Advanced Computing Technology

José M. Hernández

CIEMAT Scientific Computing Unit: Our **mission**

Enable and accelerate science with computing technology

Infrastructure and support

- for particle physics, astroparticle and observational cosmology projects in the department
- Transversal research line to the scientific research lines

Data-intensive distributed computing

- Massive storage systems, High-throughput workload management frameworks
- Development and integration of Grid/Cloud/HPC Computing for HEP

Highly parallel processing and artificial intelligence

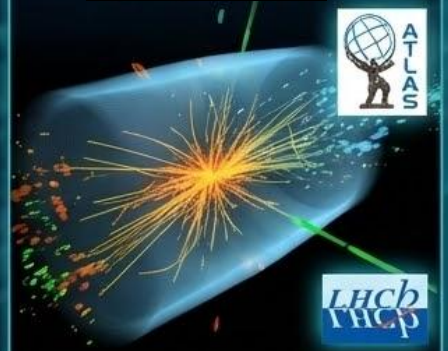
- Highly parallel computing techniques (GPU accelerators)
- Machine learning techniques (neural networks, data mining)

- We provide **hardware, software** and **human resources**
  - Specialized personal (development, integration, operations, support)
  - Data centers in Madrid (CIEMAT) and Barcelona (PIC)

Computing infrastructure and support for fundamental physics

Particle Physics

- **LHC:** CMS, ATLAS & LHCb, Tier-1 and Tier-2 Analysis Facility



Underground & Astroparticle

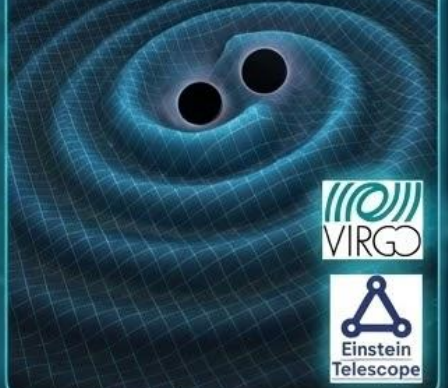
- **Neutrinos:** DUNE
- **Dark matter:** DArT, DarkSide, Deap-3600
- **Astroparticle physics**
  - Charged cosmic rays: AMS
  - Gamma rays: CTA (one of four data centers) MAGIC (Tier-0)

Cosmology

- **PAU** (main data center)
- **Euclid** (Spanish Science data center)
- **Rubin/LSST** (Data Access Center)
- **DESI** (European data mirror)

Gravitational waves

- **Virgo**-LIGO
- **Einstein** Telescope



Infrastructure

PIC-Barcelona

- WAN network: 2x100 Gbps
- CPU: 21k cores (340 kHS23)
- Disk: 60 PB
- Tape libraries:
  LTO8/09: 7000 slots (80 PB), 25 drives
  LTO10: 5000 slots (150 PB), 21 drives

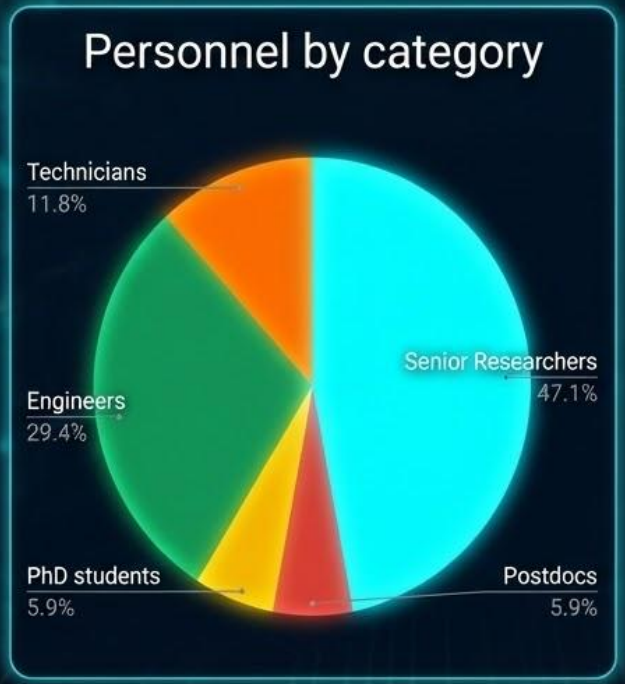
CIEMAT-Madrid

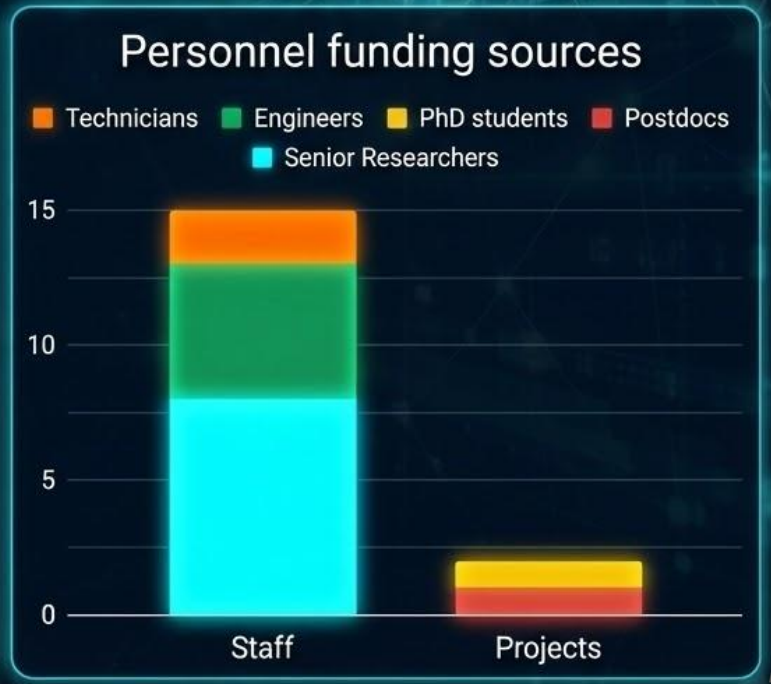
- WAN network: 2x10 Gbps
- CPU: 6k cores (60 kHS23)
- Disk: 8 PB

People

Personnel is our **most important asset**

- 17 people
- Recent staff stabilization process





Highlights in the mid 2023 - mid 2025 period

MareNostrum 5 Integration (EuroHPC)

CIEMAT led the integration of MareNostrum 5 at Barcelona Supercomputing Center (BSC) into the CMS computing system, enabling large-scale data processing and simulation through innovative network and storage solutions.

