PIC
port d'informació
científica

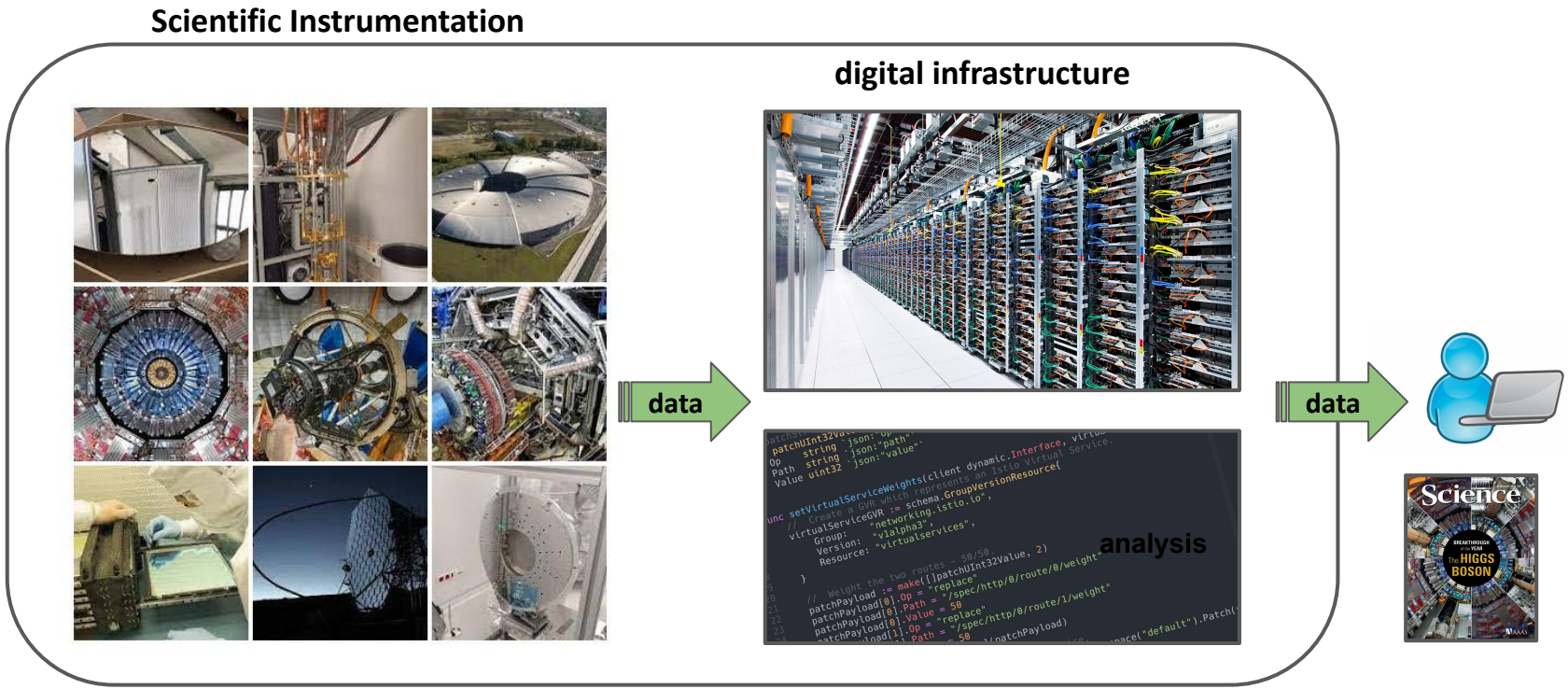
Research Infrastructures and Big Data

3rd ESFRI Open Session on RIs and Big Data

G. Merino, merino@pic.es

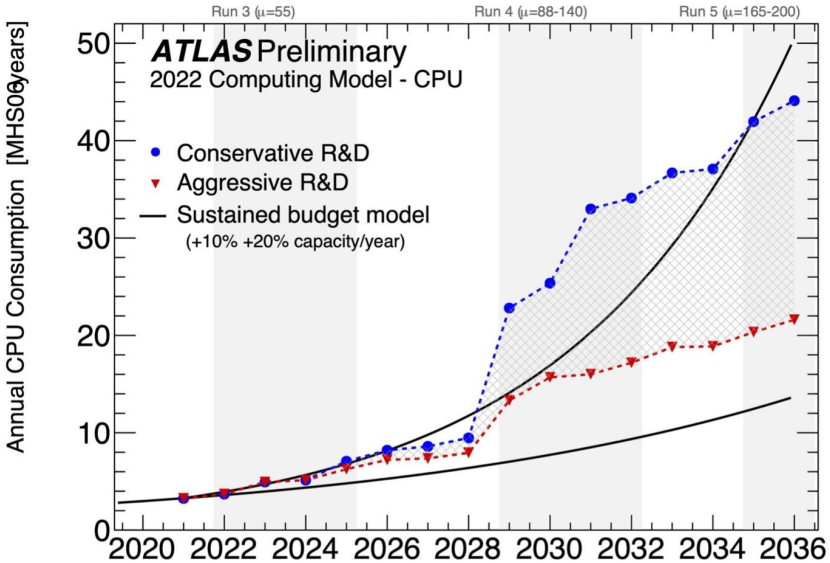
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas
cfp CIEMAT física de partículas
IFAE Institut de Física d'Altes Energies
EXCELENCIA SEVERO OCHOA
BIST Barcelona Institute of Science and Technology

Digital Infrastructure as integral part of Scientific Instrumentation



HL-LHC future computing needs

10x more data, 10x more complexity



Source: CERN-LHCC-2022-005 report

It will be needed to increase personpower doing R&D to meet the needs with a flat budget.

+20% yearly capacity increase is not guaranteed. Sizeable risk.

Operative costs for digital infrastructures increasing - electricity

Computing landscape at a turning point

For decades, Moore's law and Dennard Scaling enabled exponential performance improvements.

This has now changed, and improvements come at the expense of increased complexity:

- Parallelization - multi-core, new architectures such as GPUs …
- Federated infrastructures
  - The future is likely to be **distributed** - use resources wherever they are.
  - The challenge is automation. Orchestration to manage complexity.

High Performance Computing

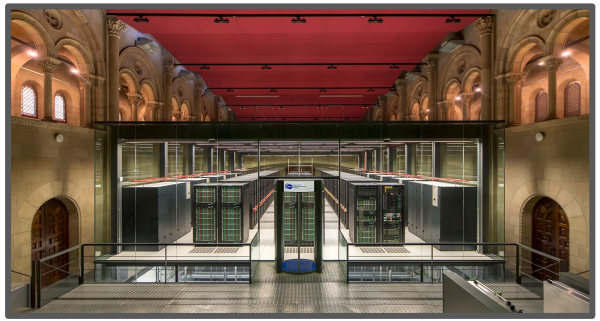
Big investments in strategic agendas of several countries.

Need to find the best way to use HPC for scientific experiments, BUT there are barriers - need stronger collaboration.

- Access policies - quarterly calls for short projects
- WAN not designed for massive data in/out
- No outgoing connectivity from the worker nodes
- There will be need to access HPC in real-time

Heterogeneous HPC landscape. Some will be usable, some won't

- Success in exploiting GPUs will play an important role.
- Also an opportunity for experiments to access opportunistic CPU and mitigate its shortage risk.



Cloud

Cloud advantage is **elasticity**

Need to understand TCO (experiences at large-scale already under way: LSST, ATLAS ...)

Keep control over the data - avoid vendor lock-in

https://www.ocre-project.eu/

ARCHIVER
HELIX NEBULA
ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

https://archiver-project.eu/

Collaboration

Lots of science programs are spanning decades - sustainability of the computing infrastructure is as big a challenge as raw capacity.

- Increased collaboration is one of the keys to **sustainability**.
  - Scientific exchanges and technological cooperation in areas of common interest such as R&D for s/w and computing.
  - Exploiting synergies between different fields

Many of the large scientific computing facilities support multiple experiments.

