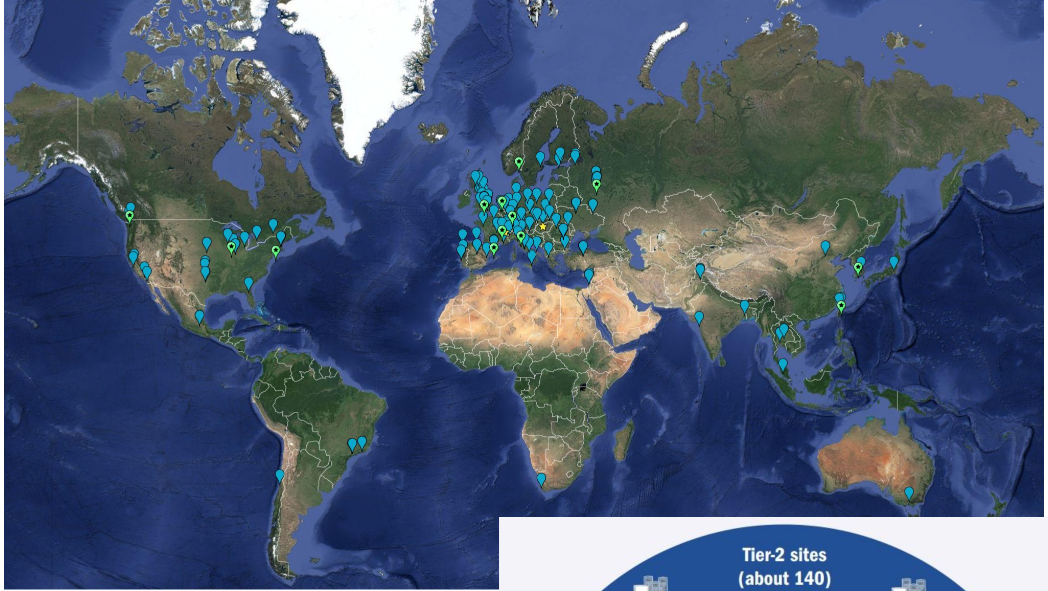
Prospects for LHC computing in Spain for the next 3 years

José M. Hernández
CIEMAT, Madrid

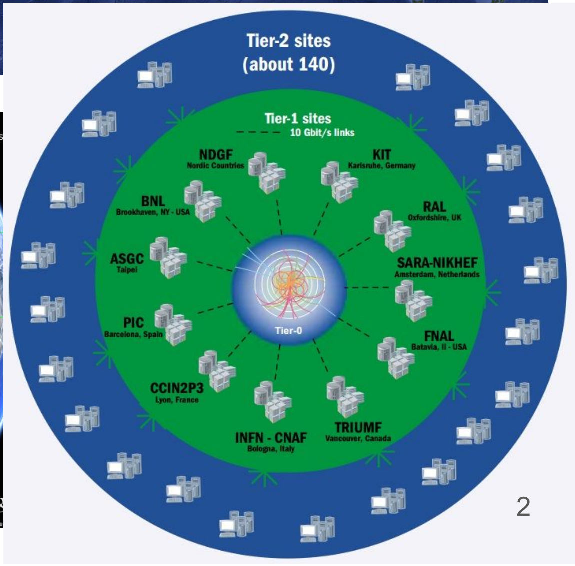
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas
cfp CIEMAT física de partículas

Worldwide LHC Computing Grid (WLCG)



Distributed high-throughput computing infrastructure to store, process & analyze data produced by LHC experiments

- 167 sites, 42 countries, 63 MoU's
- ~1 million CPU cores
- ~900 PB disk storage
- ~1500 PB tape storage
- Optical private network (LHCOPN) and overlay over NRENs (LHCONE) with 10/100 Gbps links
- ~Tbps LAN bandwidth between compute and storage nodes at sites



Major features and capabilities of LHC computing infrastructure

- Networks
  - International and national, private and public
- Data management
  - Key to success, data transfers, storage systems, data management tools and data organization
- Compute
  - Provision of resources and workload scheduling, execution and monitoring
- Security
  - Authentication and authorization, federation, single sign on, etc
- Operations support
  - Security, incidence response, problem tracking, daily operations, upgrade campaigns
- Diverse experiment-specific services and tools, applications