CMS Analysis Facility at CIEMAT

An innovative high-performance computing infrastructure was deployed at CIEMAT to support modern LHC data analysis, combining optimized hardware, advanced software, and expert user support within the CMS ecosystem.

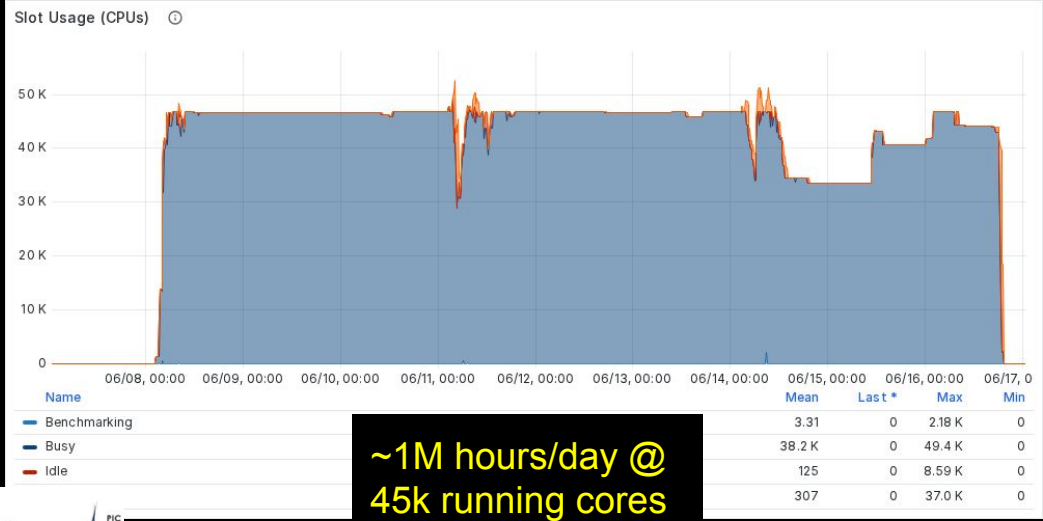
CosmoHub Platform Recognition

The PIC-developed CosmoHub platform was recognized by the European Open Science Cloud macroroadmap in 2024 for its pivotal role in enabling accessible, scalable, and collaborative data analysis for the scientific community.

Barcelona Supercomputing Center & LHC Computing

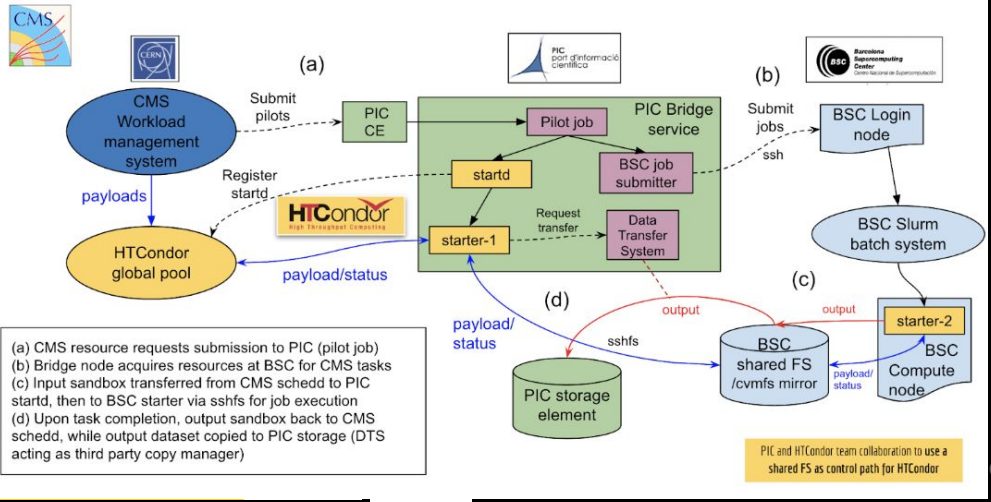
- Spain's Largest Supercomputing Center
  - MareNostrum4 (150k CPU cores)
  - MareNostrum5 (750k CPU cores, from 2024)
- BSC - WCLG-ES Agreement (2019)
  - LHC computing as a BSC "strategic project"
  - Access to dedicated resources (up to 7% of MareNostrum4)
  - Providing CPU for LHC simulation (~50% of WLCG-ES CPU, ~100M hours/year)

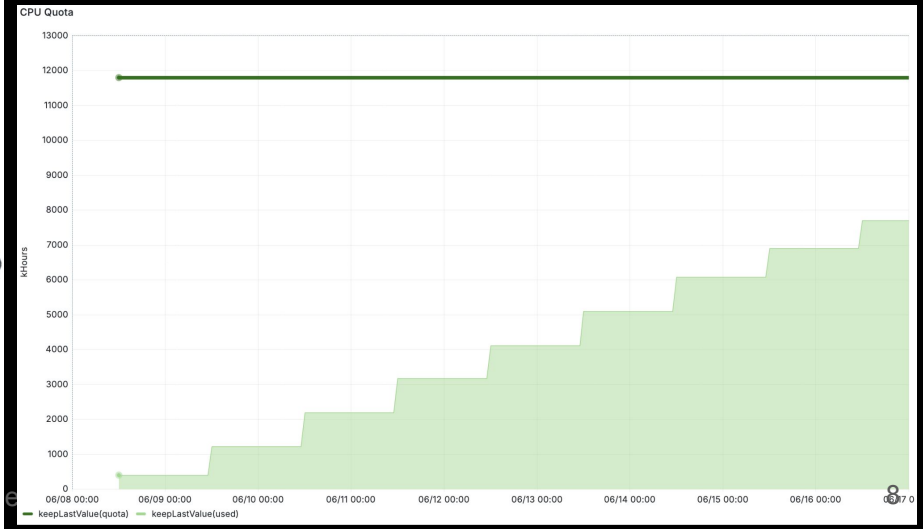
MareNostrum5 exploitation



~1M hours/day @ 45k running cores

BSC use by CMS: intense R&D programme





CMS Analysis Facility

The LHC Challenge & Current Workflow

LHC Run3 & HL-LHC Phase Challenge

- Higher pile-up, higher trigger rate, more Monte Carlo simulated events

Analysis at HEP & The Laptop Bottleneck

- Iterative data reduction to manageable dataset
- Growing dataset sizes (15-20x) demand new paradigm

