

# Cost modelling of tape storage

*System Tape and Archival WG meeting  
2019-01-18  
Renaud Vernet, CCIN2P3, Lyon, FR*

**How to store massive amounts of data in an efficient and performant way, keeping TCO as low as possible ?**

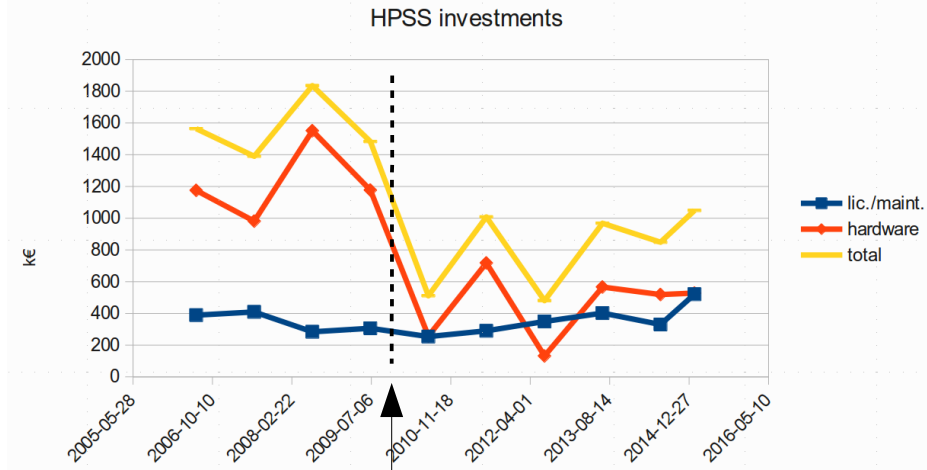
- Need metrics about performance & expenses in particular
- Here we focus on **expenses**
- Current activities within System Performance & Cost modelling WG

- From the literature
  - Market trends of tape cartridge densities and costs
  - Their mid-term future roadmaps provided by companies
- In practice, in data centers (sites)
  - Observed costs (robots, drives, licences, maintenance...)
  - Access to information is uneasy
- Get some global picture
  - → then forecasts for HL-LHC become possible

**Need to understand current costs  
and evolution over time**

# First attempt @ CCIN2P3 (2014)

- Dug into procurements history
- Licence + maintenance ~ constant over time
- Hardware (robots, drives, movers, tapes) : not so much
  - Early expenses = new robots+drives = initial investment
    - Need to spread over some time
  - Later times = expenses ~ constant



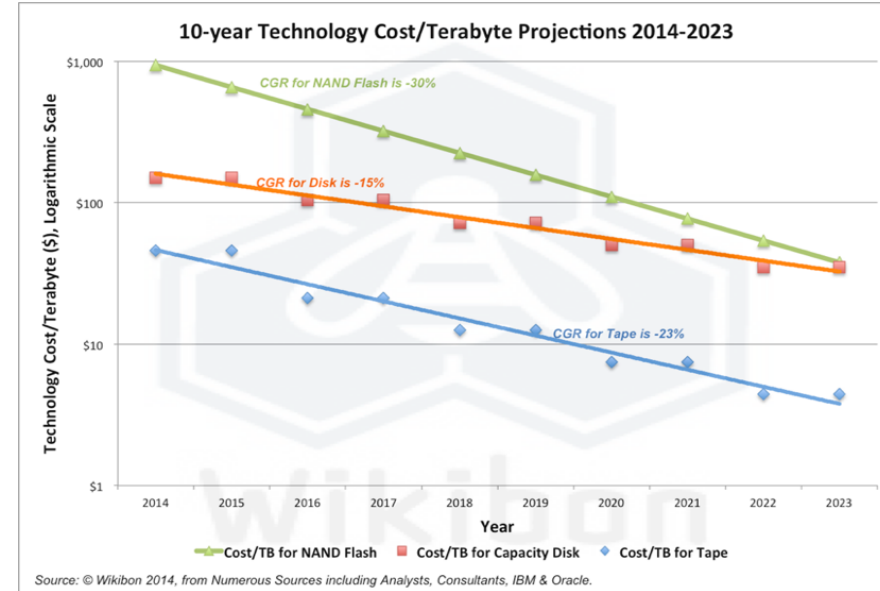
**Need model with constant expense year after year**