Distributed data-intensive high throughput computing (HTC)

Precursor of Big Data processing and Cloud computing

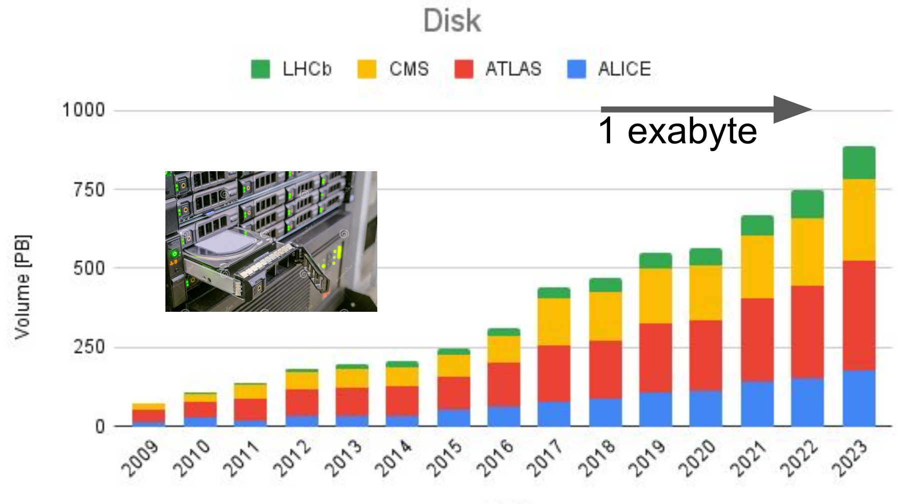
WLCG Computing resource evolution

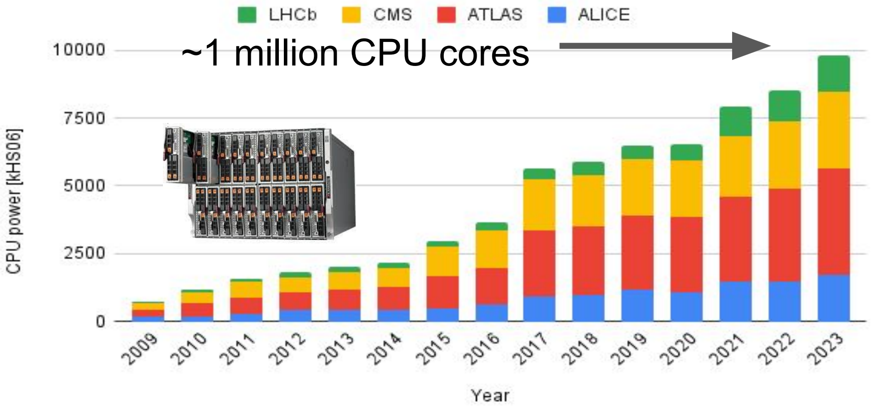
Currently:

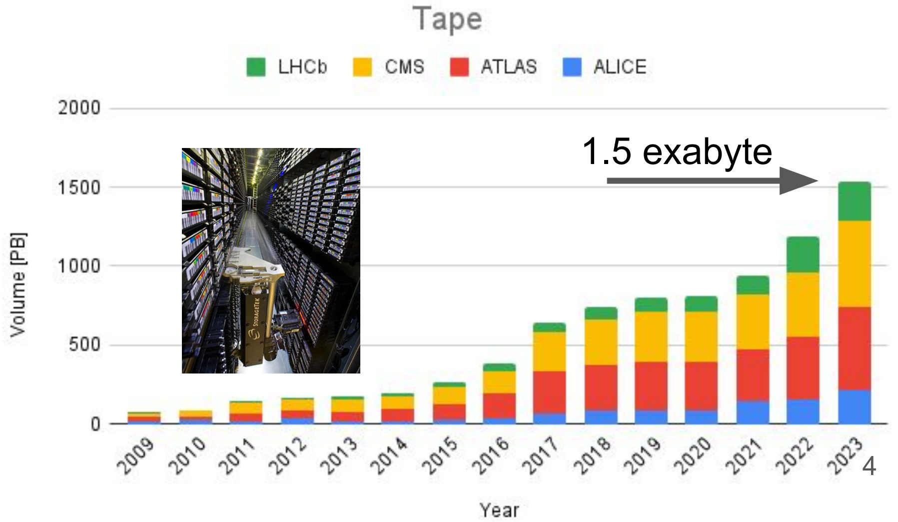
~1M CPU cores, ~2.4 exabyte storage

~15% annual growth with ~flat funding

(hardware getting cheaper)