The New Paradigm & Software Evolution

The High-Performance Local Facility

- Key objectives: ease of use, performance, scalability, sustainability

Evolving Software & Paradigms

- New centrally produced reduced data formats (NanoAOD)
- Modern programming interfaces, trivial parallelization
- Growing use of Python, Jupyter notebooks, Machine Learning

CMS Analysis Facility

PID2020-113807RA-I00, Servicios innovadores de análisis de datos para el experimento CMS, 09/2021 - 08/2025

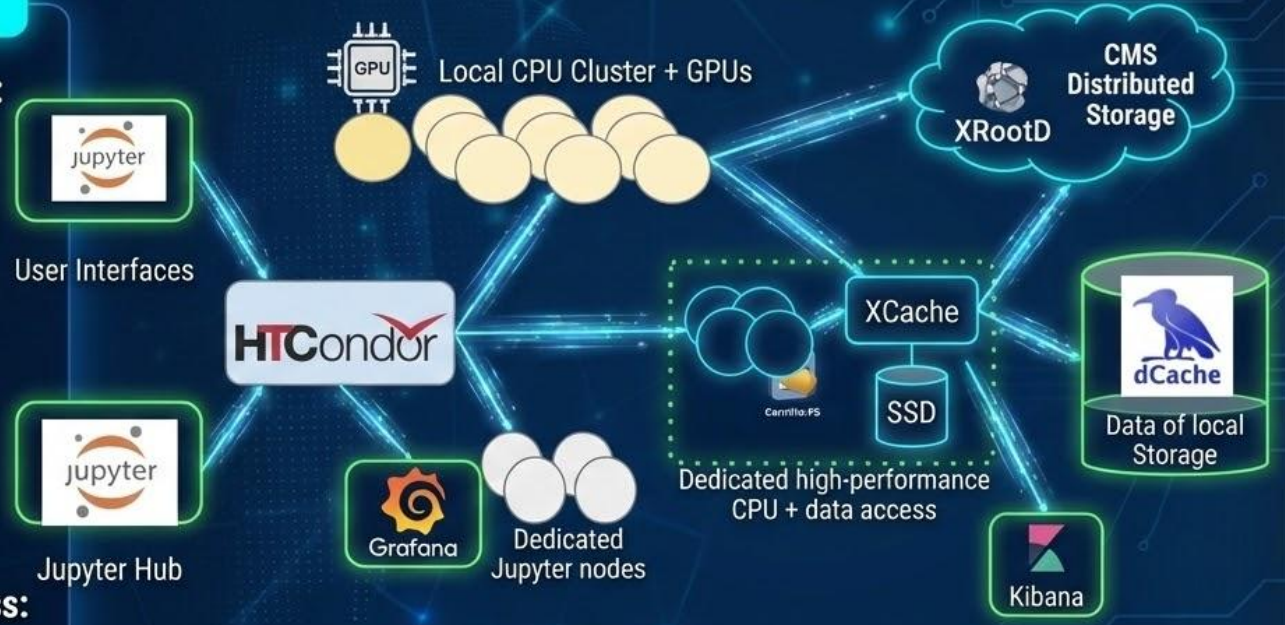
Facility Components

High Throughput Storage:
- 400 TB NVMe
- XCache

GPU Servers:
- 4x Nvidia A100
- 4x Nvidia H200

Dedicated CPU Batch:
- 400 + 576 CPU cores

Interactive & Batch Access:
- Jupyter & Batch



PhD Thesis: Cecilia Morcillo (Supervisors: Antonio Delgado & Antonio Pérez Calero) | To be presented in Q2 2026 at UCM

CosmoHub

CosmoHub.pic.es

eosc.eu/use-case/cosmohub

Funding source	In-Kind value	Timeframe	Target group	Scale	Year of reporting	Country
National, Regional	€500k–€1M	2013–ongoing	Researchers	International	2024	Spain

Case study

CosmoHub is a web-based service designed to facilitate an interactive exploration, visualisation, and distribution of large astronomical datasets. Powered by a highly specialised big-data infrastructure, it supports users while providing excellent performance. CosmoHub officially collaborates with numerous European and international projects, such as Gaia, Euclid, PAUS, DES, MICE, LSST and CTAO.

Added value

Providing a **centralised open data repository,** adhering to FAIR principles.

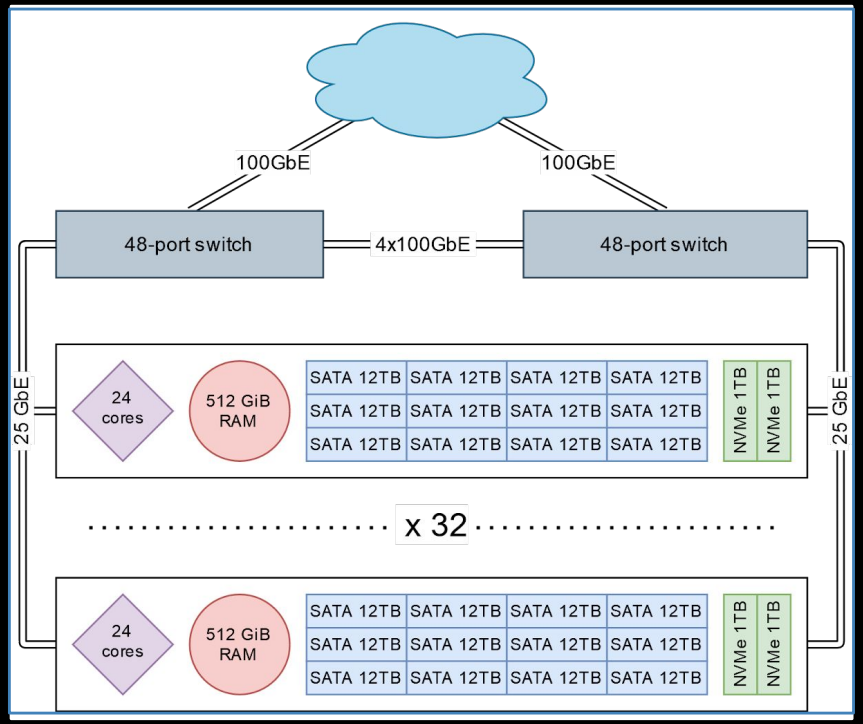
Interacting with and distributing massive, structured datasets through the **fastest platform** in the astronomy and cosmology community.

