



# PYCOA

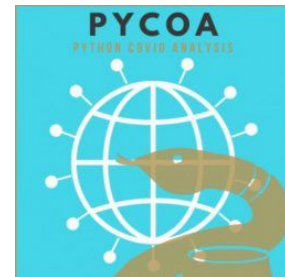
effectuer sa propre analyse Python des données Covid-19

14eme Journées Informatiques IN2P3/RFU nov. 2022 @ Croisic

Olivier Dadoun

Tristan Beau LPNHE / Sorbonne U. / Univ. Paris

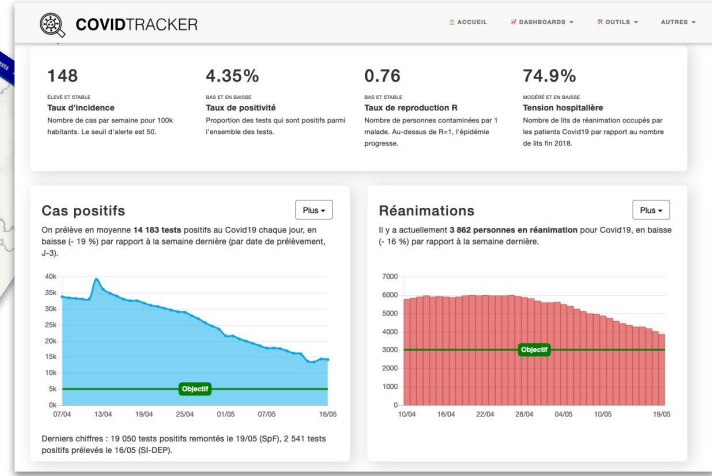
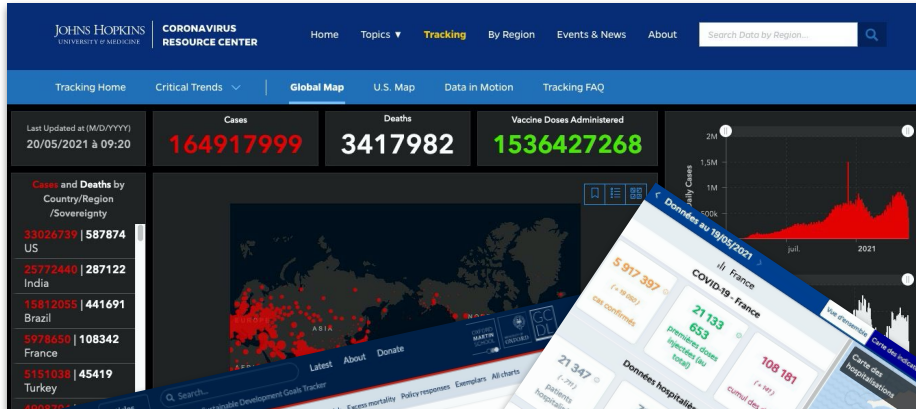
Julien Browaeÿs MSC / Univ. Paris





# Le Covid19 et les dashboards ... beaucoup de dashboards

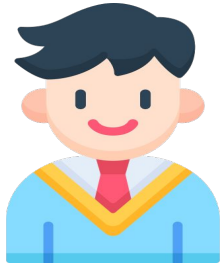
- + Utilisation simple
- + Belles cartes et graphiques
- Interaction limitée
- Mono-base de données
- Pas de réutilisation possible



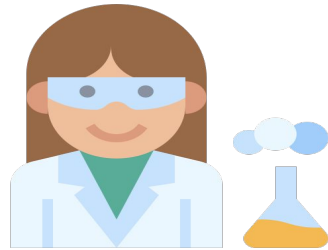


# Que souhaitons-nous promouvoir et pour qui?

- I. L'accès à une information de référence
- II. La capacité à produire en autonomie ses propres représentations
- III. Les logiciels libres et gratuits



Lycéen·nes et  
étudiant·es



Scientifiques



Data journalistes



Analystes  
stratégiques

Une solution unifiée pour permettre la collaboration



Accès en autonomie à l'information pour le plus grand nombre

## Nos objectifs

- Pas d'installation ou peu
- Python au cœur du programme
- Navigateur web (jupyter notebook)
- Belles cartes et beaux graphiques interactif (bibliothèque Bokeh)
- Une interface, frontend simple



