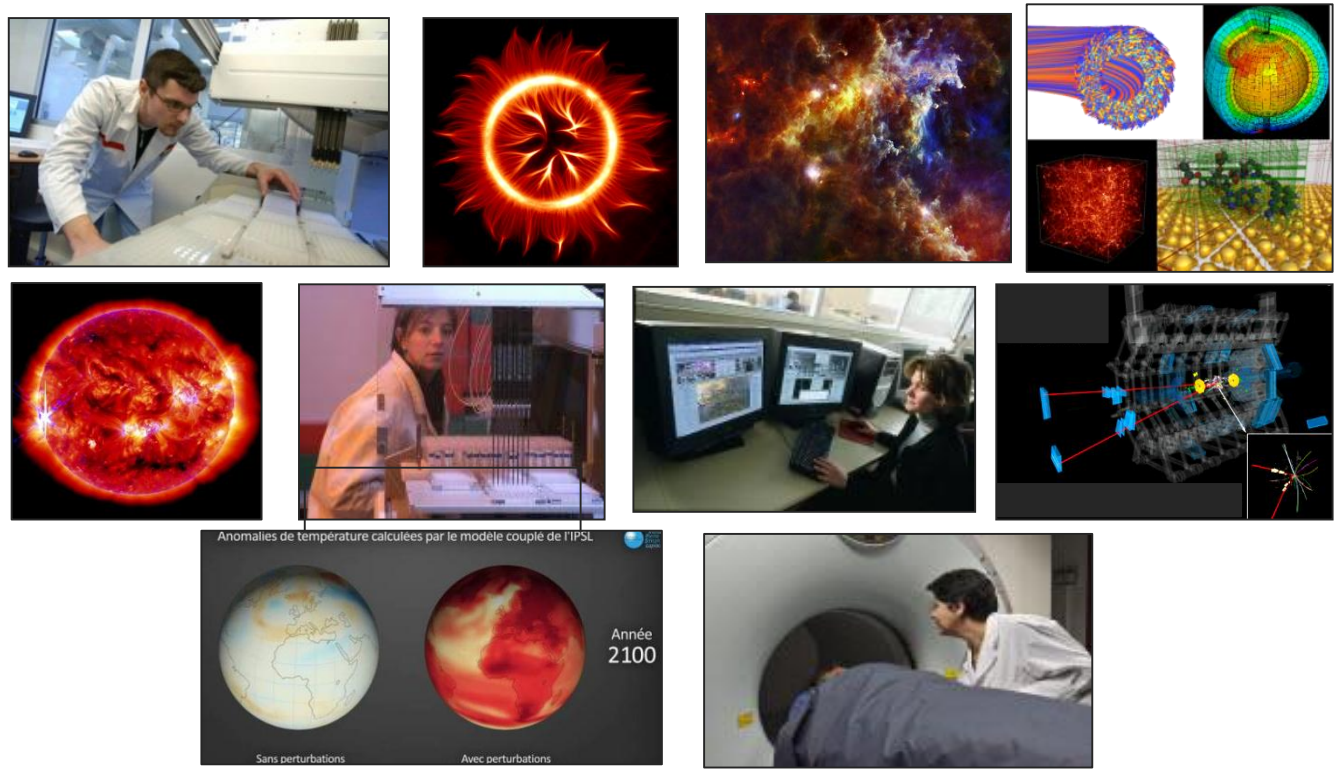
Exascale challenges and opportunities for fundamental research



Fundamental Research @ CEA

Part of Fundamental Research at CEA

- Contributes to CEA strategic framework
- Pushes forward the frontiers of knowledge

Knowledge

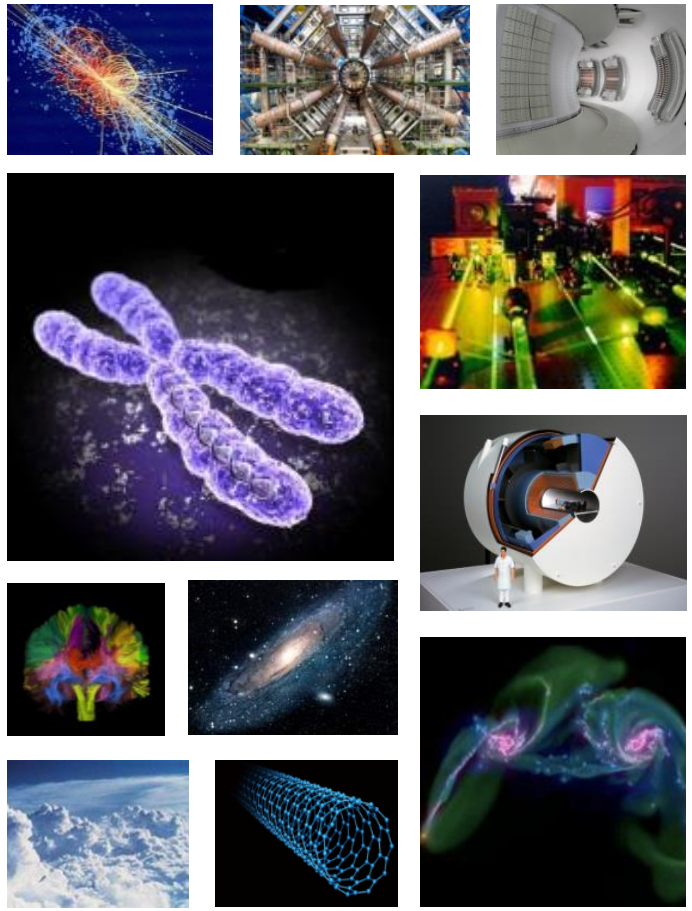
- Physics
- Chemistry
- Material science
- Biology et biotechnologies
- Health
- Climate et environment

Know-how

- Cryomagnetism
- Lasers
- Nanotechnologies
- Sensors
- Imaging
- Numerical simulation

Our missions

- Contribute to the development of knowledge at the best international level in its areas of expertise
- Contribute to the promotion and transfer of knowledge, skills and technologies
- Contribute to training
- Design, implement and manage scientific and technological platforms
- Participate on behalf of the CEA in the conduct, management or governance of major international and national facilities
- Steering, on behalf of the CEA and in line with the policy adopted by the General Administrator, cooperation with national research establishments and site universities
- Conduct an active watch for the various subjects of interest for the CEA



Key figures

5200 people

including :
2500 fulltime employees CEA,
1200 CNRS, INSERM, ….

600 M€ budget

Scientific excellence

- **1148** PhD sudents
  (50 % international)
- **211** Post-doctoral and other contracts
- **223** Collaborative European projets since 2014
- **118** European Research Coucil grants since 2007
- **13** highly cited researchers