Offering an **easy and intuitive interface** to advanced features through custom-defined functions, such as HEALPiX, array aggregation, and spherical geometry operations.

Collaborators

Port d'Informació Científica (PIC)

CosmoHub big data platform



R&D on cosmological simulations and analysis platforms for massive cosmology data

CosmoHub

Interactive exploration and distribution of massive structured datasets

- **Very fast** and unlimited visualization and custom subsets
- User friendly access
  - **Guided interface**
  - No SQL knowledge required
- Supporting embargoed and public datasets
- Expanding into a **multi-messenger science platform** (gamma rays, neutrinos, GWs)

200 active users / 200 TiB data
Success use case in the EOSC Macro-roadmap
Participation in strategic multi-messenger European initiatives such as ACME

SciPIC

Generation of **high fidelity mock galaxy catalogs** from dark matter halo catalogs

- **Very fast** and flexible pipeline
- Flagship galaxy mock
  - Largest and most comprehensive catalog ever produced
  - Fundamental for **ESA Euclid mission**
- Supporting other projects
  - DES, KIDS, ARRAKIHS, PAUS, …
- **Leadership in mock catalog development**
- Strategic capability to contribute to future surveys (LSST, Roman, DESI-II, etc.)

Euclid Flagship mock catalog v2 public release

PIC Big data common service: 768 CPU cores, 15 TIB RAM, 3PIB net storage, 60 TB NVMe cache