... et on garde le backend pour un public averti





... quand on dit simple

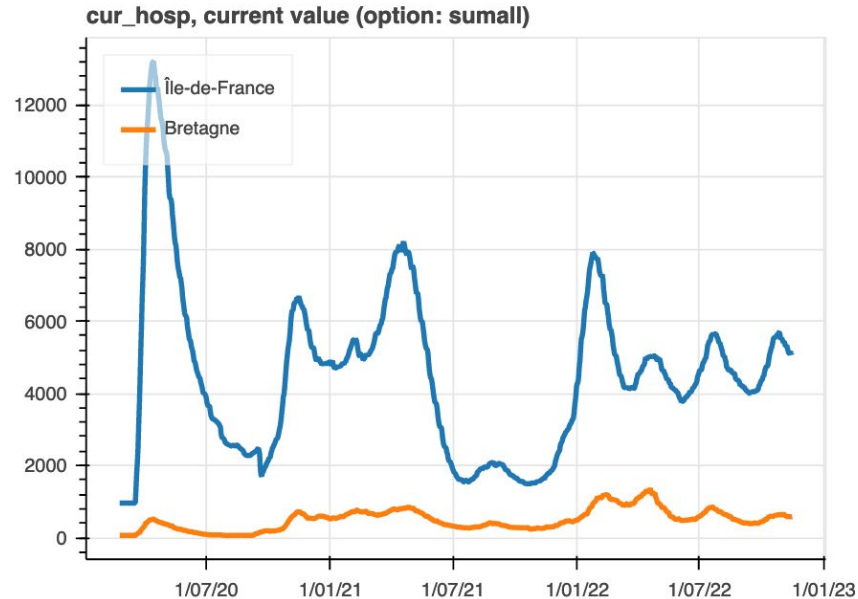
→ Fonctions

- ◆ plot
- ◆ hist
- ◆ map
- ◆ get

→ Mots-clefs

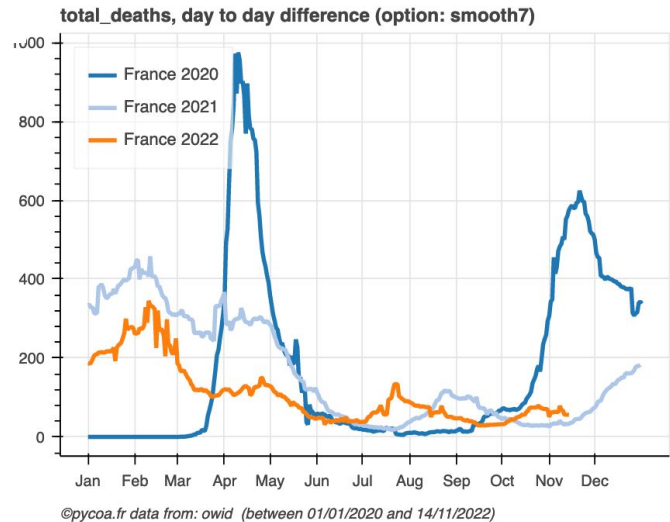
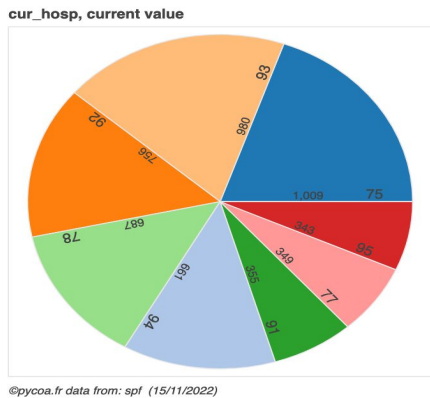
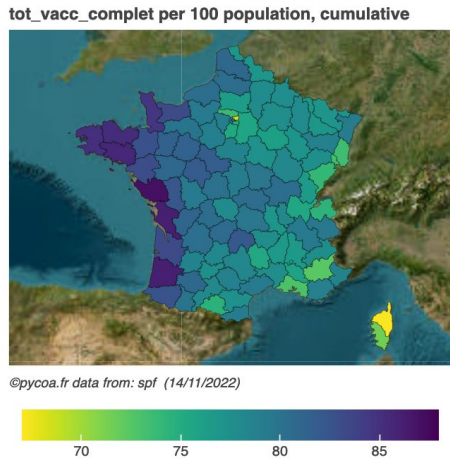
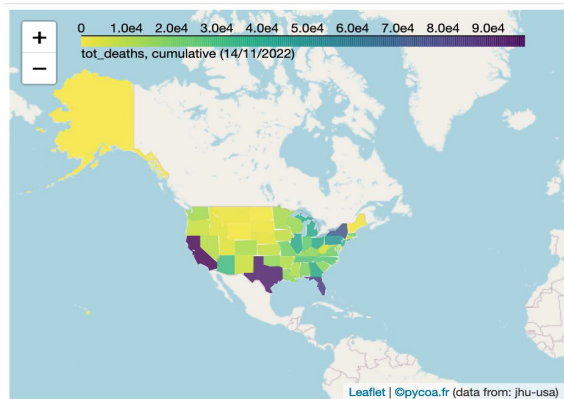
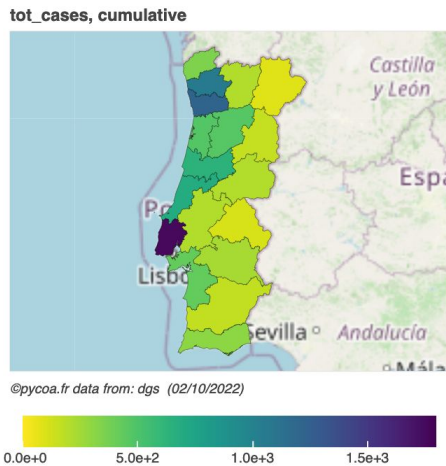
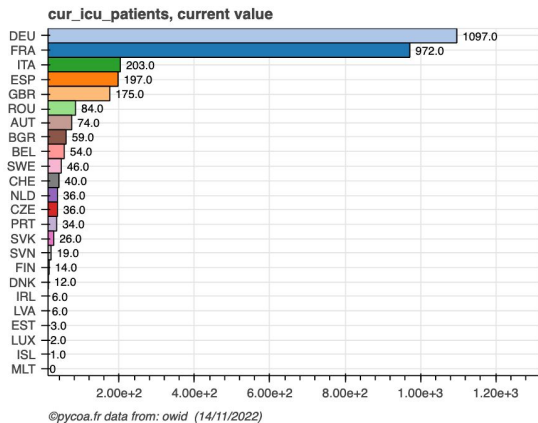
- ◆ where
- ◆ which
- ◆ what
- ◆ options

```
import coenv as pycoa
pycoa.plot(which='cur_hosp',where=[['Île-de-France'],['Bretagne']],option='sumall')
```



©pycoa.fr data from: spf (between 24/02/2020 and 15/11/2022)





# En guise de conclusion

- + obtenir un (petit) soutien institutionnel (Université / CNRS / Ministère)
- + Diffusion et communication (twitter)
- + Stages
- + Pour nous suivre

<https://www.pycoa.fr> ([support@pycoa.fr](mailto:support@pycoa.fr))

<https://github.com/coa-project> (avec un super wiki)

[https://twitter.com/pycoa\\_fr](https://twitter.com/pycoa_fr)

- + Pour tester maintenant [bit.ly/pycoa\\_fds22](https://bit.ly/pycoa_fds22)





DE LA RECHERCHE À L'INDUSTRIE

cea

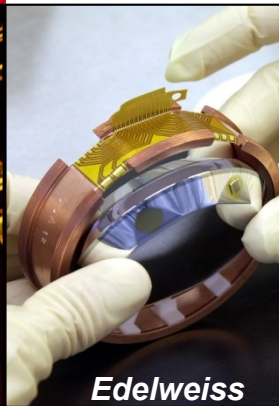
# Tiny ML/DL : Voice Recognition on ESP32