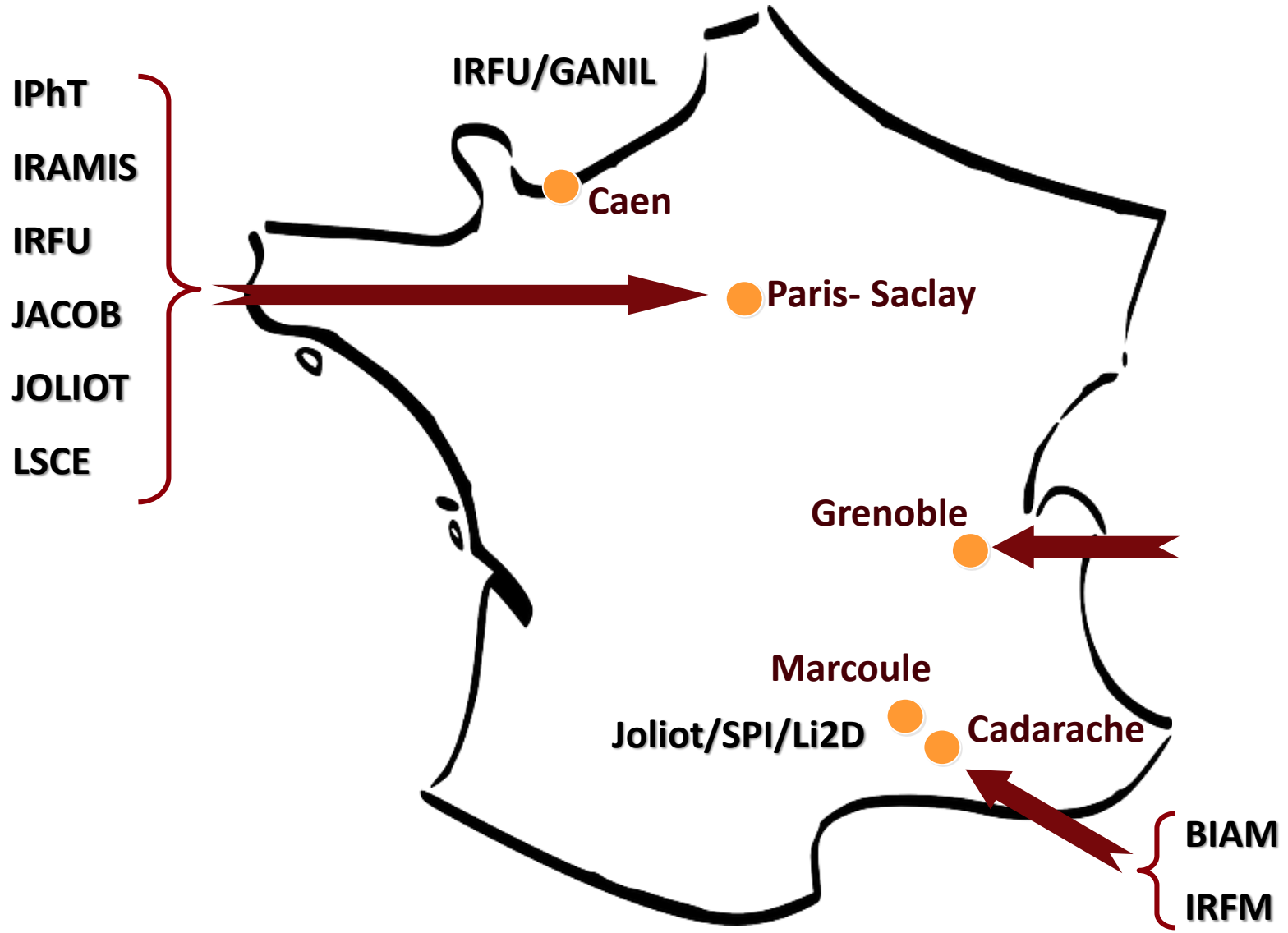
Spirit of openness and cooperation

- **36** mixed research units
- **3 700** scientific papers/yr
  (70 % with international collaborations)

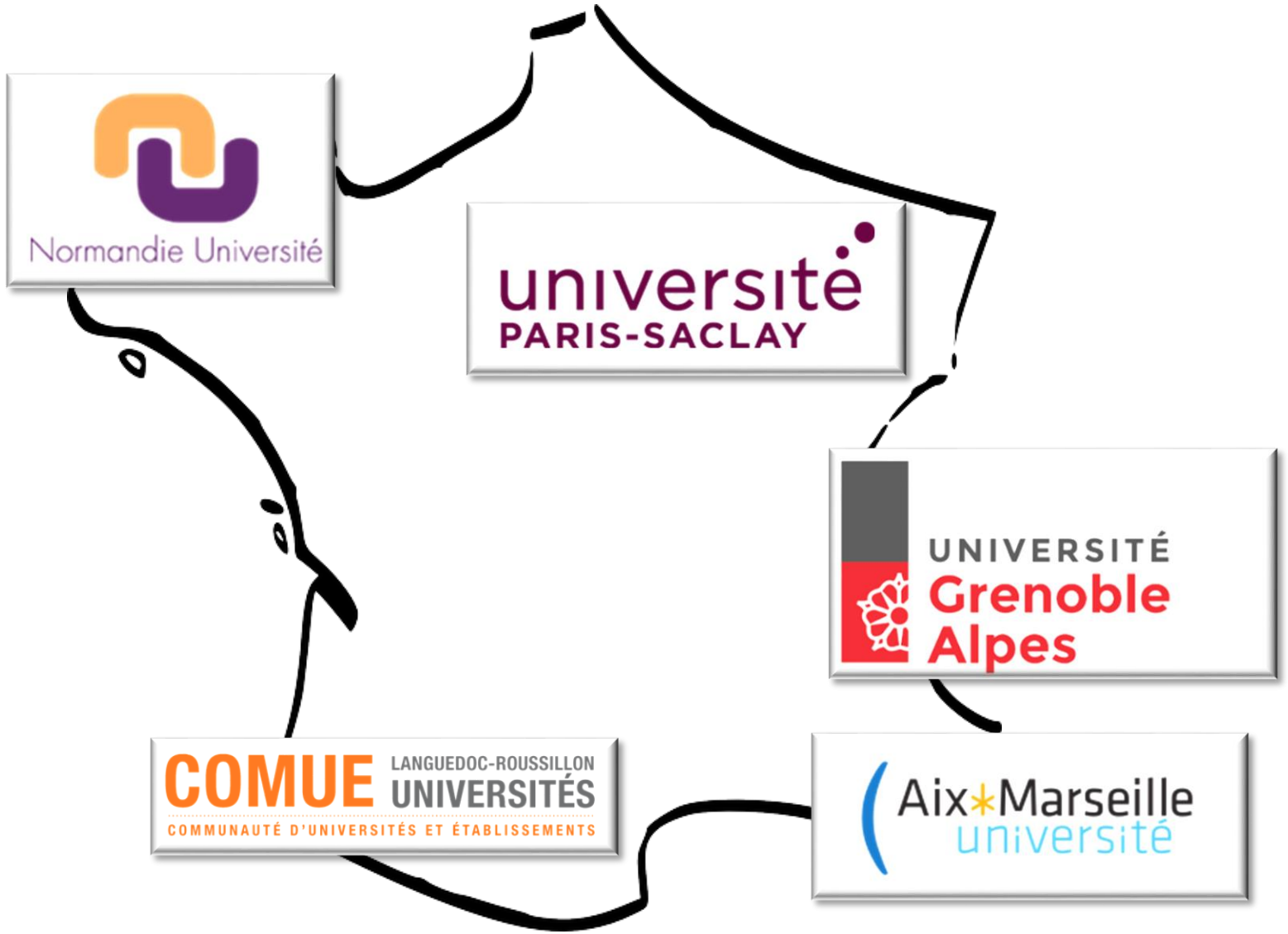
Active policy for IPP & industrial transfert

- **230** industrial contracts
- **566** active patents
- **38** start-up since 2000