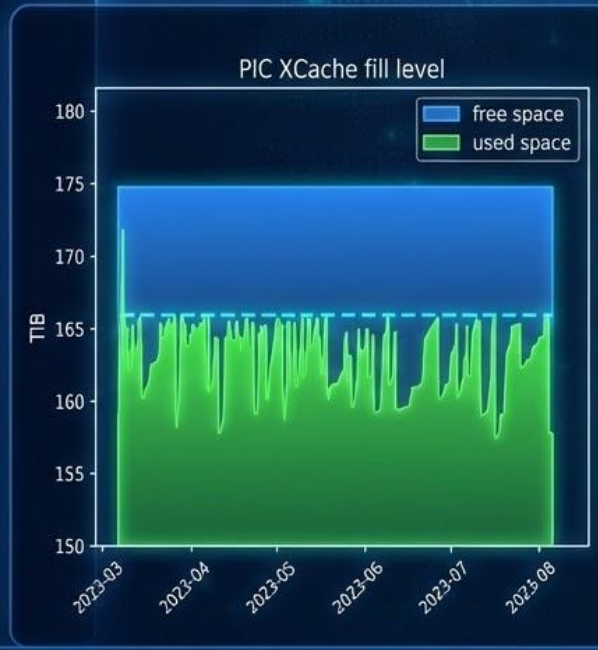
R&D in data caches

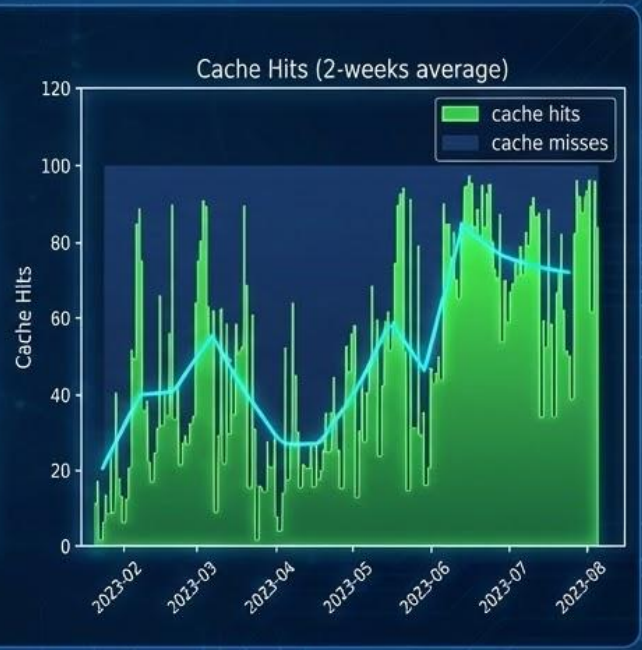
XRootD Data Cache at PIC

175 TB

Used for remote data reads

Jobs running at PIC and CIEMAT





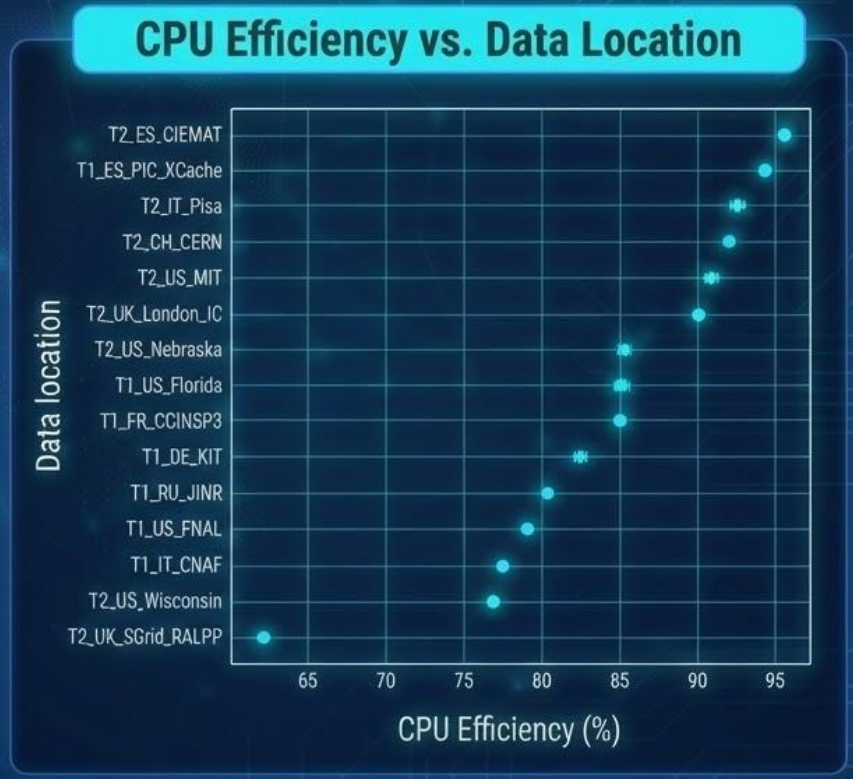
R&D in data caches

Improves job CPU efficiency and reduces WAN traffic

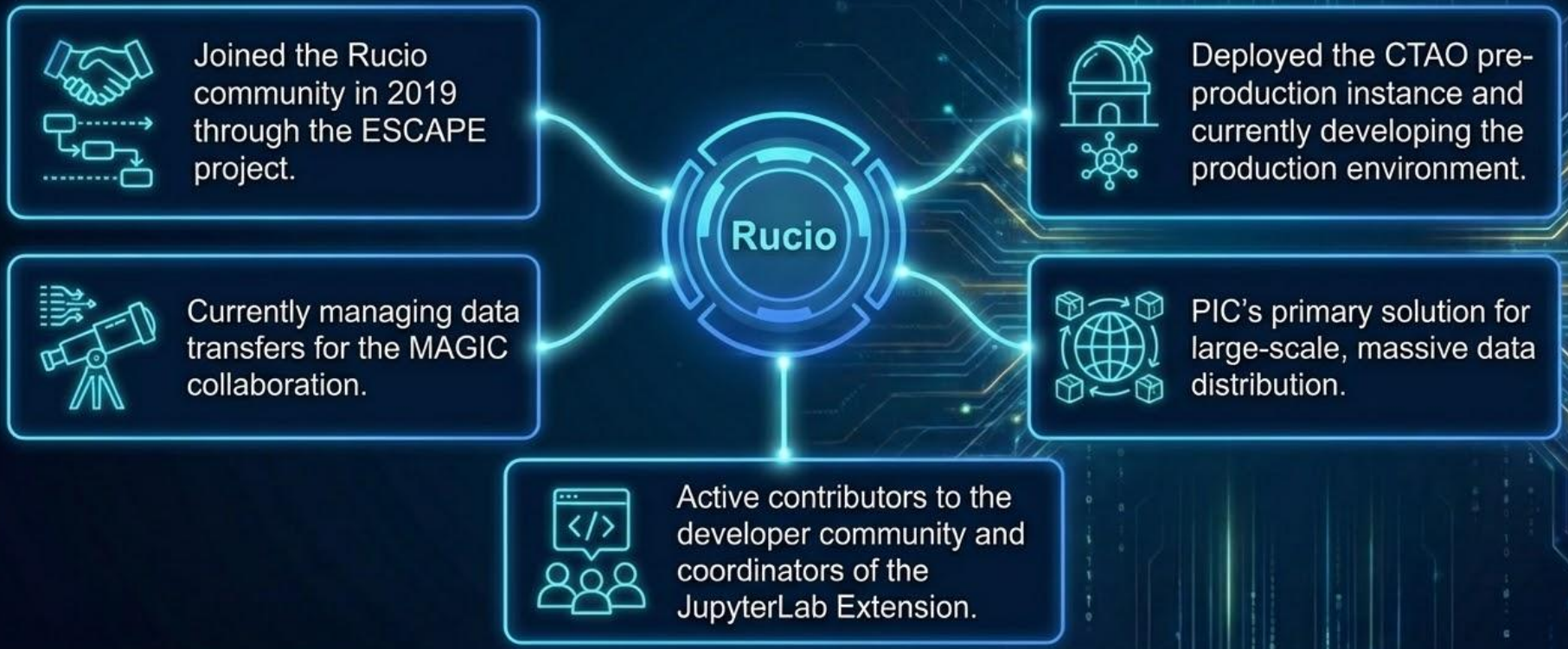
PhD thesis:

Content Delivery Network solutions for the CMS experiment: the evolution towards HL-LHC

Carlos Pérez Dengra | UAB, Feb. 2024 | Supervisor: Pepe Flix



Managing data transfers: Rucio



Applied AI activities

Applications of ML/DL, evolutionary algorithms, explainable AI

- **Astronomy**: redshift estimation, AGN classification, and image denoising using DL models
- **Gravitational waves**: LIGO event discriminator
- **Particle physics**: CMS muon trigger
- **Dark Matter searches**: Signal/background discrimination in Deap3600, position reconstruction in DarkSide-2k

Quantum computing + AI

Within the Quantum Spain project: AI-based optimization of control pulses for quantum systems, with simulations performed on PIC's infrastructure (ongoing Ph.D. thesis)

BSC AI factory

The adoption of Artificial Intelligence within the corporate sector

Exploration of more complex AI models

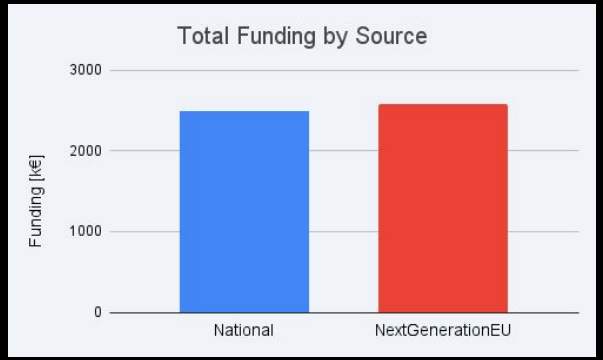
Foundation Models learning semantic representations of data