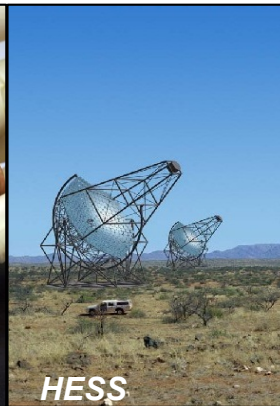
Double Chooz



ALICE



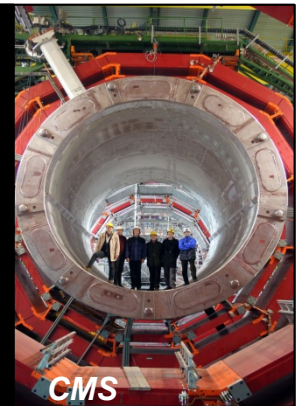
Edelweiss



HESS



Herschel



CMS

*Déchiffrer les rayons de l'Univers*



I. MAGROUNE

16/11/2022

## ➤ **Contexte**

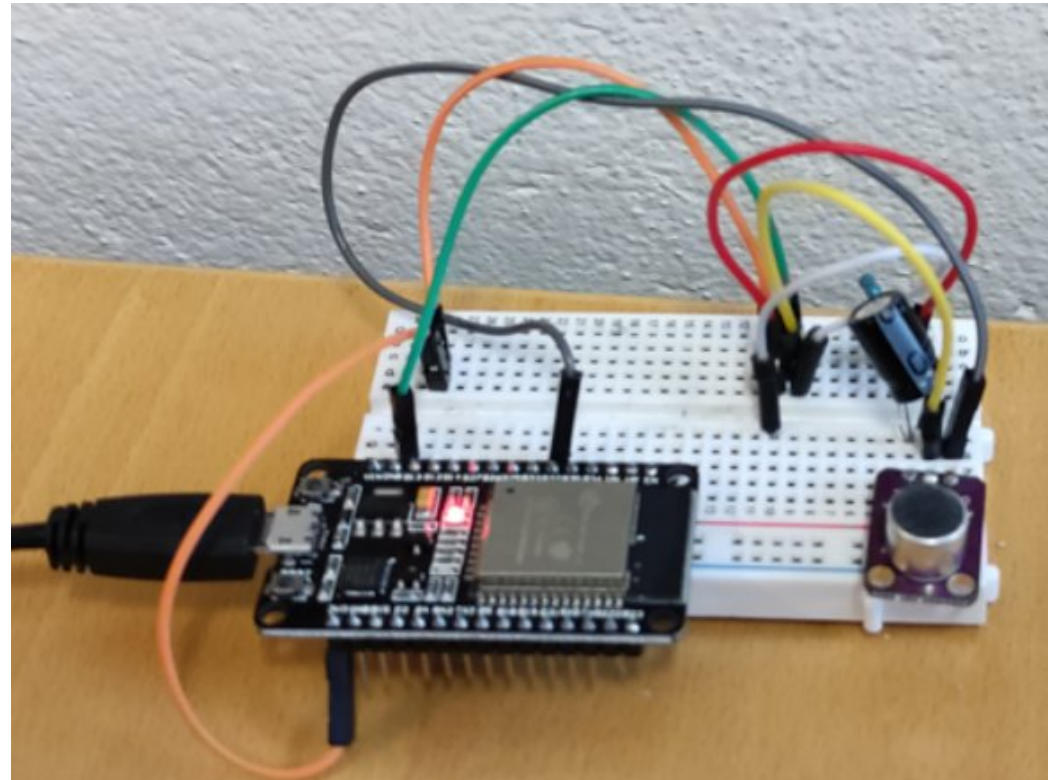
- *IoT @ AIoT*
- *Faire plus avec moins*

## ➤ **Objectif**

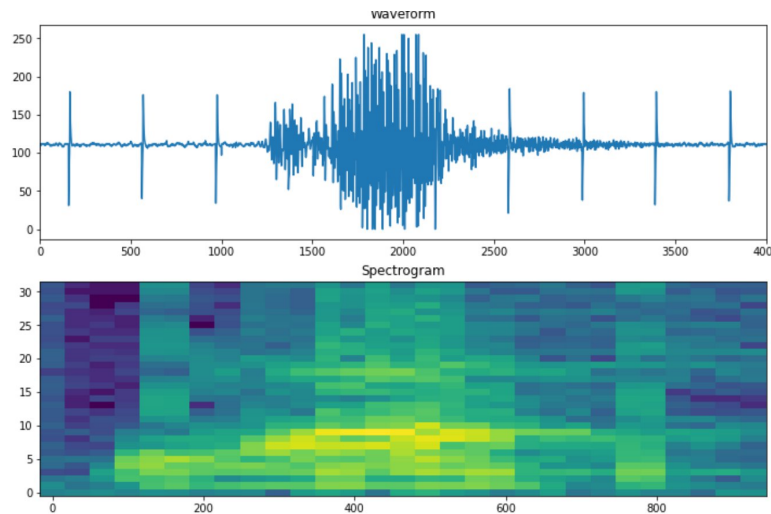
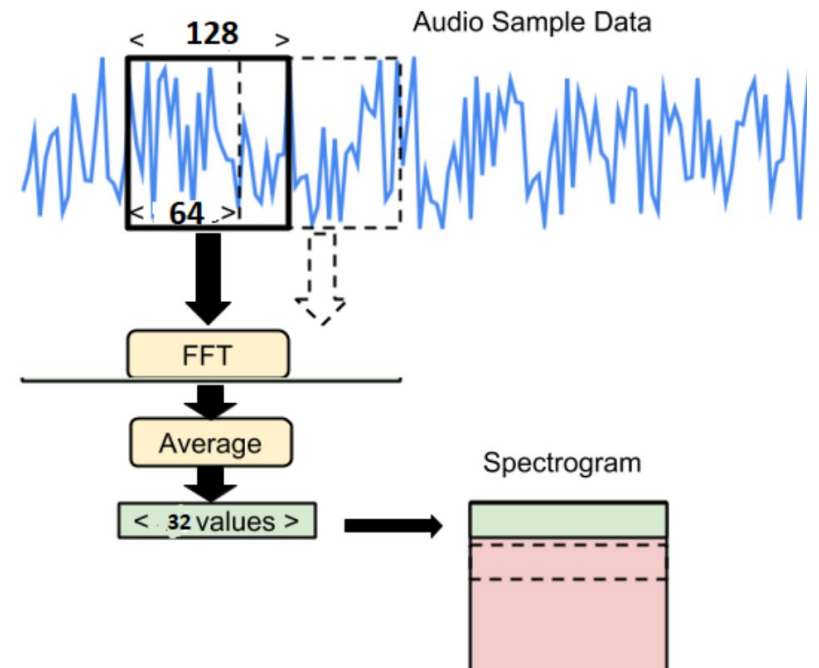
- *ML/DL sur des MCUs*
- *Model inférence sur MCU*

## ➤ **Contraintes**

- *ARM® Xtensa LX6 ® RISC-V*
- *520 Ko SRAM*
- *Ressources limitées*
- *Budget encore plus frugal*



- **Signal sonore Analogique**
  - **Fréquence** (entre 20 Hz et 20 kHz)
- **Echantillonnage**
  - **44100 Hz**
  - **Se limiter à 8000 Hz**
- **Spectrogramme**
  - **Extraction des fréquences**
  - **FTT**



- **Problème de classification d'images**

## ➤ **Choix de la carte de dev (MCU)**

- **Arduino Nano Sense (25 € - aujourd'hui 40 € - )**
- **ESP32 (5€ - MCU à 1 € - )**