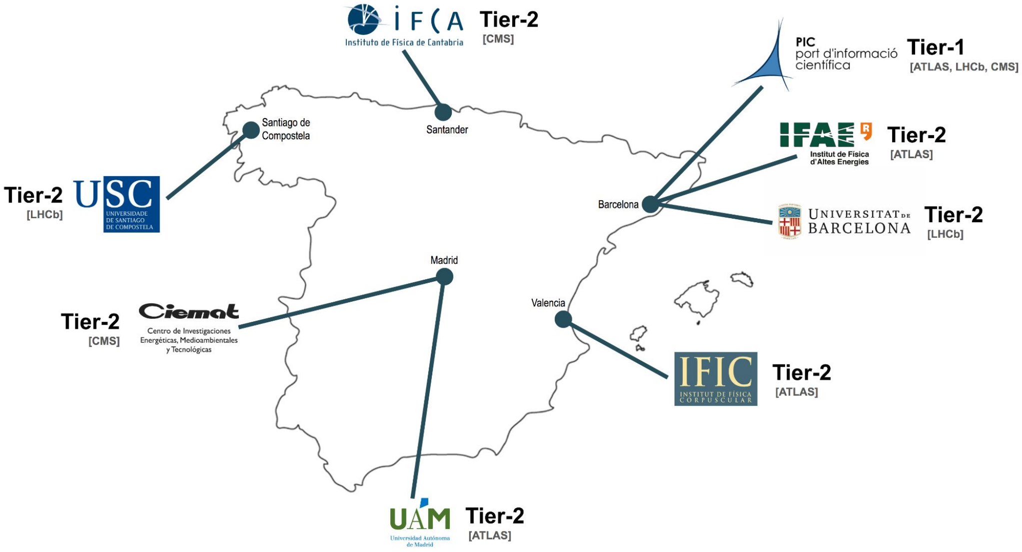




Spain in WLCG

Spanish contribution:

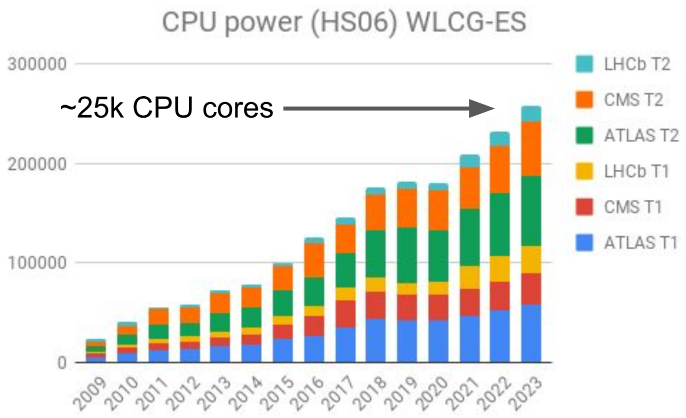
- ~5% resources T1 & T2 (MoU)
- 1 Tier-1 center (PIC, CIEMAT-IFAE)
- 6 Tier-2 centers
  - CMS: CIEMAT-Madrid (75%)
    IFCA Santander (25%)
  - ATLAS: IFIC-Valencia (60%)
    IFAE-Barcelona (25%),
    UAM-Madrid (15%)
  - LHCb: USC-Santiago
    UB-Barcelona (decommissioned)