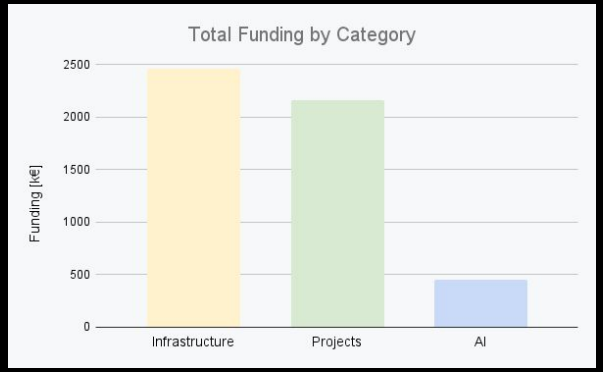
Knowledge and technology transfer to CIEMAT

Oncology, environment

Funding

Call	Funds	Project	Period	Funding [k€]
National scientific-technological equipment	National	PIC high performance storage system	2022-2023	923
Unique Scientific and Technical Infrastructure equipment	NextGenerationEU	PIC data preservation system	2023-2026	540
Unique Scientific and Technical Infrastructure equipment	NextGenerationEU	PIC data processing system	2024-2026	1000
National Program R&D&I	National	CMS Analysis Facility	2022-2025	372
National Program R&D&I	National	LHC Computing Tier-1/Tier-2	2023-2027	634
Ministry Direct assignment	NextGenerationEU	LHC Computing upgrade	2023-2026	593
Ministry Direct assignment	National	PIC CTA data center	2024-2026	460
National Program R&D&I	National	Simulation and data analysis for cosmology	2025-2028	105
Ministry Direct assignment	NextGenerationEU	AI infrastructure for quantum computing	2022-2025	200
Ministry Direct assignment	NextGenerationEU	BSC AI Factory	2025-2027	250
			Total:	**5077**
HORIZON-JU-EUROHPC-2026	Horizon Europe	Scaling HEP Applications and Knowledge at Exascale HPC		160
National Program R&D&I	National	LHC Computing Tier-1/Tier-2		2600





Scientific production

In addition to providing computational support to the scientific projects of the department, the computing group produces as well scientific results, mostly innovation in data-intensive computing. Last three years:

Responsibilities in WLCG

We maintain a significant and highly visible presence in the WLCG organization

Gonzalo Merino is the WLCG LHCC lead reviewer since 2024

JMHC is the Chair of the WLCG CRSG since 2025

Pepe Flix has been the chair of the WLCG Deployment Board (2020-2024)
He is now member of the WLCG Technical Coordination Board

Pepe Flix is Chair of the worldwide HEPiX forum since 2024

Pepe Flix is a Member of the WLCG Management and Overview Boards since 2013

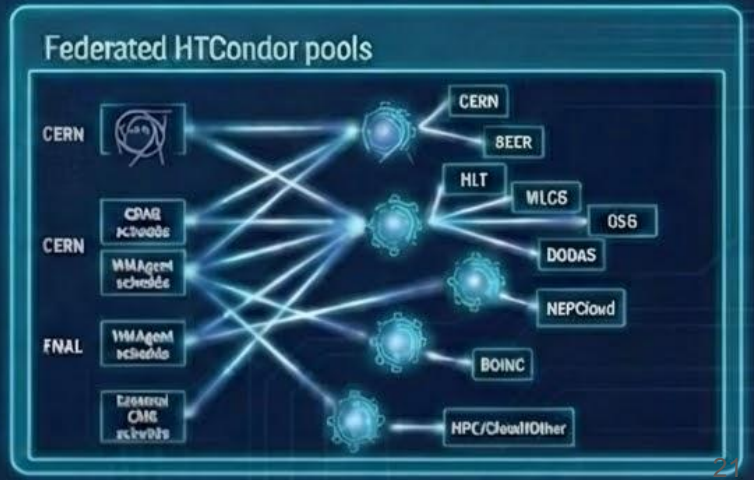
CMS Submission Infrastructure

Provisioning

CMS L2 area co-led by Antonio Pérez-Calero since 2016. Computing resource provisioning system for CMS workflows, aggregating ~500k CPU cores distributed worldwide from Grid, HPC, and cloud providers.

R&D Initiatives

Pilot-based dynamic multi-core scheduling. Integration of heterogeneous resources (HPC, GPU, Cloud, HLT farm).



CMS resource management

Pepe Flix: Resource Management Office L2

Co-leads area since 2013.
Biannual resource utilization and resource requests reports to the WLCG CRSG.
Modelling for computing resource needs projections.

JMHC: CMS Computing Resources Board

Co-chair (2020-2025).
Act as intermediary between resource providers and CMS computing.
Make sure resource needs are matched by institutional pledges.

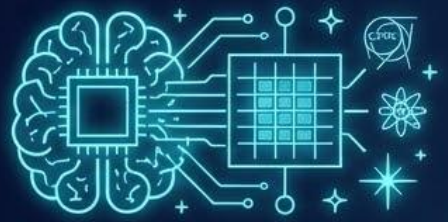
CMS 2024-2025 Usage

CMS		Pledged	Used	Used/Pledged	Avg. CPU EFF
CPU [kHS23]	Tier-0	980 (720)	1,622 (1,172)	166% (162%)	78% (72%)
	Tier-1	1,020 (916)	1,462 (1,236)	143% (135%)	78% (80%)
	Tier-2	1,526 (1,313)	2,200 (2,388)	144% (182%)	68% (72%)
	Total	**3,526 (2,949)**	**5,284 (4,796)**	**150% (162%)**	**74% (73%)**
Disk [PB]	Tier-0	54.0 (45)	47.1 (39.6)	84% (81%)	
	Tier-1	115.7 (96.6)	91.1 (78.3)	79% (81%)	
	Tier-2	136.7 (109.7)	110.6 (98.0)	81% (99%)	
	Total	**366.4 (215.5)**	**248.8 (215.5)**	**81% (86%)**	
Tape [PB]	Tier-0	320.0 (228.0)	300.7 (212.1)	94% (93%)	
	Tier-1	353.9 (303.7)	279.9 (214.3)	79% (71%)	
	Total	**673.9 (531.7)**	**580.6 (426.4)**	**86% (80%)**	