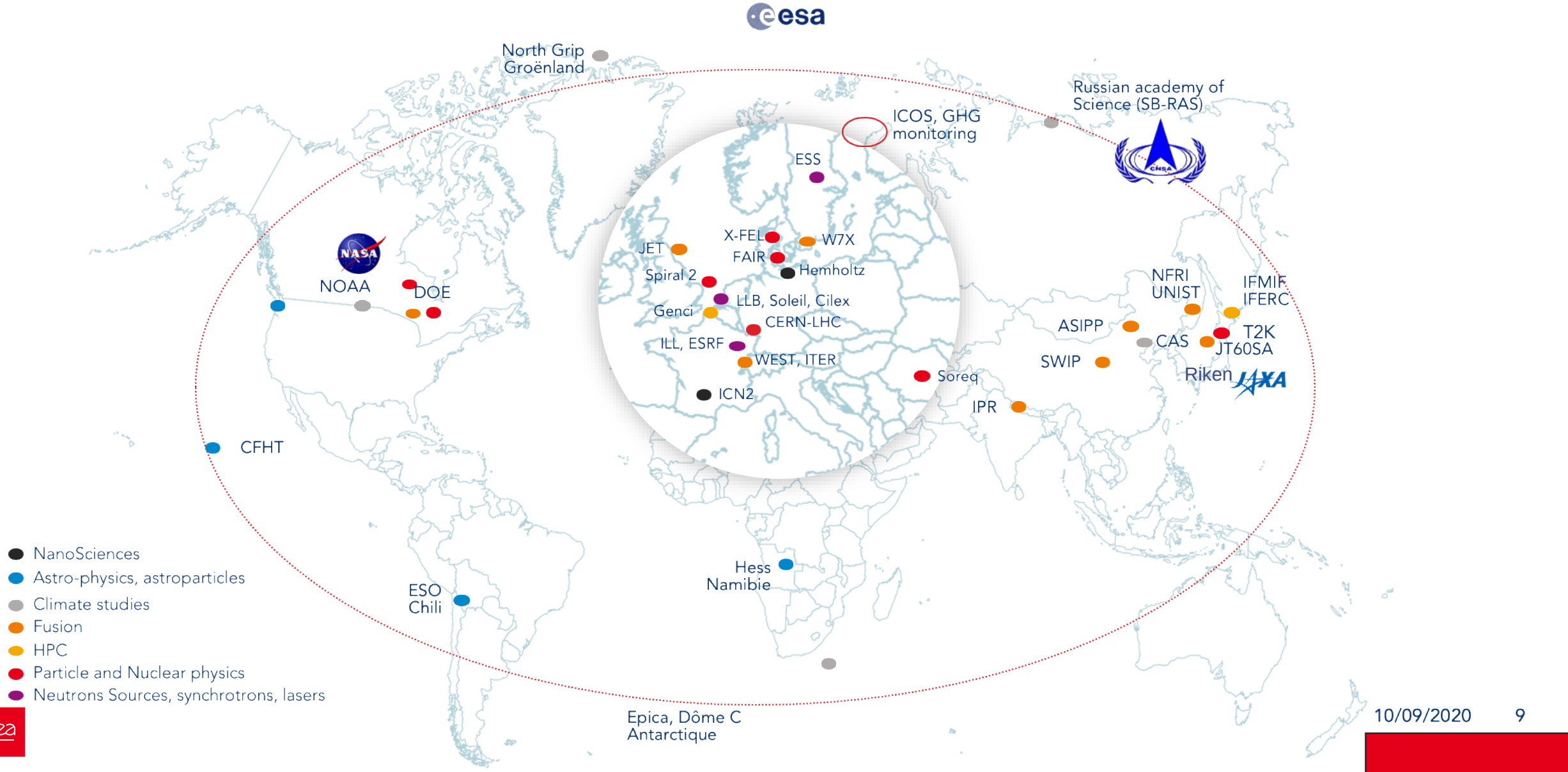
Main facilities



Strong involvment in Universities



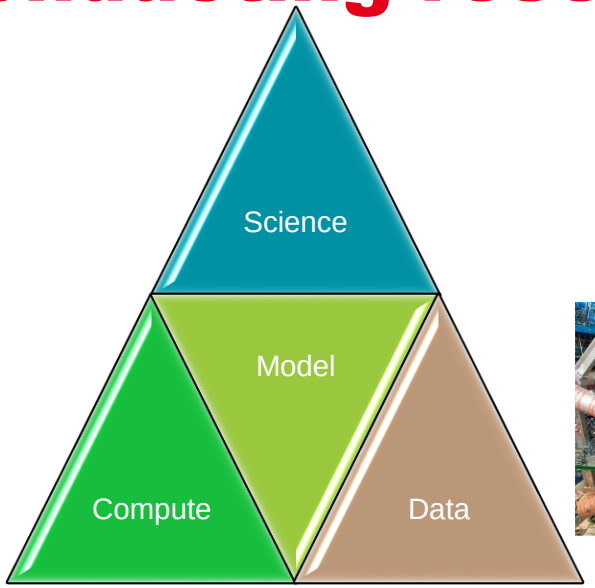
Large experimental facilities and international collaborations

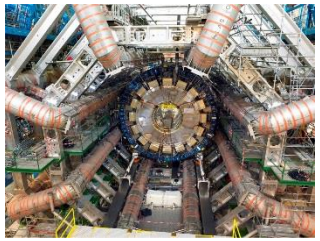


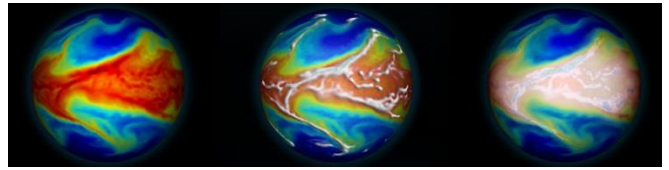
High performance digital for fundamental research

Cornerstones for conducting research and achieve our objectives

- High Performance Digital
  - High Performance Computing
  - High Performance Data Analytics / AI
  - QC-HPC







Data analysis and numerical simulations are now fully nested to achieve our research

