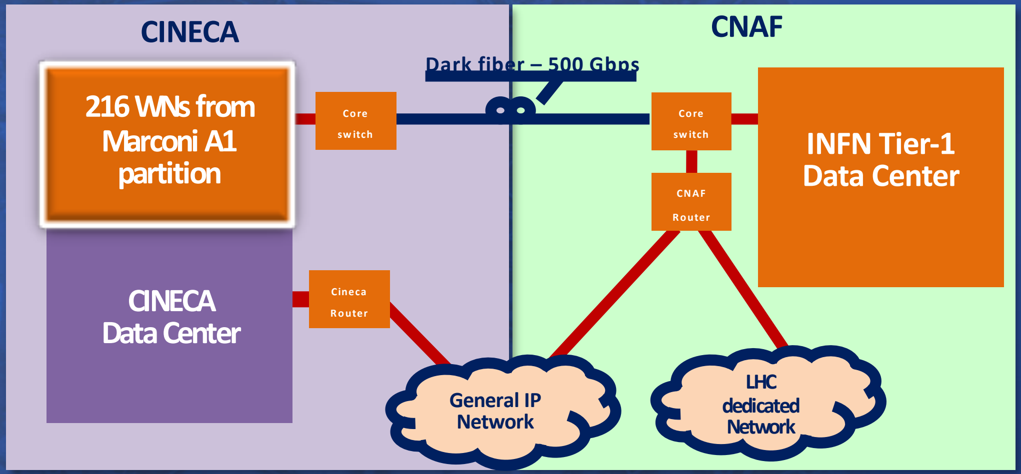
Murphy's law: unique data stored on this system only ….

Farm recovery

- During February we started reopening the services
  - LSF masters, CEs, squids etc…
- Not all experiments at the same time (depending on storage availability)
- Performed upgrade of WNs
  - Middleware, security patches (i.e. meltdown etc..)
- Only part of the local farm powered on (only 3 chillers in production)
  - ~150 kHS06 (out of ~200kHS06 available)
- But exploiting the CNAF farm extension to provide more computing power
  - Remote farm partition in Bari-RECAS (~22 kHS06)
  - Remote extension farm (~ 180 kHS06) at CINECA – In production since March

Farm remote extensions

- ~180 kHS06 provided by CINECA for 2018-2021
  - CINECA, located in Bologna too (~15 Km far from CNAF), is the Italian supercomputing center and Tier-0 for PRACE
  - 216 WNs (10 Gbit connection to rack switch and then 4x40 to router aggregator) managed by LSF@T1
- Dedicated fiber directly connecting INFN Tier-1 core switches to our aggregation router at CINECA
  - 500 Gbps (upgradable to 1.2 Tbps) on a single fiber couple via Infinera DCI
- No disk cache, direct access to CNAF storage
  - Quasi-LAN situation (RTT: 0.48 ms vs. 0.28 ms on LAN)
- In production since March
  - Slowly opening to non-WLCG experiments (CentOS 7)
  - Efficiency comparable to partition @CNAF



Lesson learned (1)

- Consider also low probability events
  - ~~In the project for our data center foreseen all possible incidents~~
  - In the project for our data center foreseen most probable incidents
    - E.g. fires, power cuts,…
  - The only threat from water was supposed to be intense raining, not a large pipe breaking with a robust water flow
    - Waterproof doors had been installed some years ago (after an heavy rain)
    - The municipal water supply company, apparently aware of the problem since it happened, needed several hours to stop the flow….

INFN

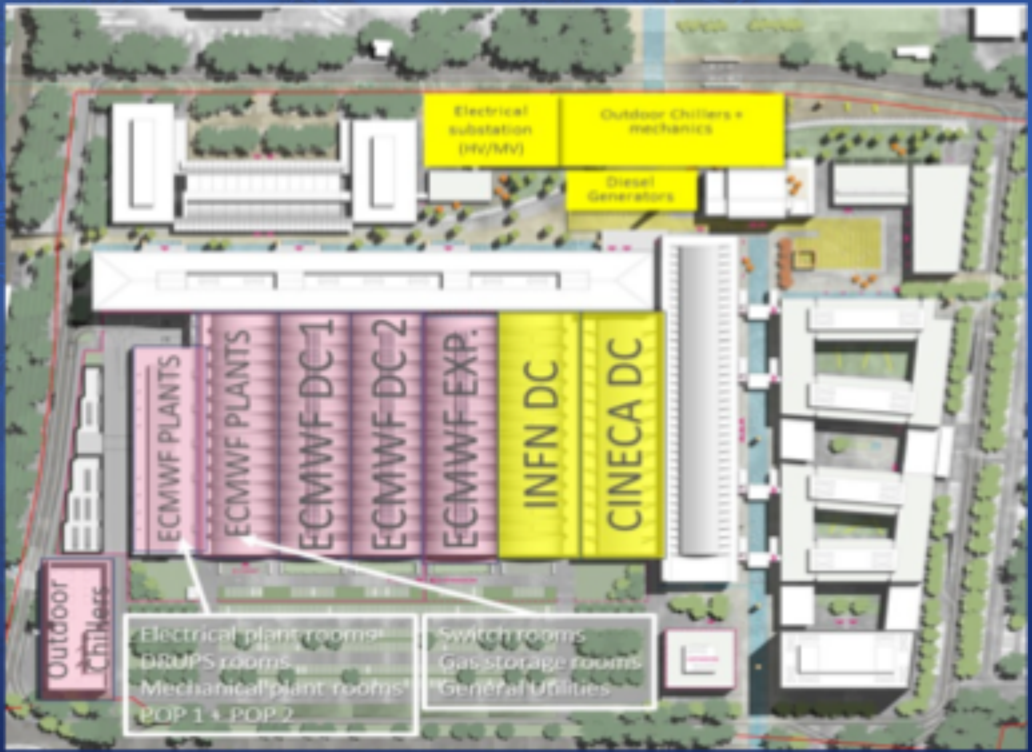
Lesson learned (2)

- Wet disks and tapes are not definitely lost: clean & dry them carefully
  - Disks should be powered on only for the needed time to copy the data
  - The cost of recovering a wet tape in lab is ~500 € to be compared with the estimated cost of reproducing its content (> 2 k€, not to mention the human effort)
- No experiment should base its computing on a single site
  - And, even worse, store the data in a place only….
    - Most probably this should be a feature implemented in the computing infrastructure
  - Some small collaborations had on user area even their official code
  - In fact WLCG experiments could compensate "easily" using other sites

A possible future for the INFN Tier1: towards the HL-LHC Data Lake

Looking for a new location for the Tier-1

- The plan: take into account the needs for HL-LHC (i.e. data lake) and expansions due to astroparticle experiments
- This plan has become more urgent after the flood
- An opportunity is given by the new ECMWF center which will be hosted in Bologna from 2019 in the new Technopole area
- Possibility to host in the same area also:
    - INFN Tier-1
    - CINECA computing center
- Funding promised by Italian Government to INFN&CINECA to set up the data center (2021)
    - Up to 2x10 MW for IT (2026)
    - PUE ~ 1.2



Conclusions

- INFN Tier-1 fully operational since March
  - Some hiccups at the restart
- Some systems not completely recovered yet
  - Still working on 2nd power line (needed for redundancy)
  - Strategy for continuity on the 2nd line not decided
- We are currently redesigning our alarm system to really be reactive in case of flood
- In the short term also activity ongoing to improve the isolation of the data center perimeter
- We plan to move, in the medium term (2021), our data center to another location
  - Also to take into account requirements for HL-LHC era

Practicing for next time ☺