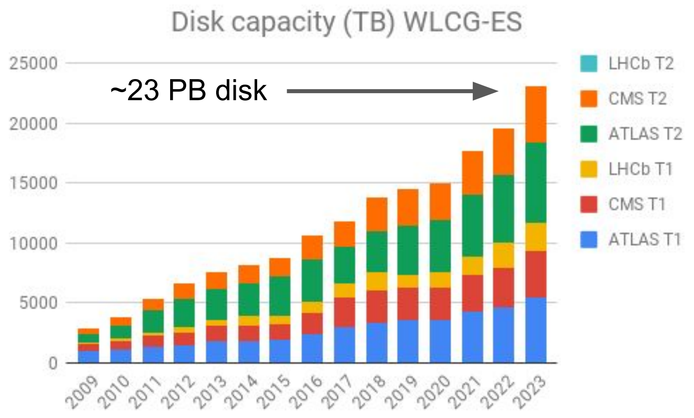


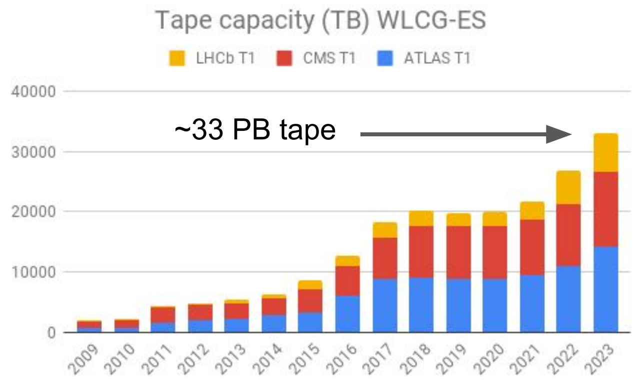
WLCG-ES: a success story

- Two decades contributing to LHC distributed computing infrastructure WLCG and R&D at the highest level
  - ~5% of WLCG resources (20k CPU-cores, 15 PB disk, 20 PB tape), ~1500M CPU hours delivered since 2004
  - Providing 1 of the 13 Tier-1 sites worldwide (PIC)
  - Federated Tier-2 sites for ATLAS (IFIC, IFAE, UAM), CMS (CIEMAT, IFCA), LHCb (USC, UB)
  - Among the most reliable sites in WLCG
- A large effort from HEP community and institutions
  - ~26 M€ funding (direct costs) from HEP national program since 2001
  - Funding from institutions of the same order
    - Funding personnel, electricity, infrastructure
- Large community of experts in distributed high throughput computing
  - Contributions to LHC computing, development, integration, operations, management
  - **Technology transfer** to other projects in HEP/astro/cosmo (CTA, MAGIC, DUNE, DarkSide, PAU, Euclid, Virgo, etc)
  - **We have generated a big strategic asset for our community!**

Resources provided by Spain to WLCG







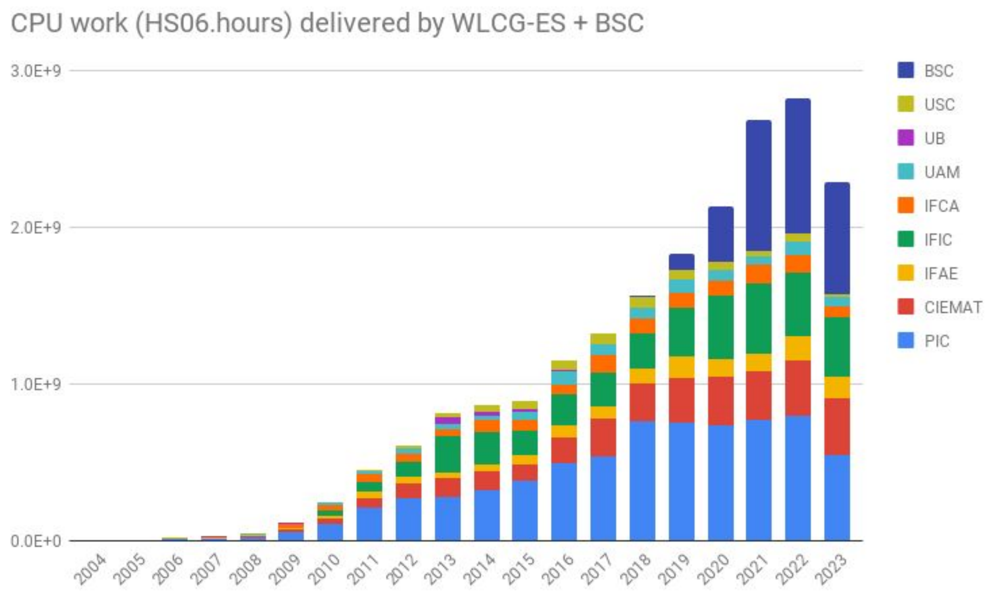
Barcelona supercomputing center & LHC computing