But we have huge challenges to face regarding extreme scale

One look is worth a thousand words

GPUs

Arm

Data

Scalability

Our applications

applications

exascale

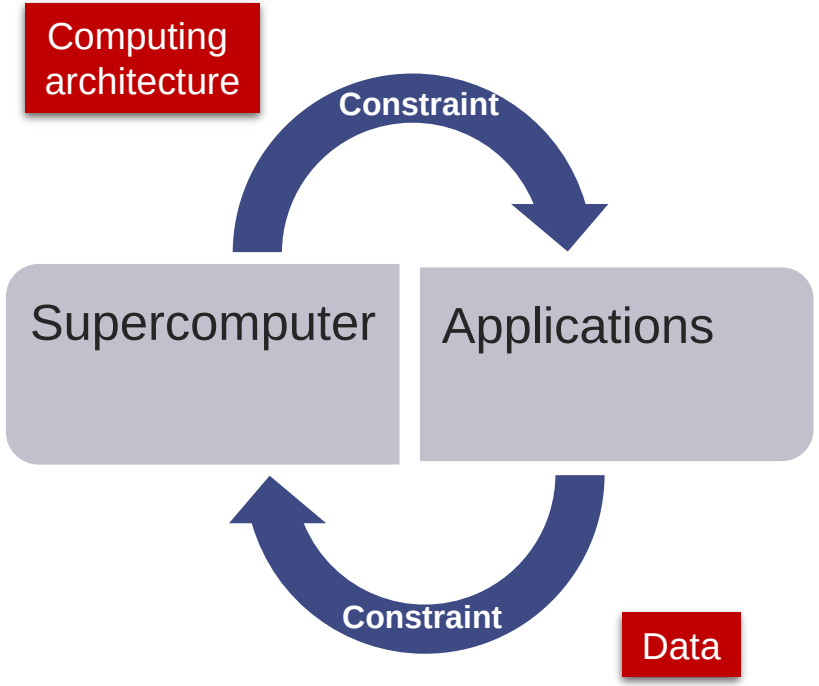
2 two main challenges

Data

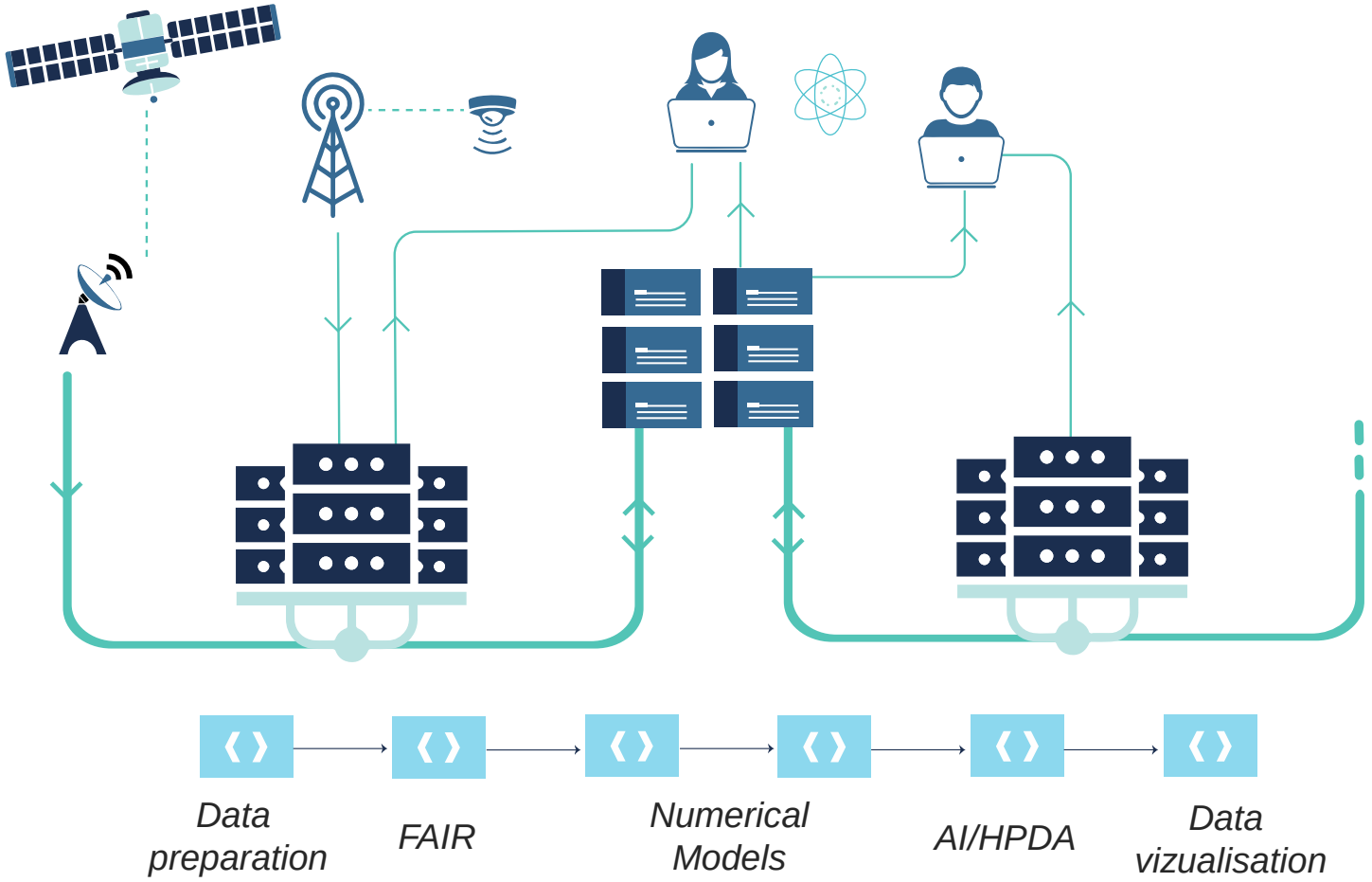
- Managing and exploiting data :
  - Data is coming from many different sources:
    - Observations, Numerical simulations, Instruments and experiments
  - Volume of data is increasing dramatically due to:
    - Needs for finer modeling, Progress in HPC usage, Instruments giving more and more information
  - Interpretation and analysis of these data:
    - Use of HPC → HPDA, Mixing of many different sources of data, Mixing exp./obs data and simulation data
  - Data needs to be open and FAIR

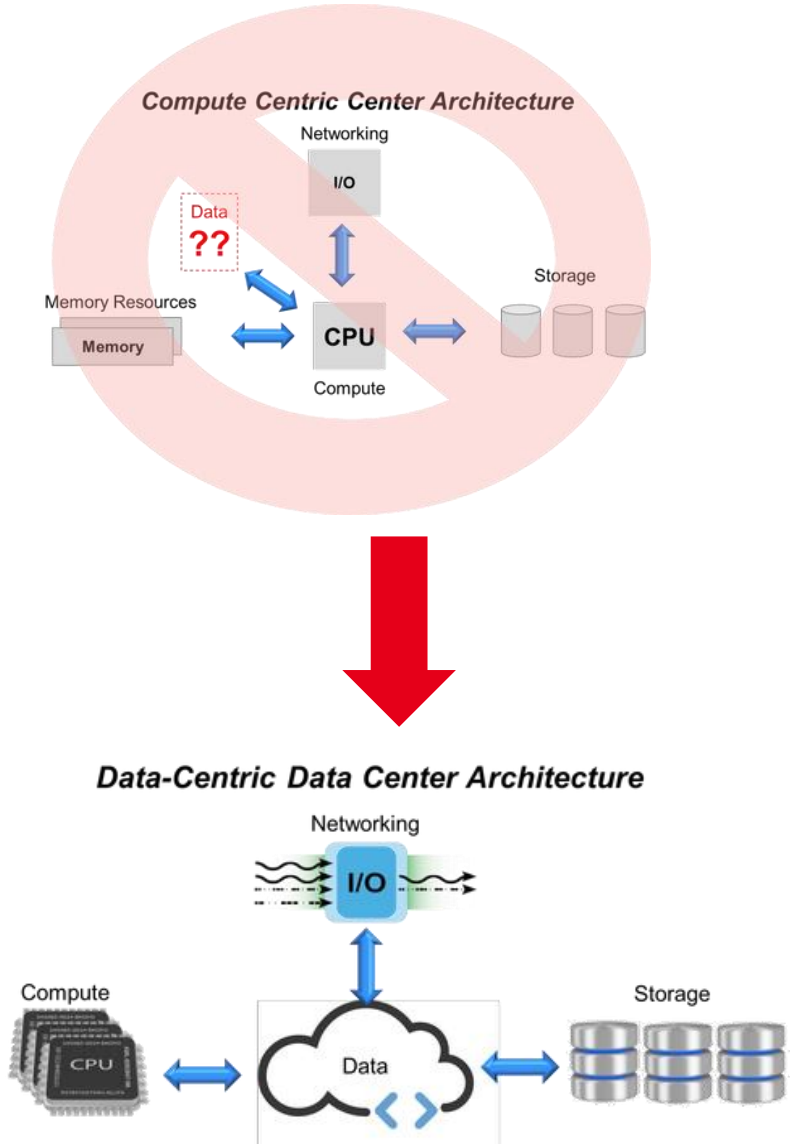
Compute

- Computing architecture are more and more complex
  - Heterogeneity (x86, ARM, GPU)
  - Memory hierarchies and computing units hierarchies
  - Huge number of cores → scalability