## ➤ **Acquisition**

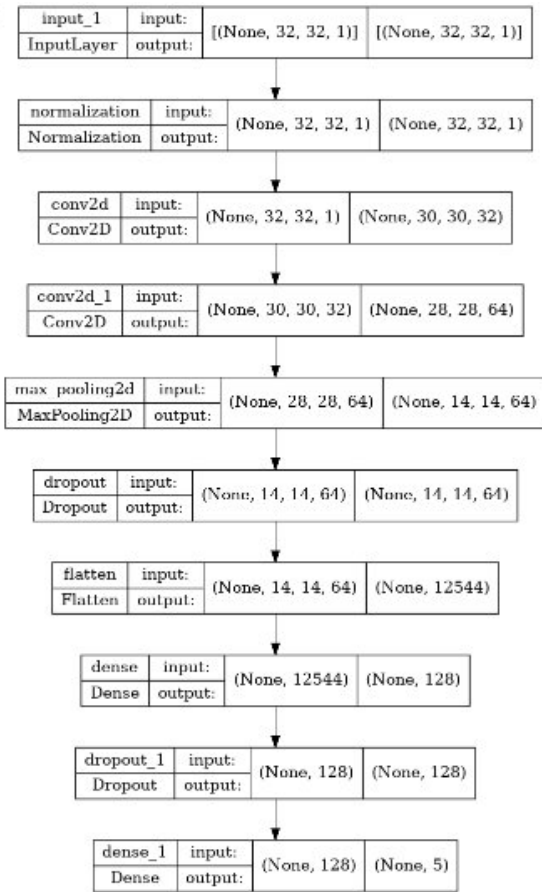
- **GY-MAX4466 (2 €)**
  - Analog
- **INMP441 (6 €)**
  - analog to digital converter
  - anti-aliasing filter,
  - power management and industry standard 24-bit I2S interface



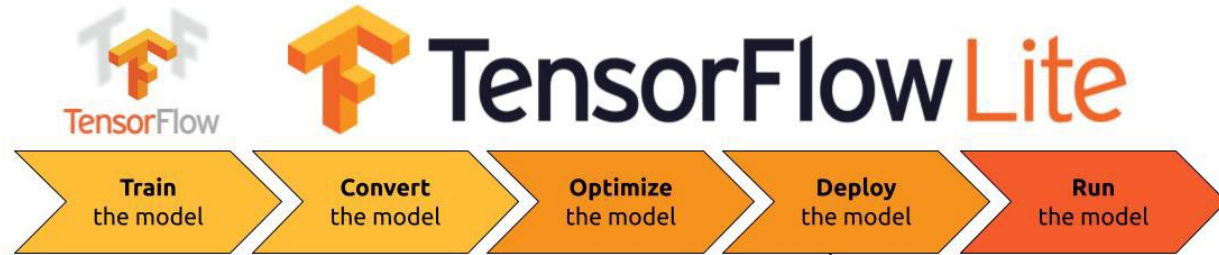
## ➤ **Alimentation**

- 3.3 V (1.2 en interne)
- Filtrer par capacité





Total params: 16,963  
Trainable params: 16,963  
Non-trainable params: 0



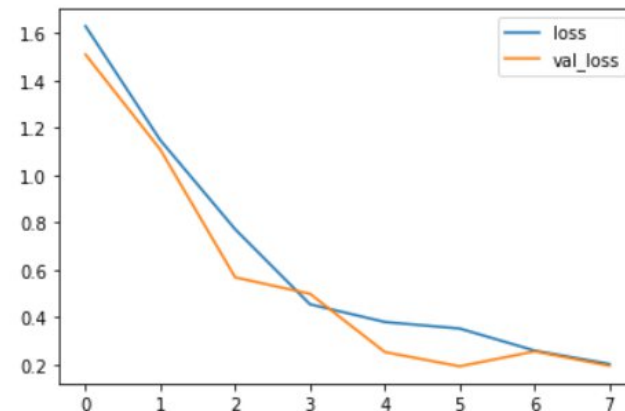
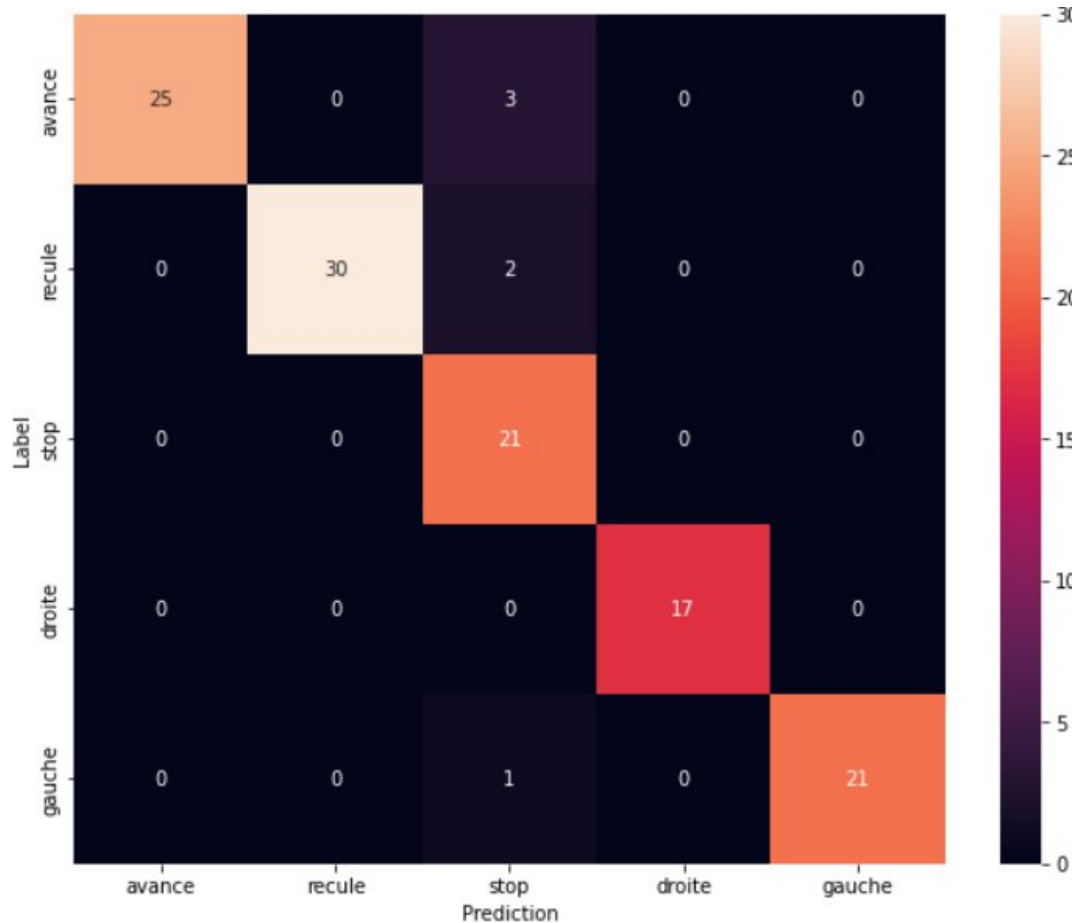
The model must be trained with spectrograms computed with the same code on esp32 MCU  
Optimized Arduino Code to compute FFT is compiled into .so and used as tf lib.