# Tape growth trend

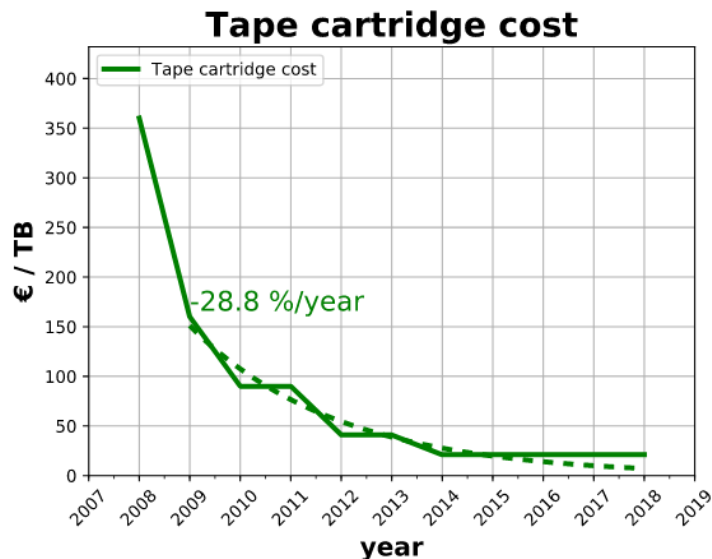
- Cartridge cost drops exponentially with time
- → online capacity grows exponentially
  - At constant budget
- Hey, I still have to purchase new drives, pay for my maintenance etc. !

Constant :  $f = \frac{\text{yearly budget for cartridges}}{\text{yearly budget for total tape system}}$

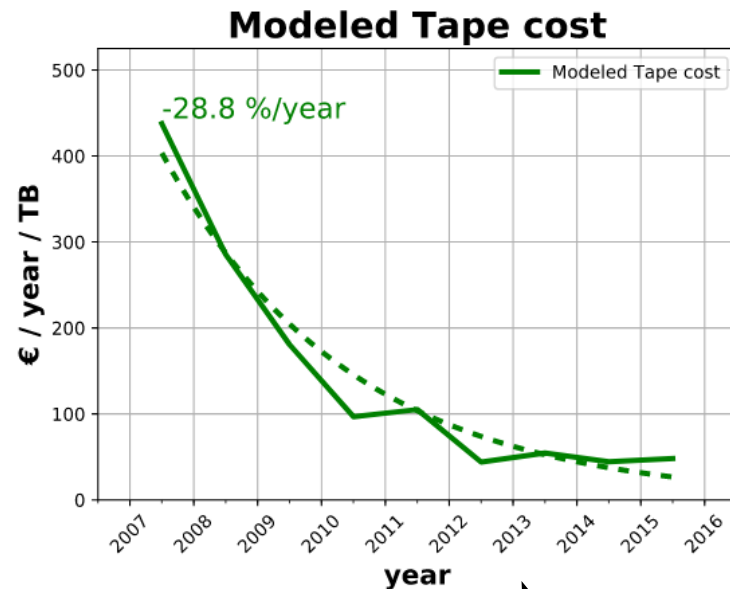
**Annual growth rate does not depend on  $f$**   
It is **only** driven by the cartridge cost decrease rate