Digital Continuum – The Data Challenge

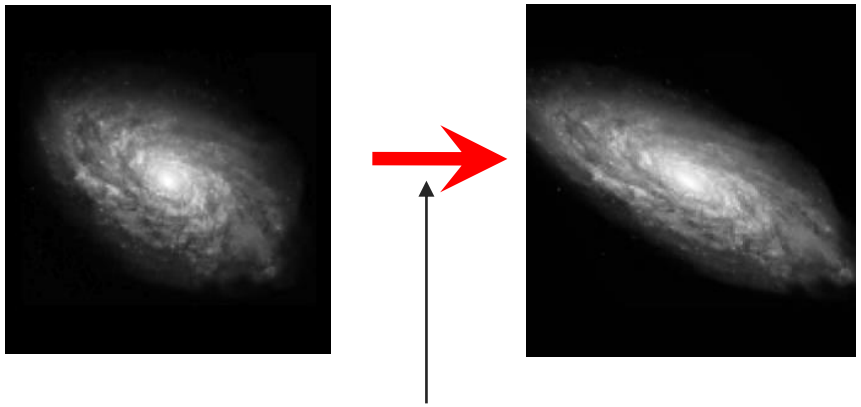




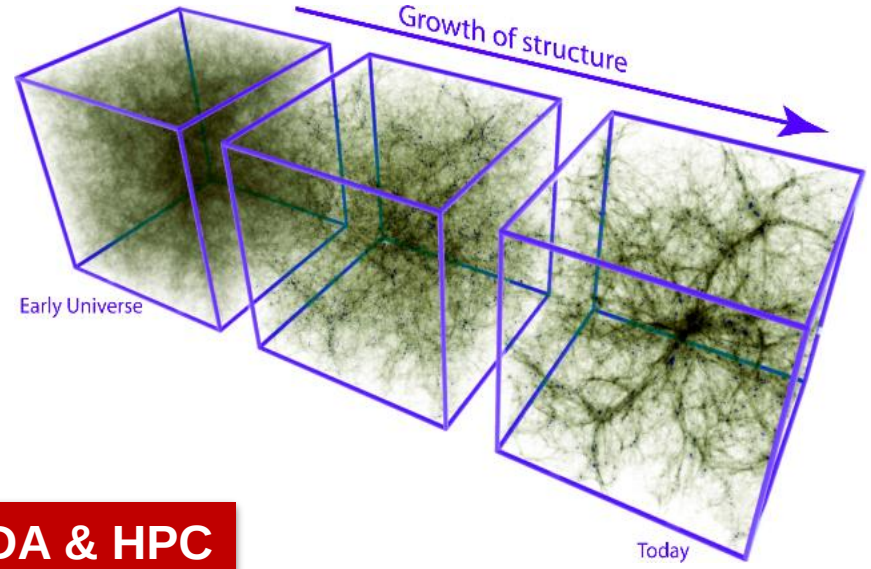
EUCLID: Searches the signatures of the dark energy and matter

The dark matter and energy can be studied by looking at:

- The geometry of the universe
  - Measure of position of galaxies as a function of redshift
- Growth of density perturbations
  - Evolution of structure as a function of cosmic time, growth rate
- Galaxy image distortion caused by dark matter bending light



Dark Matter

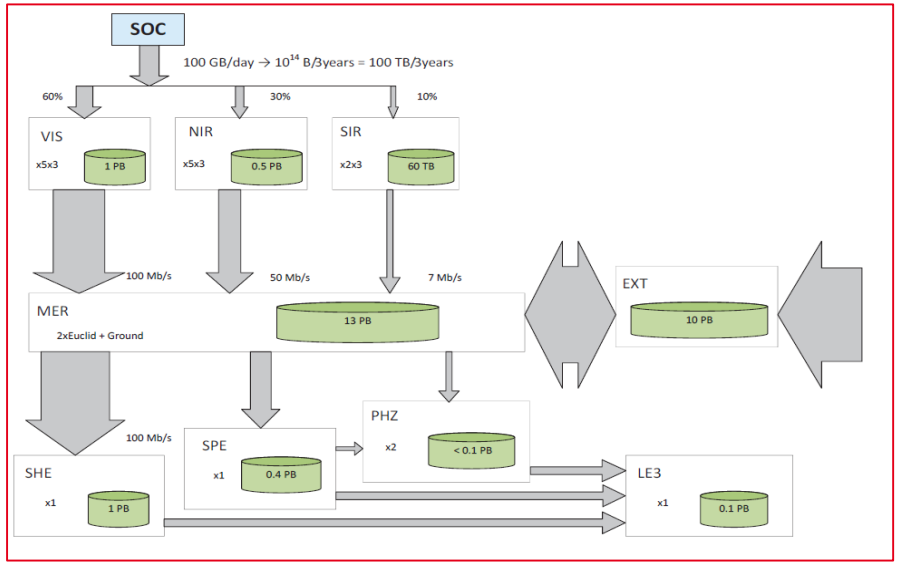


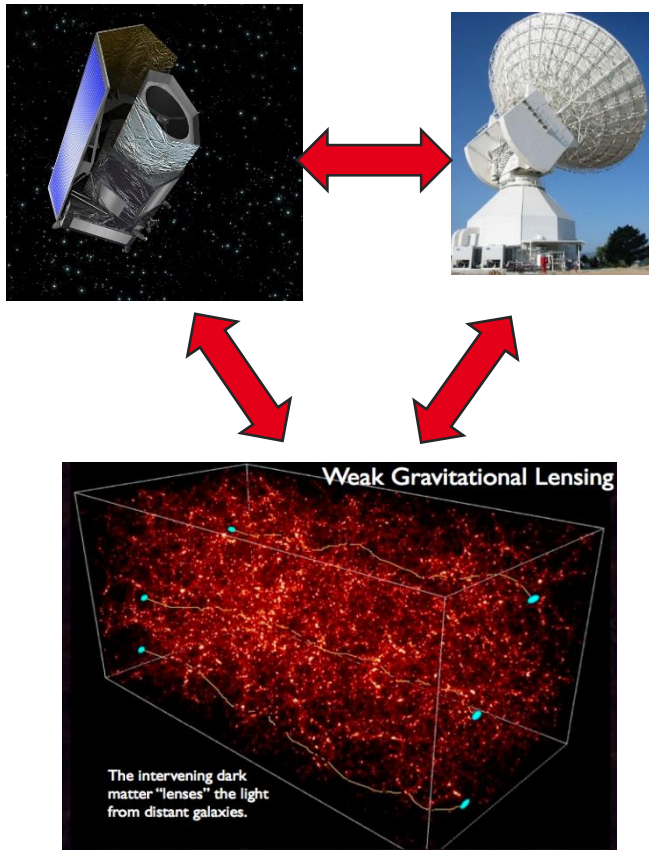
HPDA & HPC

EUCLID: Searches the signatures of the dark energy and matter

How to measure the dark matter:

- EUCLID: measure image distortion
- Cross observations from the ground: measure distances between galaxies
- Correlate these interpreted observational data with numerical simulations