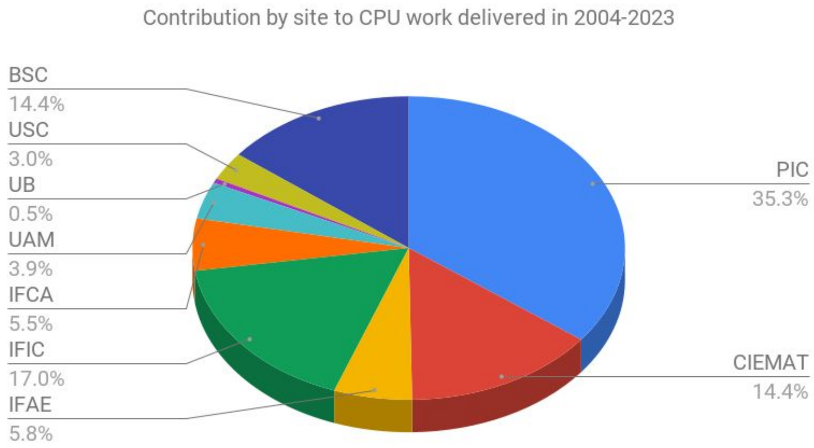
- The BSC is the largest supercomputing center in Spain
  - MareNostrum4 (150k CPU cores); MareNostrum5 10x larger (expected from 2023)
- BSC - WCLG-ES agreement
  - LHC computing designated as a BSC “strategic project” in 2020
  - Access to dedicated resources (up to 7% of MareNostrum4)
  - Providing CPU for LHC simulation (~60M hours/year, ~50% of WLCG-ES CPU)

Exploiting supercomputers for LHC

- Data intensive computing with HPC facilities is a **challenge**
  - Limited/no network connectivity in compute nodes
  - Limited storage for caching input/output event data files
  - **In practice only run CPU-bound workflows (MC simulation) with little I/O**
- Our applications are not really suited for HPC
  - No large parallelization (no use of fast node interconnects)
  - No substantial use of accelerators (GPU)
- Substantial **integration** work to make HPC work for HTC
  - No one-fit-all solution: each facility is different
  - Little effort available in the LHC experiments; in charge of the local communities
  - Experiments do not accept capacity from HPCs as pledged resources unless they can be used transparently as any other WLCG site
- Not suitable resource **allocation** model
  - We would need a guaranteed share of resources rather than apply for allocations