# Tape cost modelling results (CCIN2P3)



Evolution of cartridge cost at procurement time

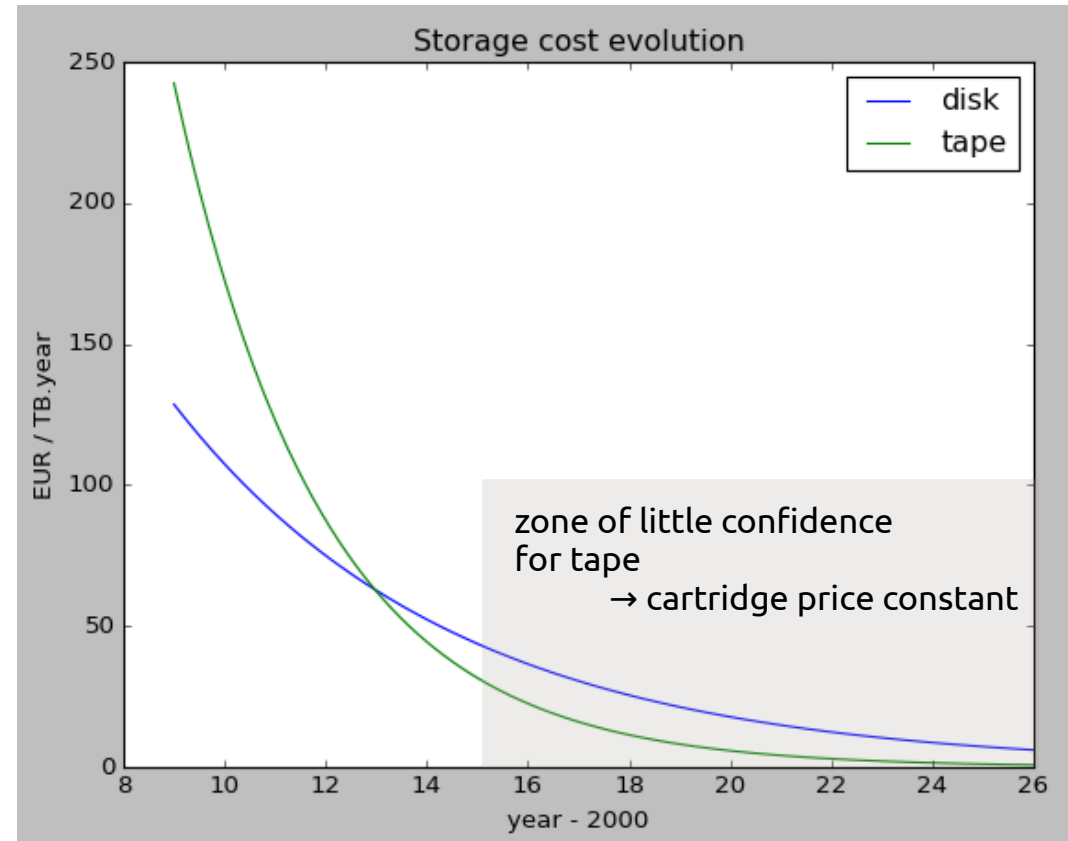


Full tape system yearly cost

$$\text{yearly TB cost} = \frac{\text{yearly budget}}{\text{online capacity} \times \text{cartridge lifetime}}$$

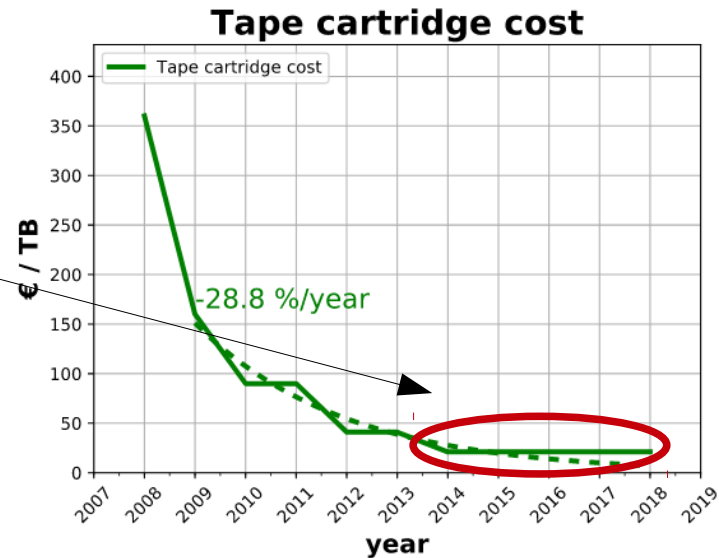
# Comparison with disk cost (at CCIN2P3)

- Before 2013, our tape system was more expensive (per online TB)
  - Including licence, maintenance, libraries, drives, cartridges
- Cost decreases faster for cartridges than for disk  
→ now, tape is cheaper
- Talking only about money here
  - Regardless of performance & other aspects



# Current situation

- CCIN2P3 Running Oracle StorageTek now
  - No more development from Oracle
  - Cartridge price steady for 4 years
- « Phase transition »
- Need input from Tier-1's
- Survey started last September





# Survey (snippet)

TAPE STORAGE (if applicable)

=====

\* How much do you pay on average for 1 TB of tape cartridge as of today?

-> can you please provide numbers related to the investment costs other than cartridges in your tape system, over a given period of time? That may include disk buffer, libraries, drives, maintenance, license, connectivity or any other significant expense. Please provide detail of those expenses.

-> can you please provide the evolution with time of these numbers over the last years?

\* Can you provide indicators on the provision of tape storage technology in your site for the next years: do you plan any significant investment or change? If yes, what are the reasons for that?