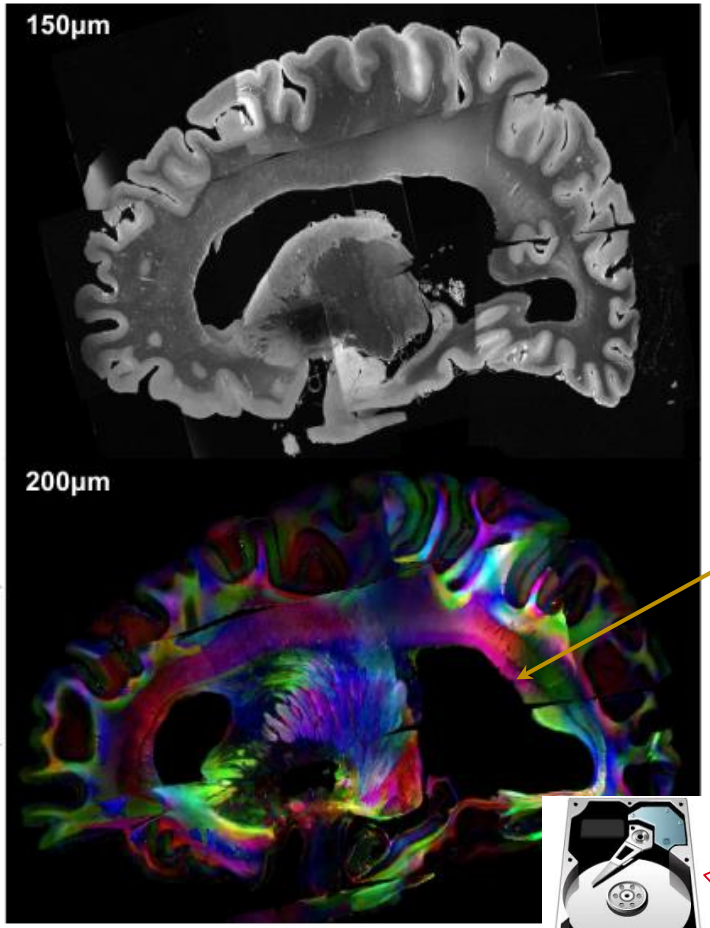


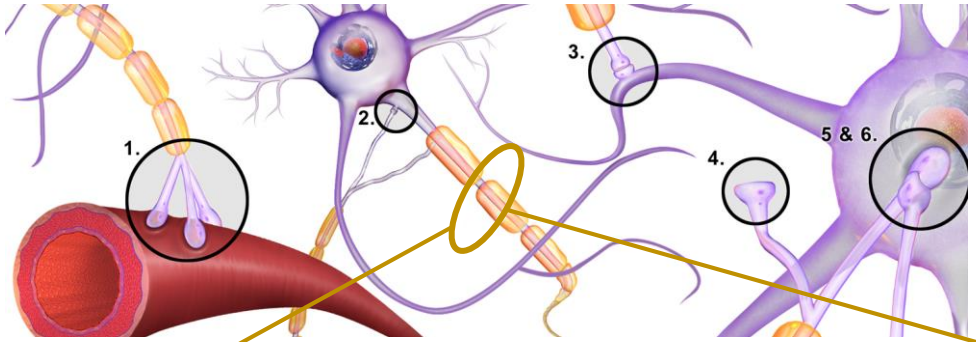


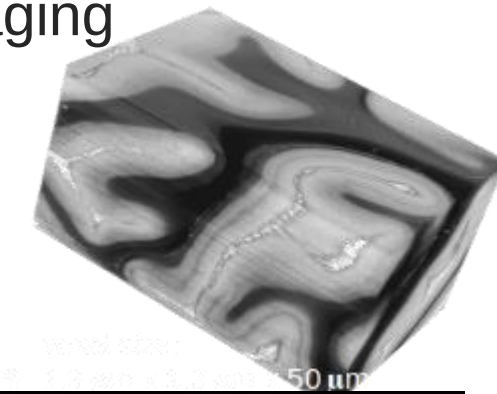
Extreme Scale Analytics and Simulation for Brain Network – scientific summary

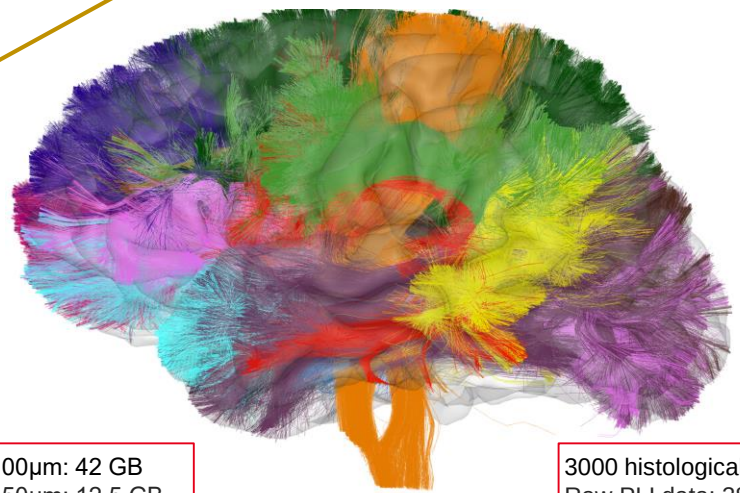
AIDAS
AI Data Analytics and Scalable Simulations

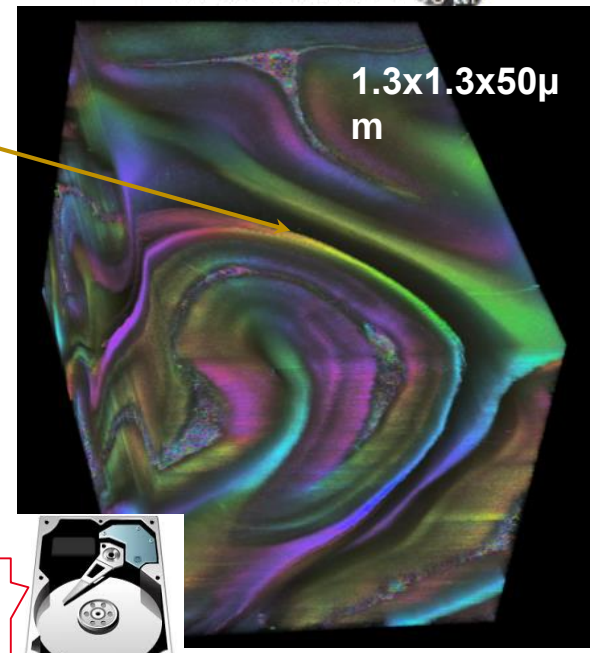
- Scientific goal: deciphering the human connectome from the meso- to the micro- scales using extreme field diffusion MRI at 11.7 Tesla and 3D Polarized Light Imaging











Anatomical MRI 100µm: 42 GB
Anatomical MRI 150µm: 12.5 GB
Anatomical MRI 200µm: 5.2 GB
Diffusion MRI 200µm: 870 GB

HPDA/AI/HPC

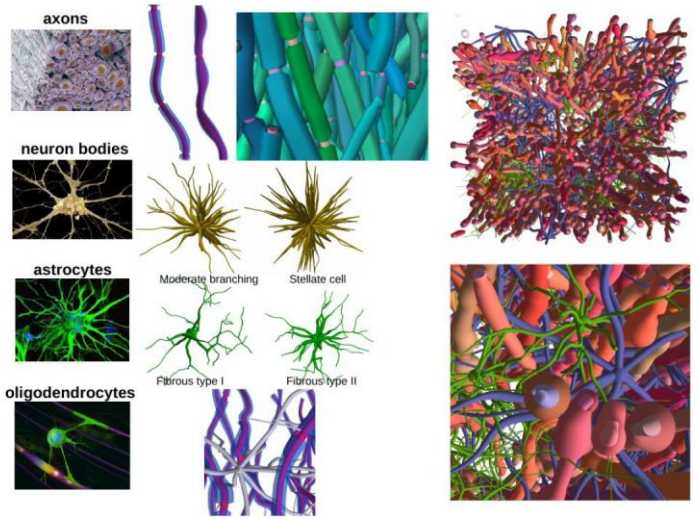
3000 histological sections
Raw PLI data: 3800 TB
Processed PLI
@1.3µmx1.3µmx50µm: 8800 TB

Diffusion MRI: 12000h scan @11.7T

PLI: 2.5 years scan

Extreme Scale Analytics and Simulation for Brain Network – scientific summary

- Big Data Acquisition
- Modeling the microstructure:
  - computational dMRI/PLI
  - computational models of cytoarchitecture from large scale simulations
- New HPC Global Tractography Algorithm from distributed data
- Mapping the human brain connectome model / white matter microstructure atlas at the meso-/micro-scales



The MEDUSA large scale simulation framework to create ultra-realistic virtual brain tissues