Model\_lite.cpp

```
const unsigned char model_data[] DATA_ALIGN_ATTRIBUTE = {0x1c, 0x00, 0x00, 0x00, 0x54, 0x46, 0x4c, 0x33, 0x14, 0x00, 0x20, 0x00, 0x1c, 0x00, 0x18, 0x00, 0x14, 0x00, 0x10, 0x00, 0x0c, 0x00, 0x00, 0x00, 0x08, 0x00, 0x04, 0x00, 0x14, 0x00, 0x00, 0x00, 0x1c, 0x00, 0x00, 0x00, ,,,,,,
```

Arduino Implémentation

TensorFlowLite\_ESP32 library





Pour plus de détails et code source :

Tuto ©

<http://magroune.net/AIoT-AI-Tensori-Flow-Lite-ESP32-Tuto5.php>

```
sum = 0.0;
for (i = 0; i < size; ++i) {
    sum += exp(input[i] - m);
}

constant = m + log(sum);
for (i = 0; i < size; ++i) {
    input[i] = exp(input[i] - constant);
}
}
```

### output samples

```
Output Serial Monitor x
Message (Ctrl + Enter to send message to 'ESP32 Dev Module' on 'COM3')
14:43:51.220 -> 0.00 0.00 1.00 0.00 0.00 stop 1498
14:43:51.748 -> 0.00 0.00 1.00 0.00 0.00 stop 1498
14:43:52.282 -> 0.00 0.00 1.00 0.00 0.00 stop 1499
14:43:52.811 -> 0.00 0.00 1.00 0.00 0.00 stop 1499
14:43:53.352 -> 0.00 0.00 1.00 0.00 0.00 stop 1500
14:43:53.881 -> 0.00 0.00 0.00 0.00 1.00 gauche 1500
14:43:54.395 -> 0.00 0.00 0.00 0.00 1.00 gauche 1501
14:43:54.926 -> 0.00 0.00 0.00 0.00 1.00 gauche 1501
14:43:55.468 -> 0.00 0.00 0.00 0.00 1.00 gauche 1502
14:43:56.025 -> 0.00 0.00 1.00 0.00 0.00 stop 1502
14:43:56.546 -> 0.00 0.00 1.00 0.00 0.00 stop 1503
14:43:57.072 -> 0.00 0.00 1.00 0.00 0.00 stop 1503
14:43:57.634 -> 0.00 0.00 0.36 0.00 0.64 gauche 1504
```

Commissariat à l'énergie atomique et aux énergies alternatives  
Centre de Saclay | 91191 Gif-sur-Yvette Cedex

Etablissement public à caractère industriel et commercial | R.C.S Paris B 775 685 019



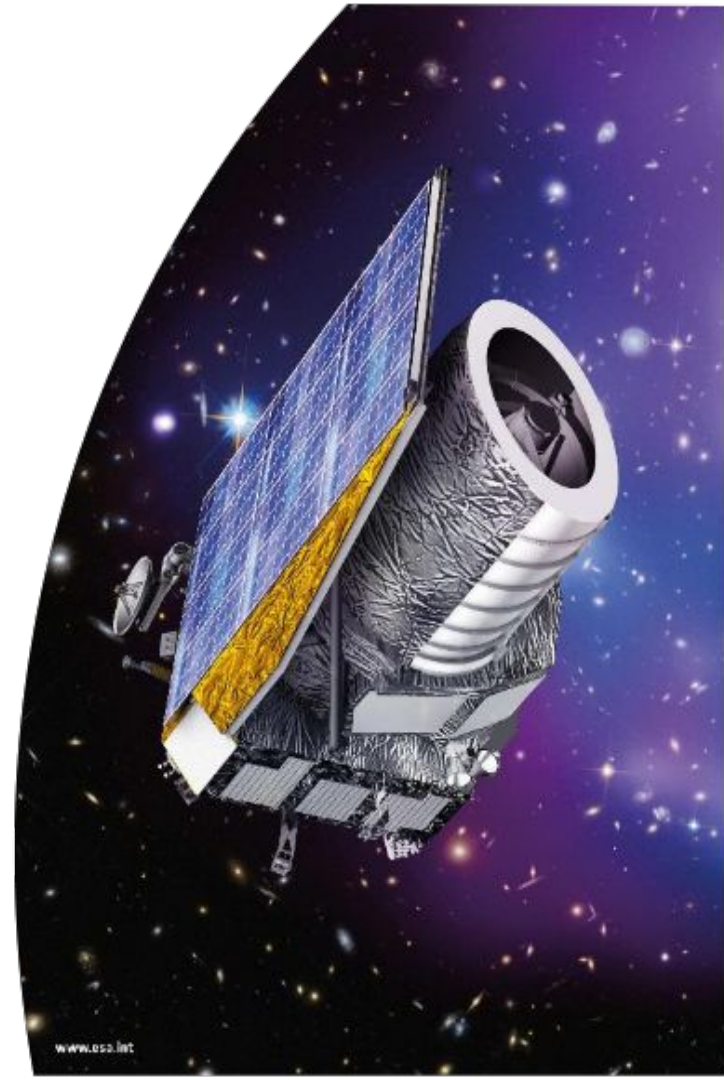
Direction de la Recherche Fondamentale  
Institut de recherche  
sur les lois fondamentales de l'Univers  
Service