- More efficient operations through economies of scale
- Reduce operational support load by:
  - Open source, open protocols & federated resources
  - Common tools

ESCAPE

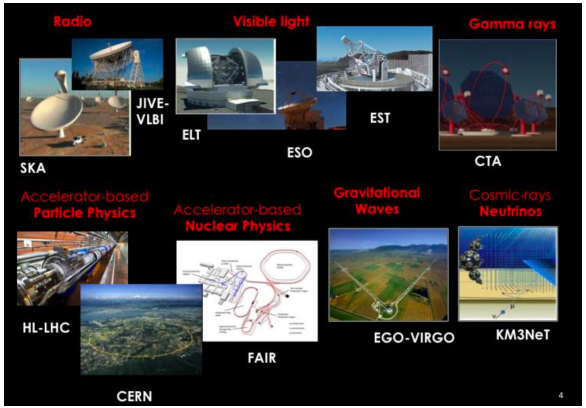
EU project in response to H2020 call to build the EOSC by connecting it to the **ESFRI RIs.**

ESCAPE: Astronomy and Particle Physics ESFRIs (2019 - 2022)

- Goal: Develop common “e-infrastructure” solutions for a wide range of particle physics & astronomy research facilities.
- Exploit synergies between both communities expertise

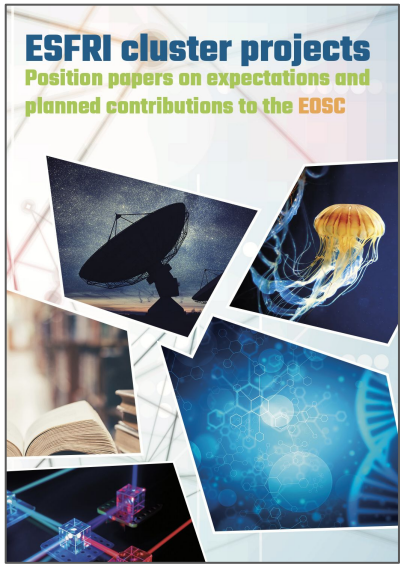
ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

https://projectescape.eu/

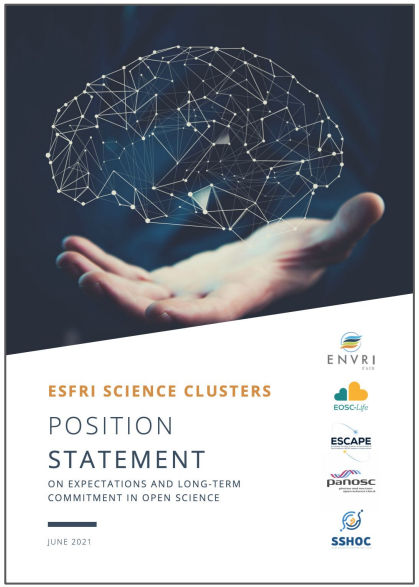


Broader synergies with other research clusters

*Five thematic **Science Clusters** funded under the same H2020 call (80% of ESFRI RIs)*



Dec 2019
doi:10.5281/zenodo.3675081.svg



June 2021
doi:10.5281/zenodo.4892245.svg

Summary

Extraordinarily dynamic environment

- Massive investments from industry - lots of innovation, fierce competition for talent
- Short cycles - need a framework and processes to manage change - iterative improvement
- Engage with industry to develop, integrate and exploit new technologies

Long-term data preservation is challenging for large science projects - multi-decade timescale

- Need organizations with sustainable funding model, with technical expertise on data management and trusted by the community

Computing (software dev. & data analysis) as a 1st class citizen in Research Infrastructures

- Incentive to contribute to computing, as compared to detector construction/operations
- Software development as an intellectual activity - career path for Research Software Engineers

Technology and needs evolve exponentially, but available people's effort does not

- Ability to scale-out is key - Federated distributed heterogeneous infrastructures
- Sustainability through Collaboration

thank you