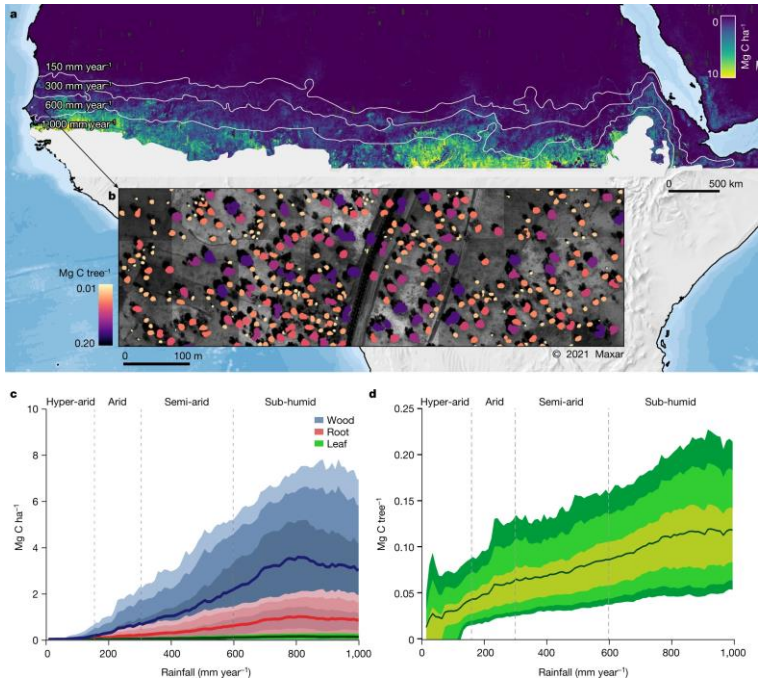
Human Brain Project

EBRAIN

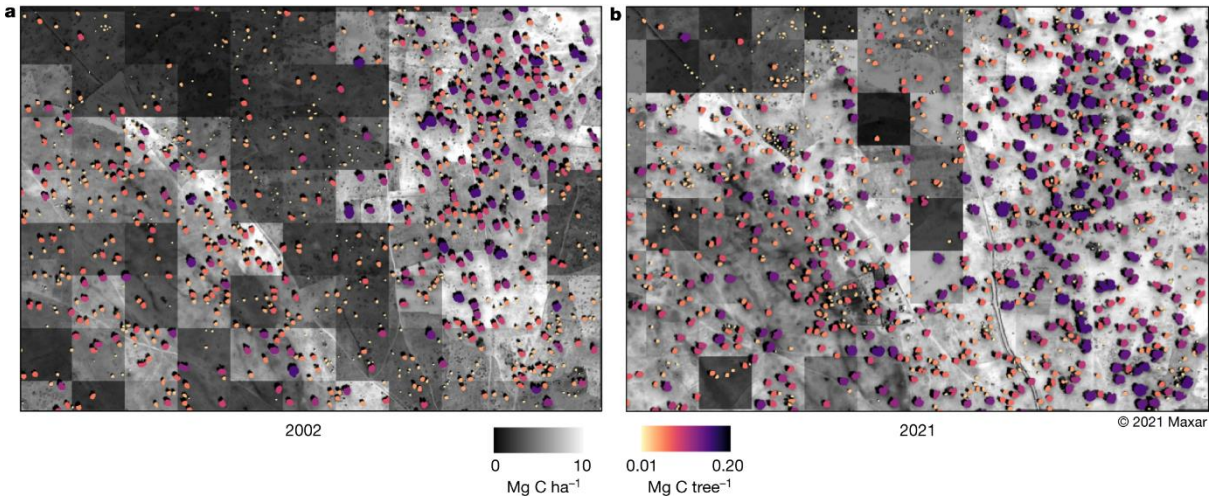
Sub-continental-scale carbon stocks of individual trees in African drylands

*Tucker, C., Brandt, M., Hiernaux, P. et al. Sub-continental-scale carbon stocks of individual trees in African drylands. Nature **615**, 80–86 (2023). https://doi.org/10.1038/s41586-022-05653-6*

The combined use of very-high-resolution satellite images and artificial intelligence made it possible to identify isolated trees and map their crown area at large scales, covering the western Sahara–Sahel–Sudan areas[1]. This approach of mapping individual trees has been extended to a 7.5-times-larger area covering the drylands across Africa, from the Atlantic Ocean to the Red Sea from 9.5° N to 24° N latitude between the 0 and 1,000 mm year^{-1} isohyets, using **326,523 satellite images at a 50-cm spatial resolution, and coupled with machine learning to map 9.9 billion trees**



Wood, foliage and root carbon of 9,947,310,221 trees with crown area >3 m^2 across 9.7 million km^2 were mapped



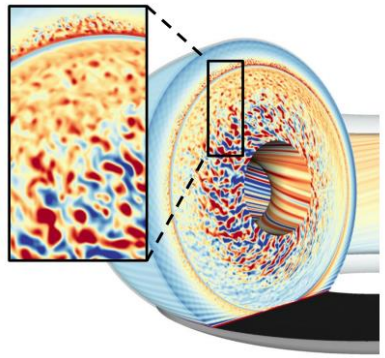
Monitoring at the level of single trees from Khombole, Senegal

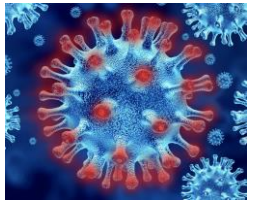
AI & HPC

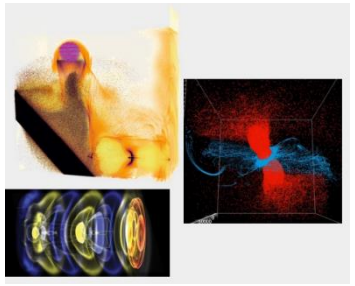
HPC Applications: new generation of codes

HPC

- **Energy - Fusion** : GYSELA-X is a semi-Lagrangian code addressing gyrokinetic full-f global simulations of flux driven tokamak plasmas
- **Materials and biology**: BigDFT is a fast, precise, and flexible DFT code for ab-initio atomistic simulation. With BigDFT, atomistic modeling is more powerful and accessible
- **Laser**:
  - Warpx is an open-source Particle-In-Cell code for the exascale era
  - SMILEI is an open-source multi-physics massively parallel scientific application designed to simulate a broad range of plasma physics scenarios
- **Astrophysics**: Dyablo is a new simulation code developed at CEA in Saclay for the modelling of magneto-hydrodynamics on AMR grids







Big DFT

AIM | Irfu - CEA Saclay Institut de recherche sur les lois fondamentales de l'Univers | cea | erc | WHOLE

Smilei

Language: Python, C++
Distributed parallelism: MPI
Node-level: OpenMP, SIMD
GPU: OpenMP, OpenACC, Thrust
IO: HDF5, OpenPMD
Documentation: Sphinx, Doxygen