CMS Projections

		2026 Approved	2027 Preliminary	Increase w.r.t. 202.
CPU [kHS23]	Tier-0	1,350	1,350	0 (0%)
	Tier-1	1,200	1,200	0 (0%)
	Tier-2	2,000	2,000	0 (0%)
	Total	**4,550**	**4,550**	**0 (0%)**
Disk [PB]	Tier-0	81	81	0 (0%)
	Tier-1	164	180	16 (10%)
	Tier-2	198	217	19 (10%)
	Total	**443**	**478**	**35 (7%)**
Tape [PB]	Tier-0	515	530	15 (3%)
	Tier-1	540	610	70 (13%)
	Total	**1055**	**1140**	**85 (8%)**

Some thoughts about the future of the research line



Keep doing Scientific Computing / Computational Science : use advanced computing capabilities to solve complex science problems

- Keep providing the necessary computing infrastructure, services and support to the scientific projects in the department
- Our methodology provides unique value to the research community, fostering agile and close collaborations with research groups, and engaging in long-term projects

Leverage data-intensive distributed computing technologies from LHC to other projects that require handling big complex data

Boost the adoption/application of Artificial Intelligence technologies

- Establish inter-disciplinary collaborations to boost AI capabilities of research groups
- Foster synergies between different domains
- Focus on the technical part of developing and deploying AI workflows: in production, at scale, tailored to the science goals

PIC strategic plan 2025-2028

Consolidation & Funding

Consolidate activities, increase funding

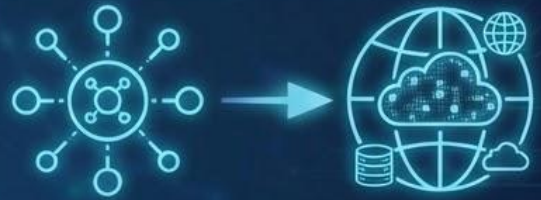
Asking support from Ministry and Catalan government

ICTS & RES Membership

Member of Spanish ICTS (Scientific and Technological Singular Facility)

Data node of Spanish Supercomputing Network (RES) from september 2020

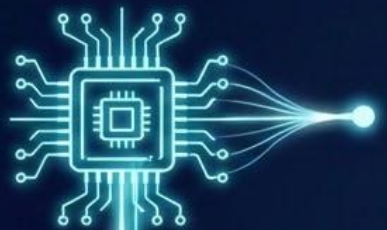
PIC



Transformation & WLCG

We are proposing to transform the RES in a distributed infrastructure using WLCG technologies

Some thoughts about the future of the research line



HL-LHC computing challenge

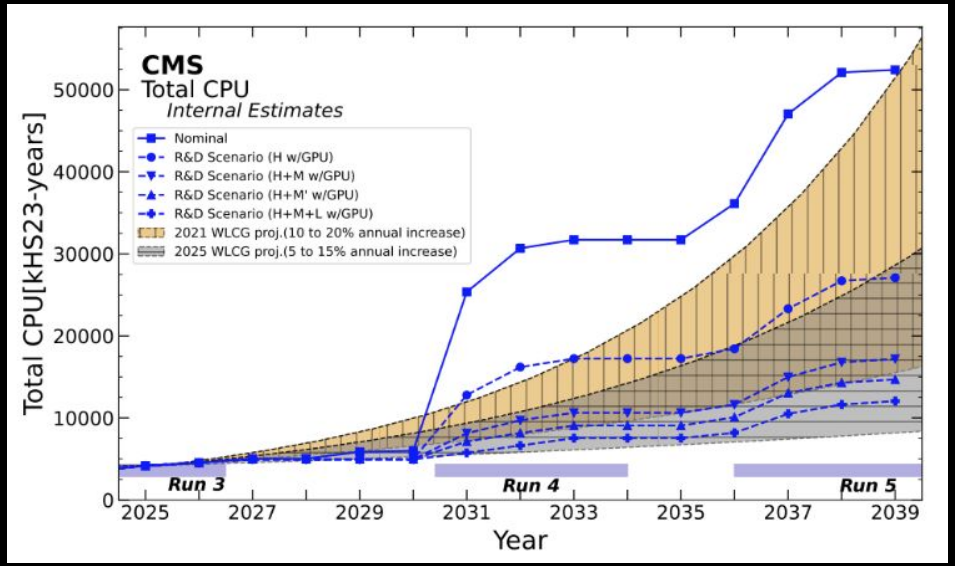
Large increase in data volume and complexity

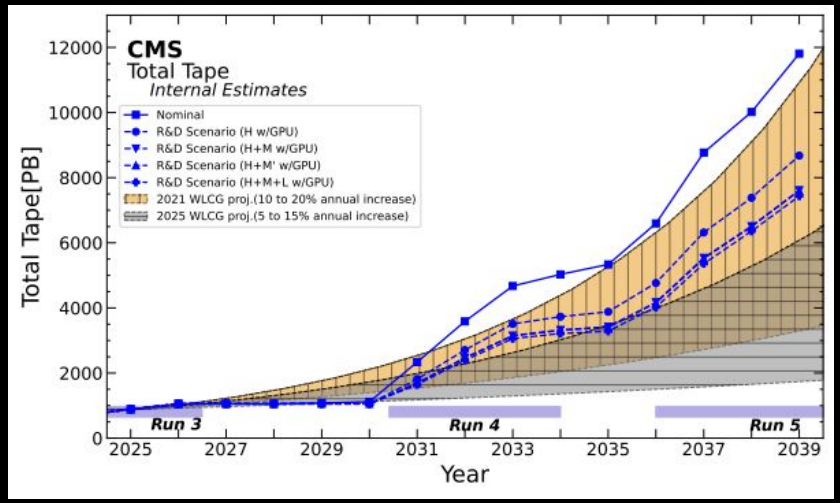
Funding challenge: at most, flat budget expected