CPU delivered by Spain to WLCG

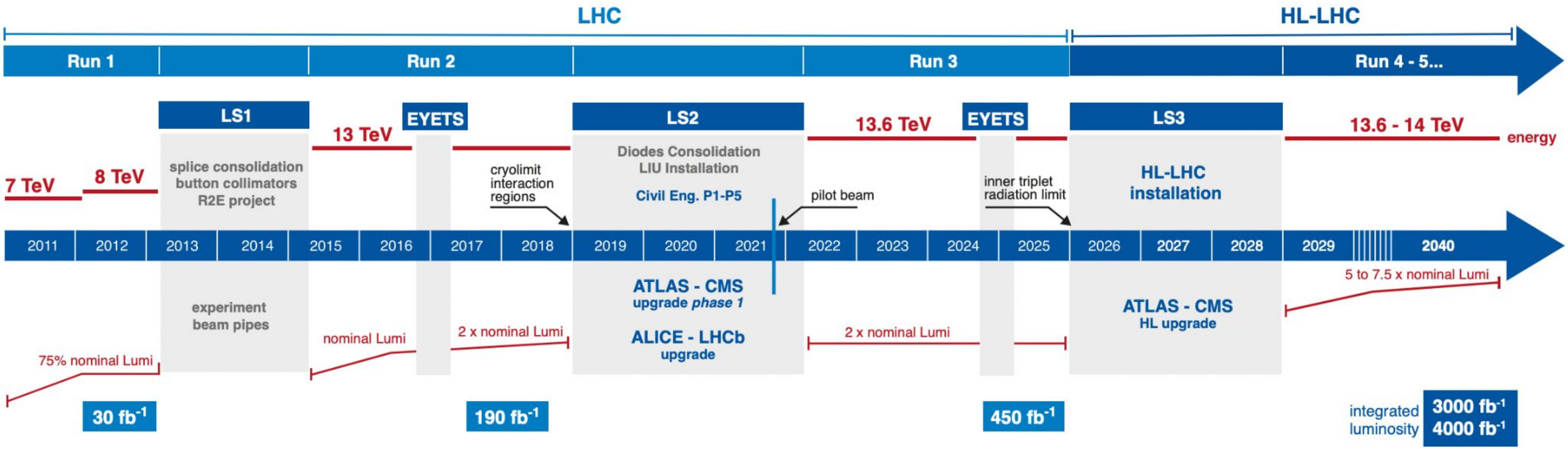




~250 Million hours delivered in 2022
~2000 Million hours delivered during 2004-2023

(20×10^9 HS06.hours; average CPU core power ~10 HS06)

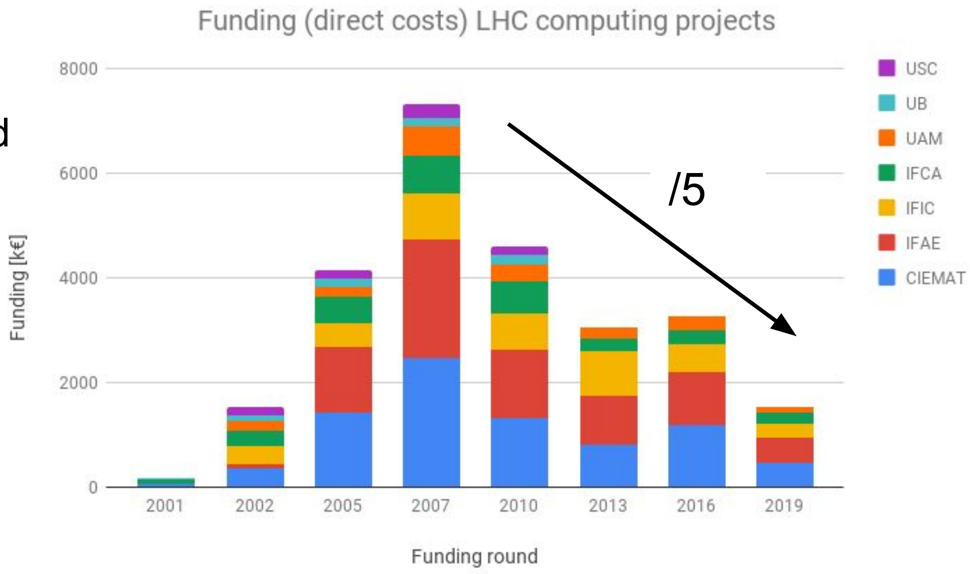
LHC / HL-LHC plan



- Run 3 (2022-2025): **~2x more data**. **Evolutionary** changes in computing models
- Run 4 (HL/LHC, 2029+): **~20-30x more data**. **Revolutionary** changes required

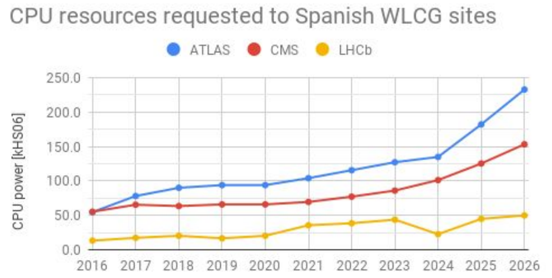
WLCG-ES sustainability challenge

- Decreasing funding from HEP national programme
  - From ~2.5M€/year in 2007 to ~0.5 M€/year in 2019
  - WLCG-ES contribution to LHC reduced from 5% to 4% to 3%
  - Aging equipment (~50% > 5 years)
- Big effort to complement funding and resources
  - From institutions
  - From national/regional scientific infrastructure calls
  - Agreement with BSC (50% CPU)
- Required funding ~1.5M€/year
  - Hardware resources and personnel