acm

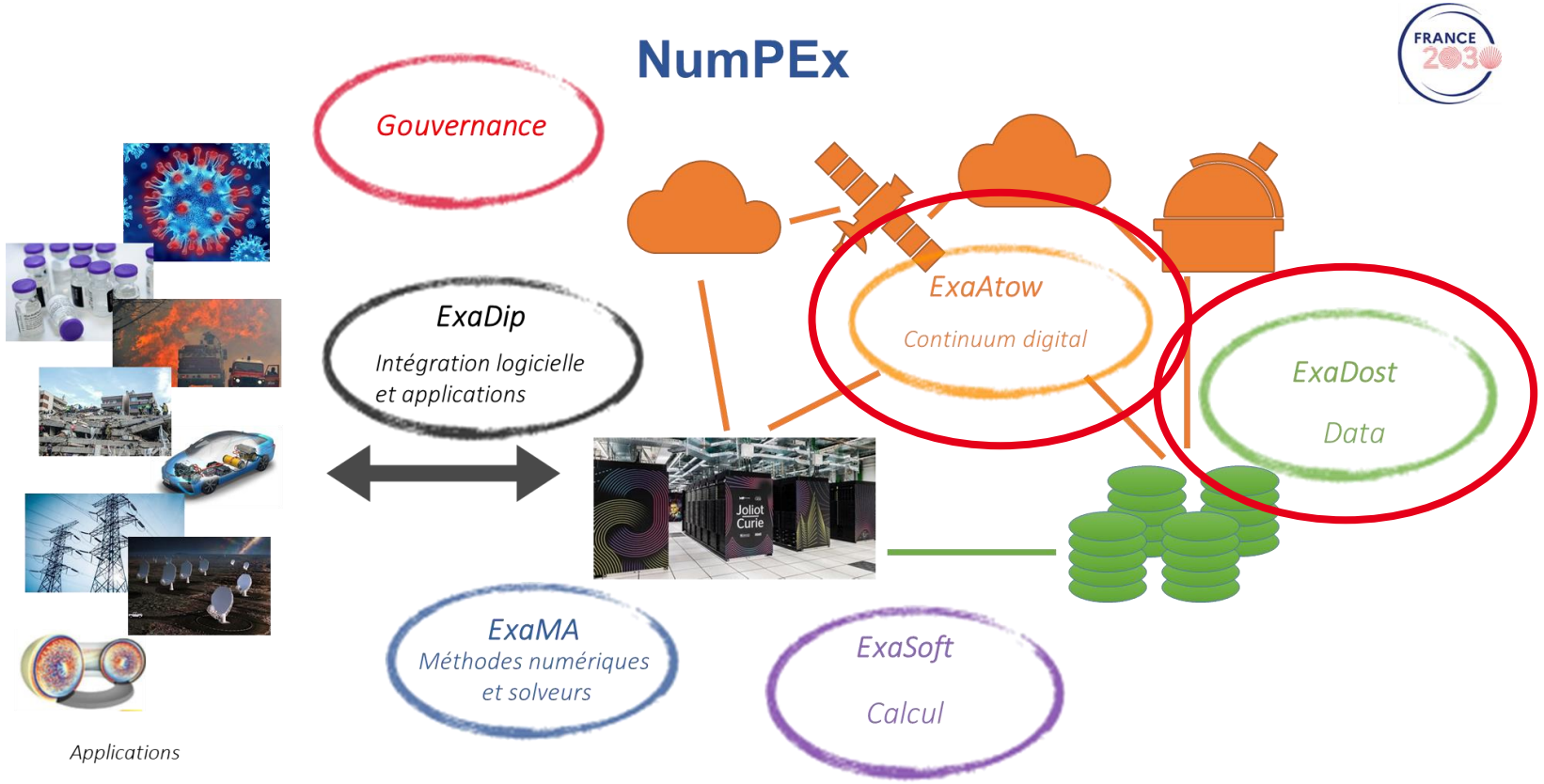
ACM Gordon Bell Prize

Recognizing Outstanding Achievement in High Performance Computing

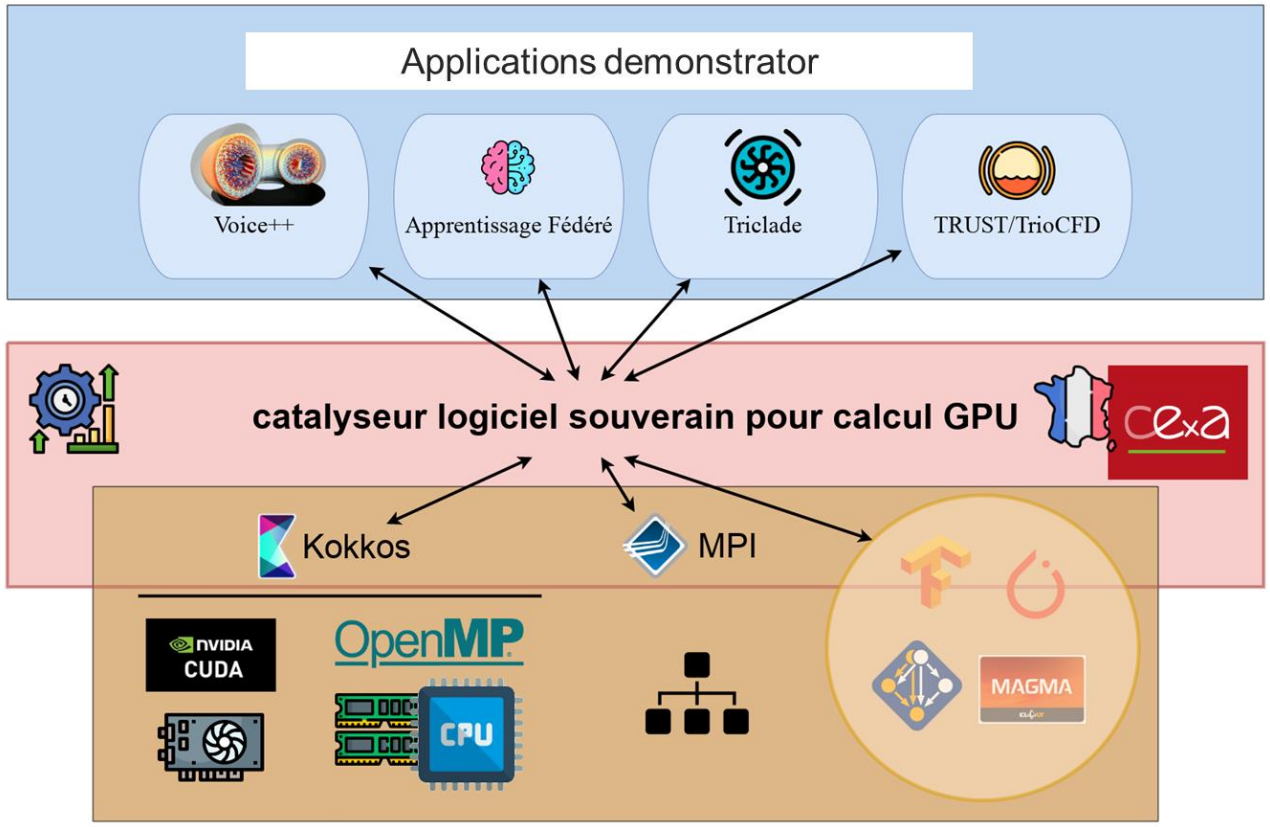
cea

Strategy

- Relies on national and European projects (NumPEx, EuroHPC)
- Launch internal project (CExA)
- New generation of simulation codes based on agnostic approaches



CExA



Hardware Features

- Distributed memory
- Heterogeneous architectures

Software Features

- Interface with data treatment tools
- Simplification of deployment on machines

Next steps

- Gather the scientific community to accelerate the ongoing work around applications portability
  - Arm architectures
  - GPU architectures
- Develop Academic & industrial ecosystem around usage and SW (French – EUR – International)

EUPEX
European Pilot for Exascale

NumPEx
Exascale computing

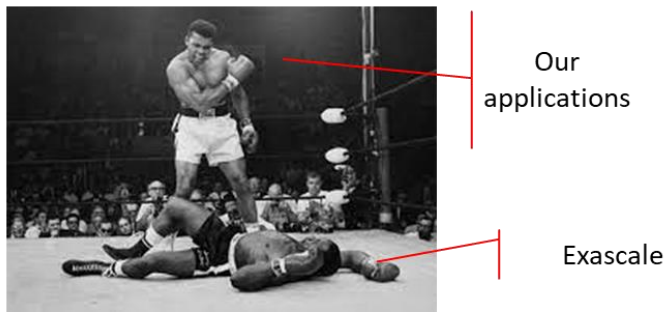
cea

AIDAS
AI Data Analytics and Scalable Simulations

RIKEN

cea

So that we come from this …



cea

Thank you !