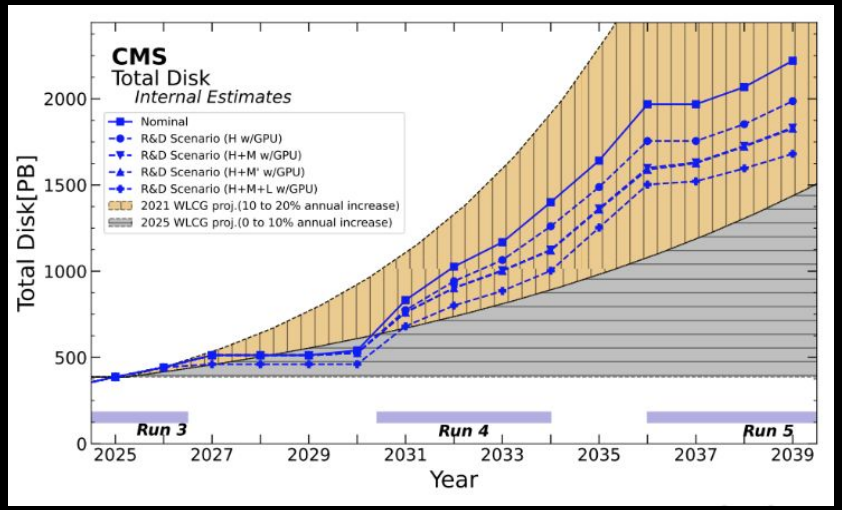
Significant R&D required to reduce computing resource needs

HL-LHC resource projections

Significant R&D needed to keep resource needs affordable with flat budget







CMS Software and Computing CDR

Conceptual Design Report in preparation to face HL-LHC computing challenges

- Computing model and software R&D to process, simulate and analyze data with available resources
  - Very active participation from CIEMAT (chapter editors)
- Detailed **R&D roadmap**, with several probability scenarios
  - Heterogeneous resources (**GPU**, HPC)
  - Software refactoring, new algorithms, GPU acceleration
  - **AI/ML** in all areas (MC gen, sim, reco, analysis)
- Computing resource estimates and projections
  - Event size increase ~4x (150-200 pileup, more granular detector)
  - Higher HLT output rate (10-15 kHz)
  - ~flat budget expected

CMS PAPER TDR-24-001

DRAFT
CMS Paper

The content of this note is intended for CMS internal use and distribution only

December 2, 2025
Archive Hash: none
Archive Date: none

CMS Offline Software and Computing for HL-LHC

Conceptual Design Report

The CMS Collaboration

Abstract

Some thoughts about the future of the research line

CTAO

- Expected completion of the 4 LST CTAO-North in 2026
- Planned completion of the full array configurations for both the North (approx. 2028) and South (approx. 2029–2030)

DUNE

- Expected start of Phase I atmospheric and solar neutrino observations as the first far detector module is commissioned (2028–2029)
- Start of neutrino beam operations following the completion of the second far detector and the beamline (2031–2032)

Euclid

- Data taking until at least 2030

Vera Rubin

- Start of official 10-year Legacy Survey of Space and Time (LSST) (early 2026)

Adoption/Application of Artificial Intelligence technologies to fundamental physics

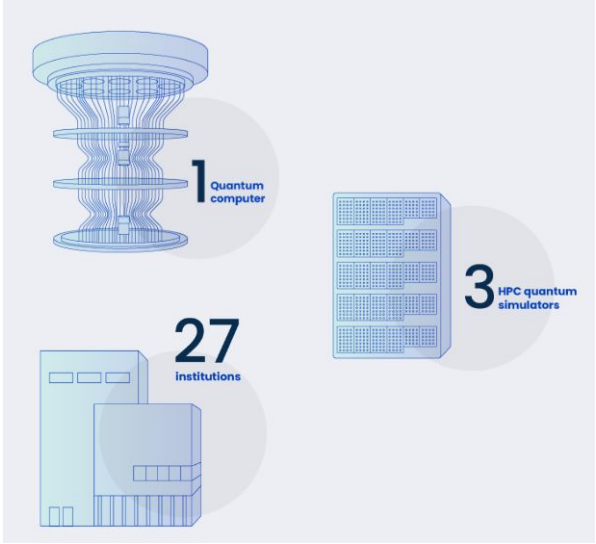
- **Unprecedented Challenges**: The next generation of fundamental physics experiments (colliders, astroparticle, cosmology) face extreme data volumes, complex detectors, stringent real-time constraints, and long operational lifetime
- **The Need for AI**: Addressing these requires a paradigm shift: **embedding Artificial Intelligence across the entire experimental lifecycle**, from detector R&D to operations and physics exploitation
  - When coupled with advanced **Big Data infrastructure** these techniques can dramatically enhance the scientific output of fundamental physics programs
- Global momentum towards AI in experimental Fundamental Physics is reflected in major European and worldwide initiatives
  - ESPP 2026, European Coalition for AI in Fundamental Physics (EUCaif)
  - Strategic White Paper on AI Infrastructure for Particle, Nuclear, and Astroparticle Physics: Insights from JENA and EuCAIF
  - US genesis mission (AI-enabled science to accelerate scientific advancement; 320 M€ initiallly)
    - White paper AI for HEP: Building an AI-native Research Ecosystem for Experimental Particle Physics
- AI4EFF will be one of the backbones of the upcoming MdM application

Backup slides

Quantum Spain project

The first quantum computing ecosystem in the country

In Quantum Spain, the **Spanish Supercomputing Network** (RES, from Red Española de Supercomputación) and leading **research institutions**, are working together to create the first quantum computing infrastructure in Spain **to explore the applications of this new computational paradigm**.



Where quantum physics and artificial intelligence come together

A whole infrastructure at the service of citizens

Researchers collaborating in Quantum Spain will develop **new quantum algorithms** and explore their **real applications**. A range of solutions and open codes **available to all users**.

A. Alou working on applying AI methods to optimize control pulses for qubit manipulation, with simulations of quantum systems carried out at PIC

AI factory

- Large-scale strategic initiative to establish European global leadership in artificial intelligence development, and deployment
- Boost the uptake and use of AI by industry
- One of seven European AI factories hosted by BSC
- 200 M€ funding
- CIEMAT will get funding for hiring two people