WLCG-ES funding for 2024-2026

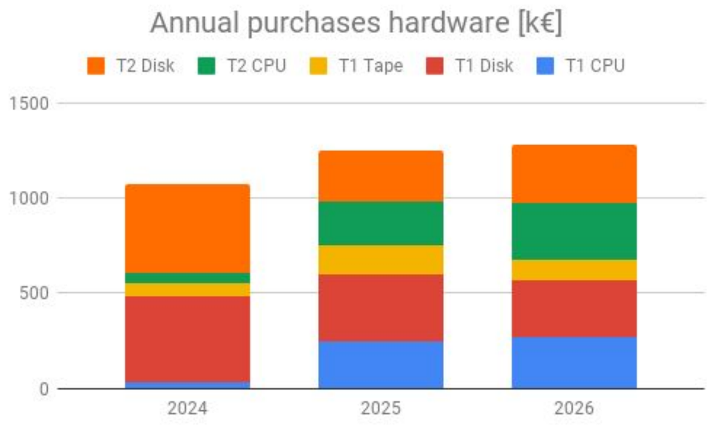
- Model for computing resource costs
  - Resource requests ATLAS, CMS, LHCb
  - WLCG-ES contribution
    - 2024: 4%, 2025: 4.5%, 2026: 5%
  - Price evolution of CPU, disk, tape
    - 10-20% reduction per year
  - Age of existing equipment
    - 5-year replacement policy
  - Fraction of CPU provided by BSC
    - 50% of total CPU
  - Contribution of institutions
    - CIEMAT 50% T1, 75% CMS T2
    - IFAE 50% T1, 25% ATLAS T2
    - IFIC 60% ATLAS T2
    - IFCA 25% CMS T2
    - UAM 15% ATLAS T2
    - No funding for LHCb T2

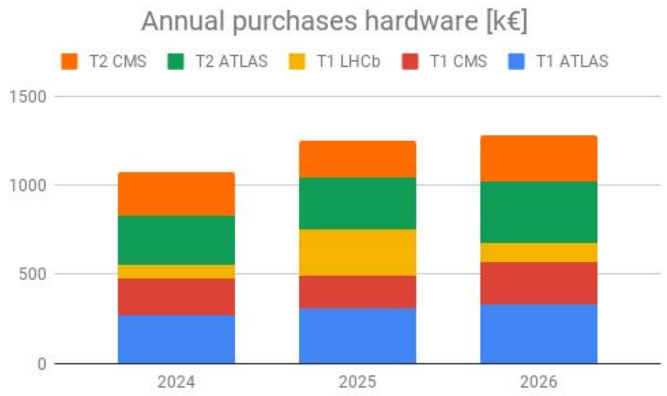


WLCG-ES funding for 2024-2026

Total cost of hardware: 3.6 M€
Total cost of additional personnel: 0.75 M€

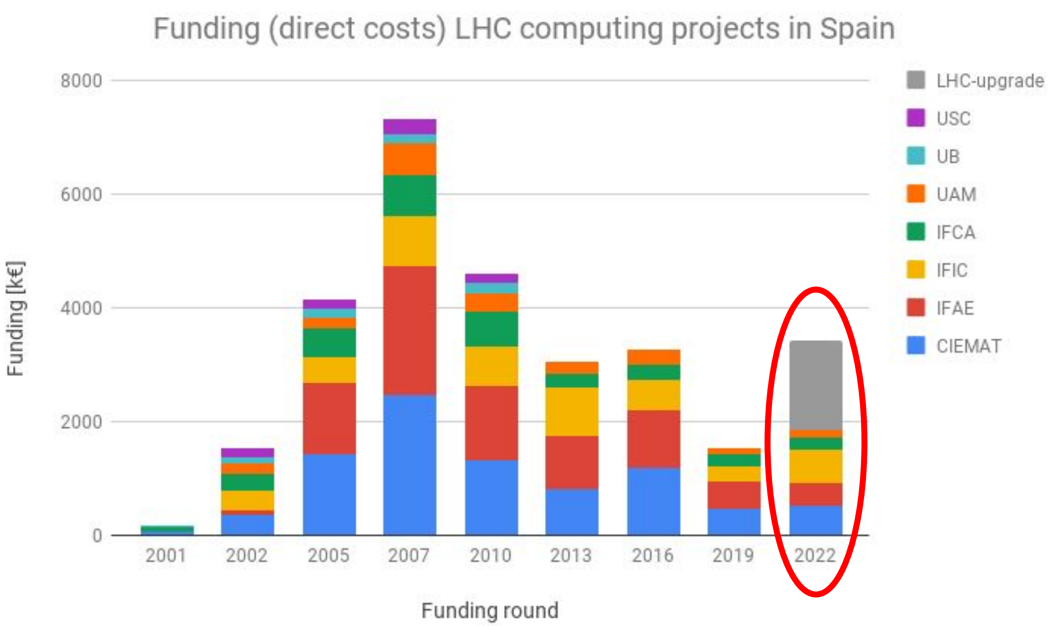
T1	1,971
75% T2 CMS	537
25% T2 CMS	179
60% T2 ATLAS	552
25% T2 ATLAS	230
15% T2 ATLAS	138
Total	3,607
CIEMAT	38.0%
IFAE	38.0%
IFIC	15.3%
IFCA	5.0%
UAM	3.8%