# CODEEN

COLlaborative DEvelopment  
ENVironnement for Euclid  
spacecraft

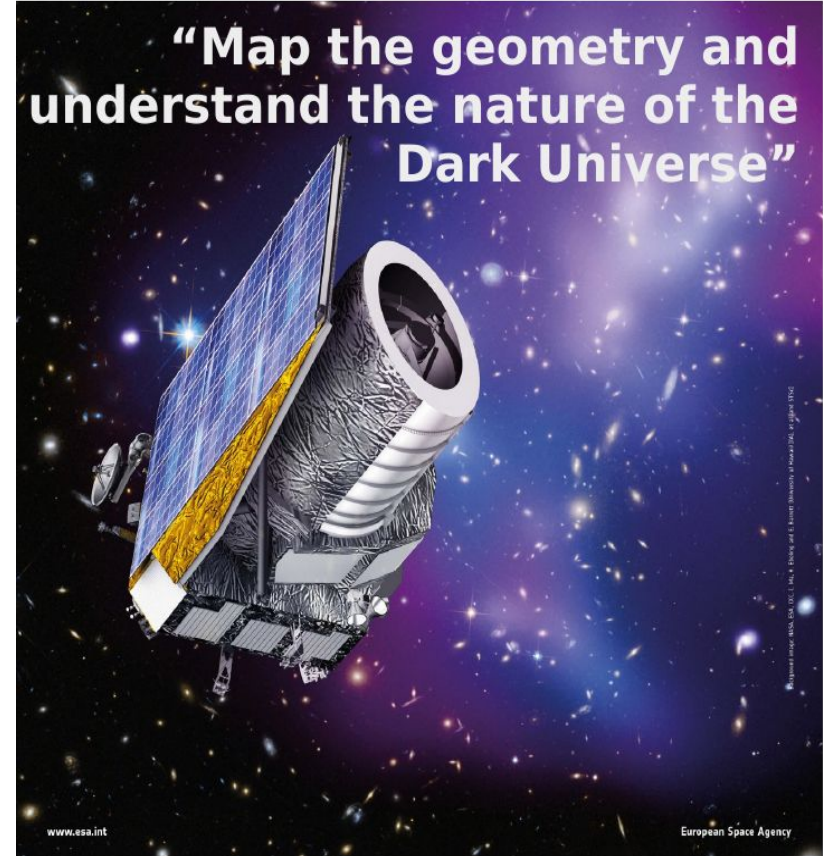


Martin Souchal (APC), Paul Zakharov (APC)

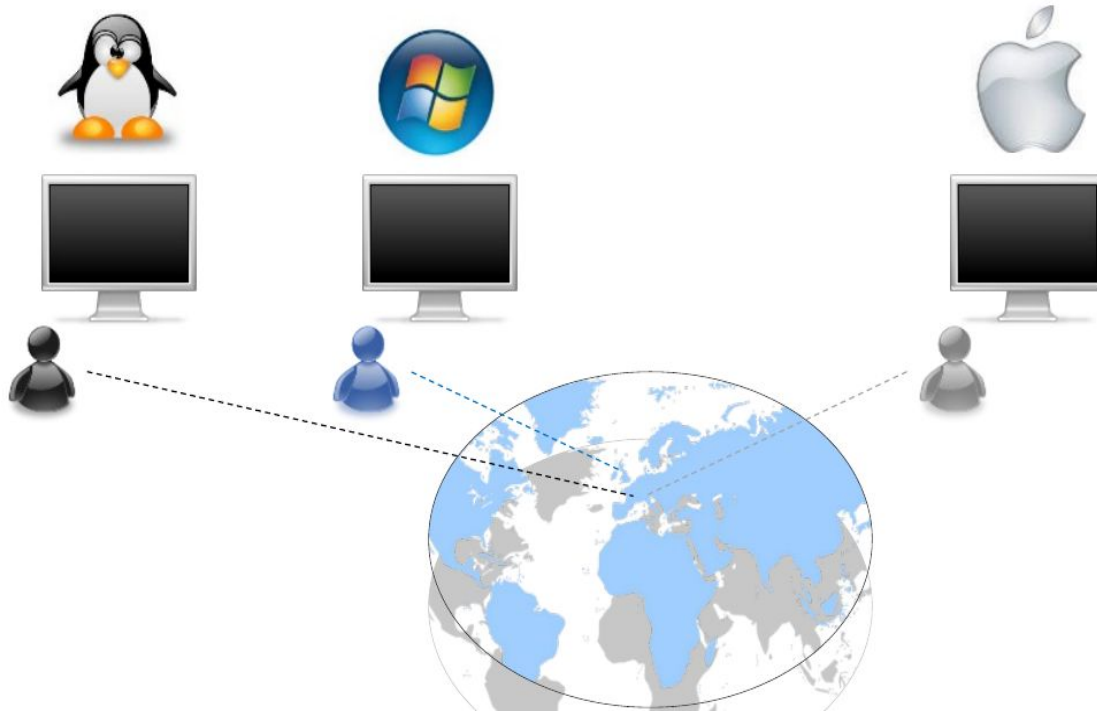


- **Euclid is an ESA medium class astronomy and astrophysics space mission**
- The imprints of dark energy and gravity tracked by two complementary probes: Weak gravitational Lensing and Galaxy Clustering
- Euclid will be equipped with a 1.2 m diameter mirror telescope feeding 2 instruments : a high quality panoramic visible imager (VIS), a near infrared 3-filter photometer (NISP) with spectrograph
- Launch is planned for 2023
- *“Aims at understanding why the expansion of the Universe is accelerating and what is the nature of the source responsible for this acceleration which physicists refer to as dark energy.”*

Source : <https://www.euclid-ec.org/>



## Comment fédérer des environnements différents?



# EDEN : Euclid development environment

- **EDEN** : Euclid development environment, is the **reference environment**. Among others, it defines the coding rules, the operating system, languages and libraries.
- **LODEEN** : local development environment. It is a **virtual machine** which implements **EDEN** and even
- **DOCKEEN** : EDEN inside a Docker container