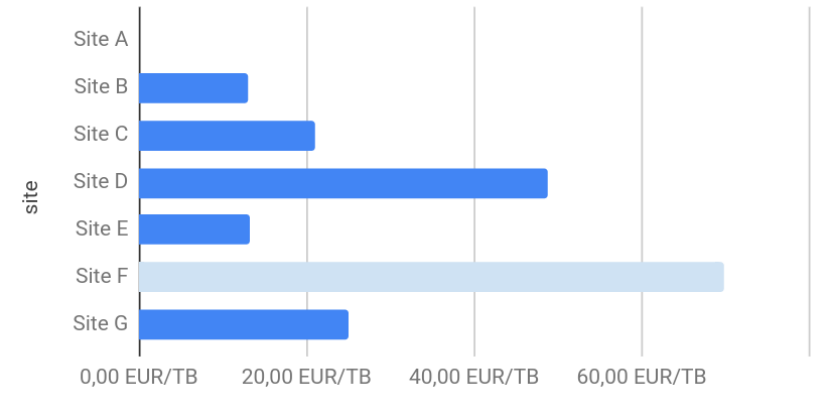
# Results (summary)

- All Tier-1's contacted
  - 6 did provide numbers about tapes
  - Others have NDAs, no time or no interest (?)
- Setups & fees vary
  - (tape techno, software...)
- National contexts
  
- Try to extract common metrics from diverse environments
  - Cartridge price per TB
  - Full cost per TB
  - Evolution with time of media price

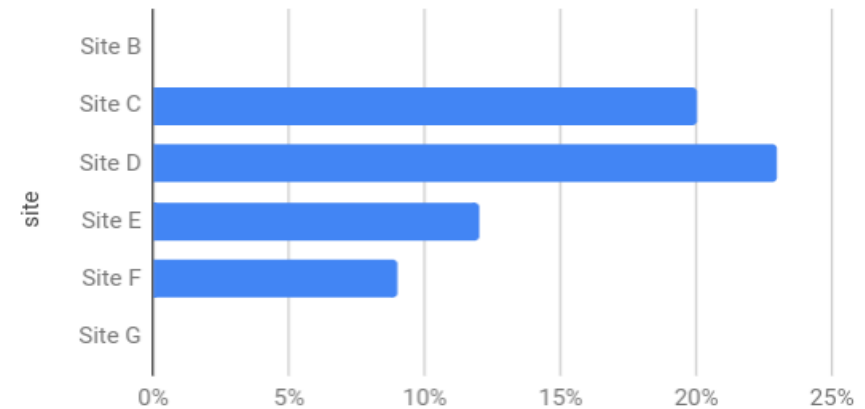
# Cartridge cost & evolution

- Some variety in cartridge price
  - Depends on techno
  - Local contexts (more important)
  - Forget about site F
- evolution over last years
  - -10 %/year .... -23 %/year
  - not as fast as seen in literature
  - Not-so-big-data-center effects
  - Situation at CERN may be better

Tape cartridge cost (2018)



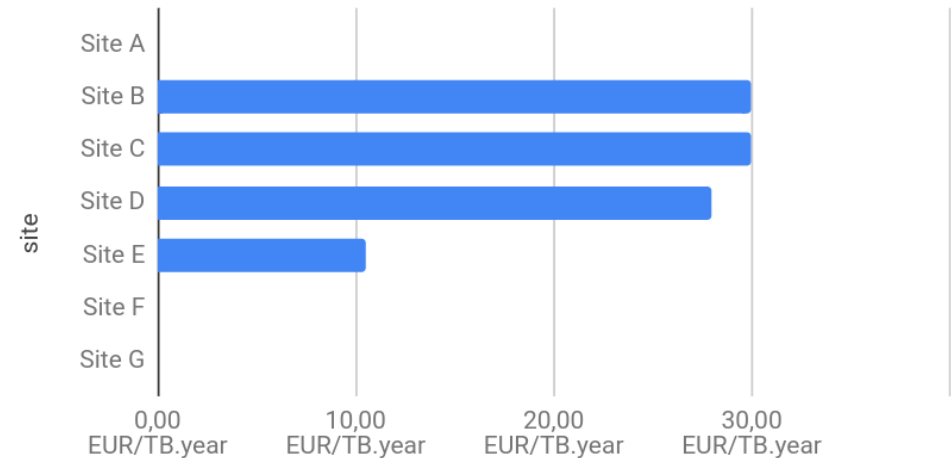
Cartridge cost yearly decrease rate



# Full tape system yearly cost

- Surprisingly similar for 3 sites
- Site E has very low costs !
  - May not represent global situation (?)
- Lacks statistics
- More sites invited to provide numbers

Tape full yearly cost (2018)



→ Have some estimate about tape full costs and trends  
→ Uncertainties remain large

# Summary

- Modelized tape system expenses
  - ~ constant budget
  - Predictions on capacity growth rate for near future
- Numbers in literature do exist
  - Translation to our community of grid sites is not straightforward at all
  - Local contexts, vendors
- Survey important to understand what sites pay in practice
  - More participating sites very welcome
  
- Next milestone
  - Combine numbers
  - Provide global cost & sigma estimates
  - Preliminary prospects for mid-term future