WLCG-ES funding for 2024-2026

- Funding received in PID2022: 1.85 M€
- Complemented with LHC-upgrade MRR funds for the Tier-1: 1.6 M€
- Funding enough for hardware resources
- Funding sustainability is a key question



Computing R&D&I activities in Spain

Spain not only contributes computing resources to WLCG, but also participates in computing development projects and in the operation and coordination of the distributed computing infrastructure

Development projects:

- Deployment of data Xcache (CMS)
- Integration of supercomputing resources (ATLAS, CMS, LHCb)
- Deployment of Analysis Facility (ATLAS, CMS)
- Evolution of workload management system (ATLAS, CMS)
- Event Index (ATLAS)
- Distributed Physics Analysis (ATLAS)
- Data preservation (CMS)
- Application of machine learning in HEP
- Technology transfer: ML for oncology, analysis of medical imaging, etc

Coordination positions:

- Workflow management coordination (ATLAS)
- Distributed Physics Analysis coordination (ATLAS)
- Event Index data production and collection coordination (ATLAS)
- Resource office coordination (CMS)
- Workflow execution coordination (CMS)
- Computing resource board chair (CMS)
- Grid deployment board chair (WLCG)
- Statistics committee chair (CMS)

HL-LHC computing challenge

- The foreseeable future funding is ~flat
  - ~10-20% annual increase from hardware getting cheaper
  - Not enough with current computing models to face HL-LHC data
- Intense R&D to drastically reduce computing resource needs for HL-LHC
  - **Less data**
    - Reduce data replication levels
      - Requires new infrastructure; data lakes, content delivery network with caching layer, aggressive use of tape technology
    - Reduce data formats for analysis, O(kB)/event
      - Requires good understanding of detectors and ~continuous reprocessing
  - **Less CPU**
    - Software improvements
      - Speed up and parallelize algorithms
    - Use additional hardware resources
      - GPUs and supercomputing resources

Analysis facilities

- Guiding principle: Help physicist minimize time-to-insight, enabling iterative exploration of data
  - In the future, when processing 10x lumi/evts, avoid physicists 10x waiting time!
  - Boost productivity and competitiveness of our physics communities
- Key features:
  - Local access to the reduced data samples with low latency from compute
  - Processing resources available with capacity and priority (CPUs, GPUs, Memory).
  - Efficient and elastic infrastructure (interactive of job scheduling in the batch system) with dynamic expansion to HPC/Cloud to absorb peaks
  - Enhanced support to software tools (ROOT + Python ecosystem).
  - Enable and encourage use of common repositories of code for analysis routines/workflows
  - Expert data/code manager: critical liaison role (neither a final user nor an infrastructure expert, however facilitating technology/access)

Summary

- Two decades successfully contributing to LHC distributed computing
  - Solid infrastructures and large community of experts in data-intensive computing
  - **Technology transfer** to other projects in particle physics, astroparticle and cosmology, and beyond!
- Funding sustainability is a key issue
  - Computing infrastructure is an extension of the detectors, needs to be funded as such
  - After a decade of decreasing funding, the trend has changed, funding for LHC Run-3 ~sufficient
- New Challenges towards the HL-LHC phase
  - Large increase in data volume and complexity
  - Optimization of data infrastructure (data lakes, content delivery network)
  - Software R&D (parallelization, artificial intelligence, performant data analysis)
  - Heterogeneous computing resources (supercomputers, accelerators, non-x86 processors)