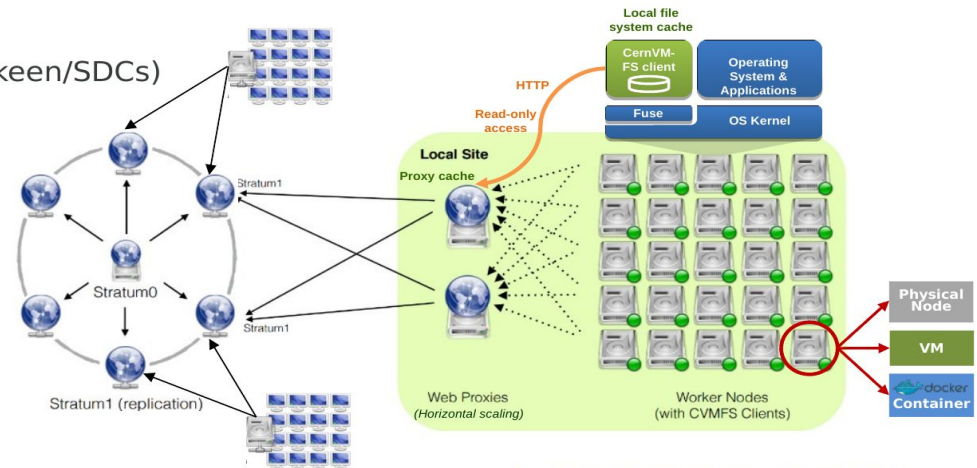




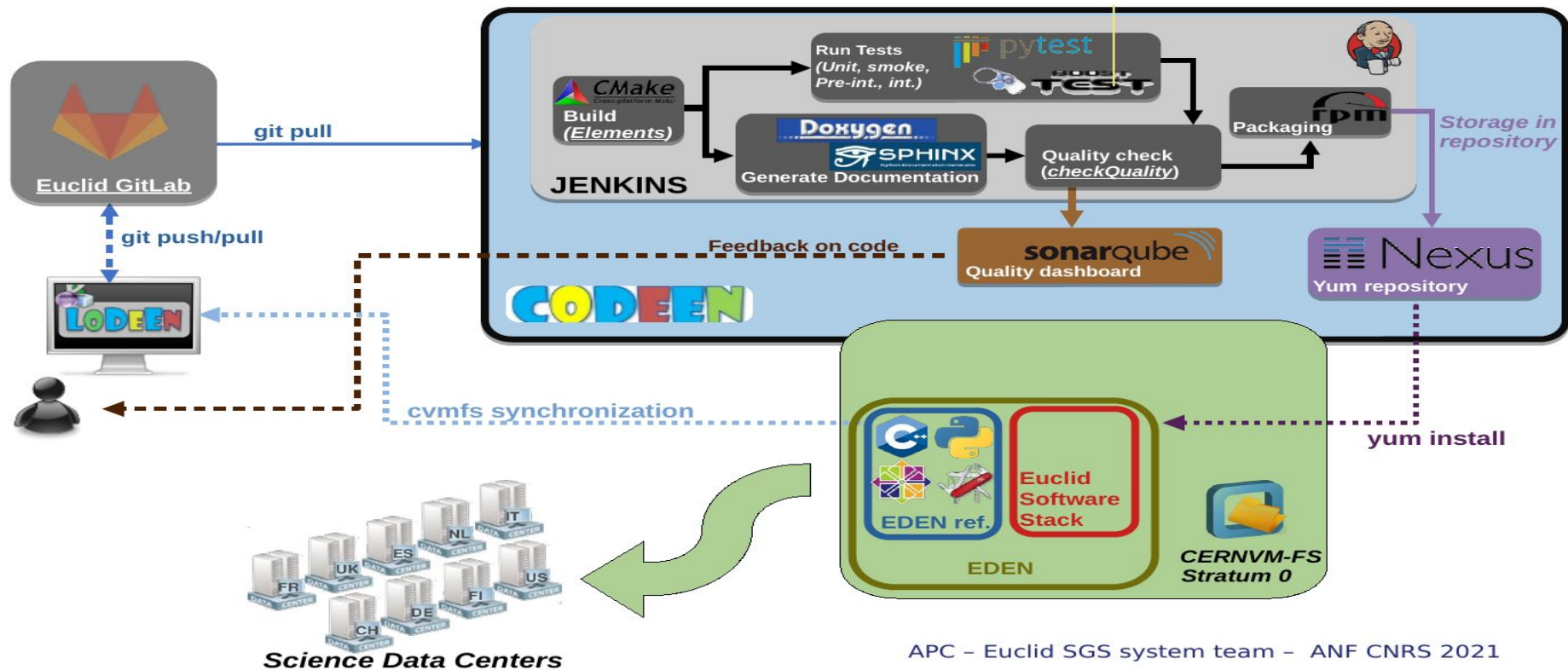
- CernVM-FS is a network filesystem, which you can mount in Linux or macOS via FUSE (Filesystem in Userspace) and on Windows in a WSL2 virtualized Linux environment.  
In some ways it is similar to other network filesystems like NFS or AFS, but there are various aspects to it that are quite different.
- Read-only filesystem over HTTP
- Developed and maintained by CERN
- The primary use case of CernVM-FS is to easily *distribute software* around the world

## Structure :

- a central *Stratum 0*
- several *Stratum 1*
- proxy servers
- final clients  
(Lodeen/Dockeen/SDCs)



## CODEEN: COlaborative DEvelopment ENvironment



## Quelques mots sur des outils : Jenkins

**Jenkins** is an *open source* automation server. It helps automate the parts of *software development* related to *building, testing, and deploying*, facilitating *continuous integration* and *continuous delivery*.

Continuous integration orchestration.

→ ~300 projets Jenkins (CODEEN)



# Jenkins

The screenshot shows the Jenkins dashboard interface. At the top, there's a search bar and user information (aboizard, log out). The main area displays a table of jobs with columns for build status (PF, S, W), job name, and build details (DG, G, D, Last Success, Last Failure, Last Duration). The 'Build Queue' on the left shows 'No builds in the queue.'

PF	S	W	Name	DG	G	D	Last Success	Last Failure	Last Duration
All									
CH									
CP									
CT									
EL									
			EL_Alexandria				3 mo 10 days - log	N/A	13 sec
			EL_ArrayLib				8 mo 0 days - log	N/A	3.9 sec
			EL_Background				5 mo 13 days - log	N/A	4.5 sec
			EL_CatalogLib				3 mo 27 days - log	N/A	7.9 sec
			EL_EuclidCatalogLibrary				8 mo 0 days - log	N/A	3.8 sec
			EL_FITsIO				3 mo 18 days - log	N/A	7 sec
			EL_imageLib				8 mo 0 days - log	N/A	3.7 sec

A **software repository** is a storage location for **software packages**. A software repository is typically managed by **source control** or **repository managers**.

NEXUS : software repository developed by Sonatype



- EDEN referential and Euclid software stack packages.

The screenshot shows the Sonatype Nexus Repository Manager interface. The top navigation bar includes the Sonatype logo, version "OSS 3.34.1-01", a search bar for components, and user information for "aboizard". A left sidebar contains navigation options: "Browse" (selected), "Welcome", "Search", and "Upload". The main content area is titled "Browse" and displays a table of repository assets.

Name ↑	Type	Format	Status	URL	Health check	IQ Policy VI...
central	proxy	maven2	Online - Ready t...	<a href="#">copy</a>	Analyze	Loading... >
conda-euclid	hosted	raw	Online	<a href="#">copy</a>	⊘	Loading... >
conda-forge	proxy	conda	Online - Remot...	<a href="#">copy</a>	Analyze	Loading... >
conda.raw.eden...	hosted	raw	Online	<a href="#">copy</a>	⊘	Loading... >
docker-reposito...	hosted	docker	Online	<a href="#">copy</a>	⊘	Loading... >
el7.eden.2.1	hosted	yum	Online	<a href="#">copy</a>	⊘	Loading... >
el7.eden.2.1.DEV	hosted	yum	Online	<a href="#">copy</a>	⊘	Loading... >

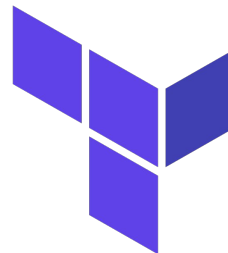
**SonarQube** is an *open-source* platform developed by *SonarSource* for continuous inspection of *code quality* to perform automatic reviews with static *analysis of code* to detect *bugs*, *code smells* on 29 *programming languages*.

- Euclid tool for **quality dashboards**  
→ base for **maturity level** attribution.



The screenshot displays the SonarQube web interface. At the top, there is a navigation bar with tabs for "Projects", "Issues", "Rules", "Quality Profiles", "Quality Gates", and "Administration". A search bar is located on the right side of the navigation bar. Below the navigation bar, the main content area is divided into two columns. The left column contains filters for "Quality Gate" (Passed: 462, Failed: 1.1k) and "Reliability ( Bugs )" (A: 1k, B: 67, C: 429, D: 2, E: 29). The right column displays a list of projects. Two projects are visible: "Alexandria\_2.17.0" and "Alexandria\_2.17.1". Both projects are marked as "Passed". The "Alexandria\_2.17.0" project shows 0 Bugs (A), 0 Vulnerabilities (A), 335 Code Smells (A), 56.0% Coverage, and 0.0% Duplications. The "Alexandria\_2.17.1" project shows 1 Bug (B), 0 Vulnerabilities (A), 335 Code Smells (A), 56.0% Coverage, and 0.0% Duplications. Both projects have a last analysis date and a version number (6.9k S C++ (Community)).

- Instance Openstack CC IN2P3
- Déploiements de services via terraform + ansible
  - Terraform : création des instances openstack
  - Ansible : configuration de l'os
- Os déployés : debian 11 et centos 7 (passage a rocky 9 en cours)



## Vue d'ensemble

A N S I B L E

### Synthèse des Quotas

#### Compute



Instances  
Utilisé 43 sur 90



VCPUs  
Utilisé 113 sur 180



RAM  
Utilisé 195,8Go sur 390,6Go

#### Volume



Volumes  
Utilisé 35 sur 50



Instantanés du volume  
Utilisé 0 sur 5



Stockage de volumes  
Utilisé 8,8To sur 9,8To

#### Réseau



# Development of a framework based on PyTorch Lightning for training Graph Neural Networks (GNNs) for tracking

14<sup>ème</sup> Journées Informatiques IN2P3/IRFU (14-17 novembre 2022)

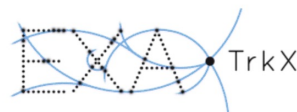
Sylvain Caillou

(On behalf of L2IT « Computing, Algorithms and Data » Team)

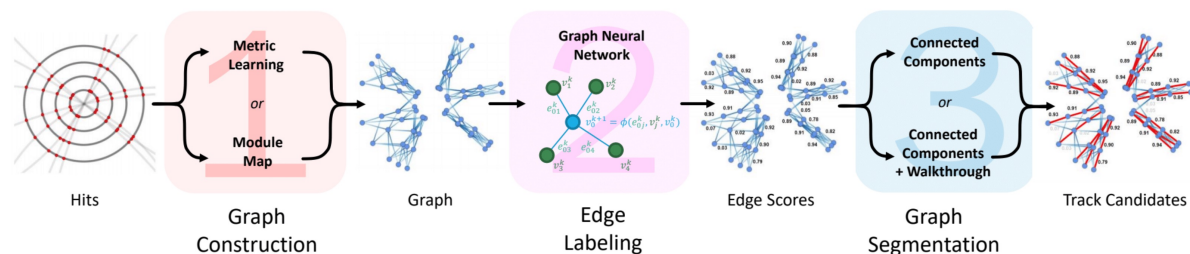
# Context

- ⇒ Geometric ML have become one of the hottest fields of AI research
- ⇒ Increasingly popular for a large number of LHC physics tasks
- ⇒ GNNs perform pretty well to learn geometric pattern of particles tracks

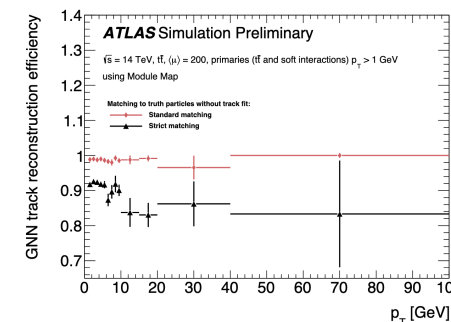
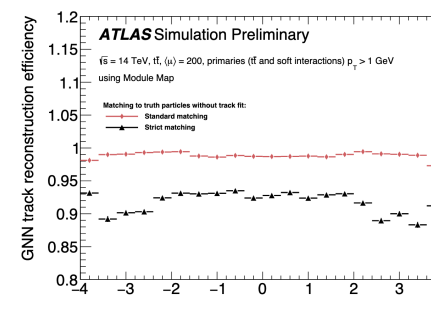
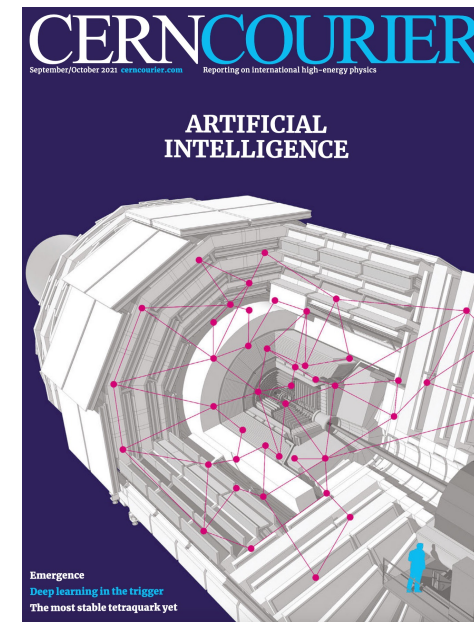
Collaboration L2IT & ExatTrkX



⇒ Construct a GNN-based track reconstruction algorithm for ATLAS Itk



- ⇒ First results on ITk published in 2022 more than encouraging
- ⇒ GNN-based algorithms now appear as competitive solutions
- ⇒ Put into production (ACTS) for the HL-LHC



[ATLAS ITk Track Reconstruction with a GNN-based pipeline, C.Rougier et al., CTD 2022](#)  
[Graph Neural Network track reconstruction for ATLAS ITk, D. Murnane et al., IML 2022](#)

# Motivation

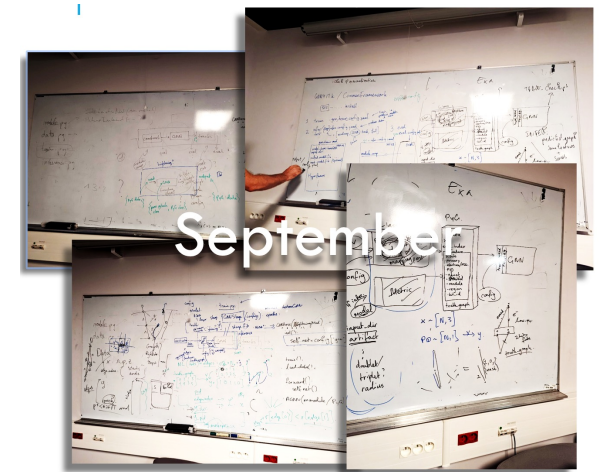
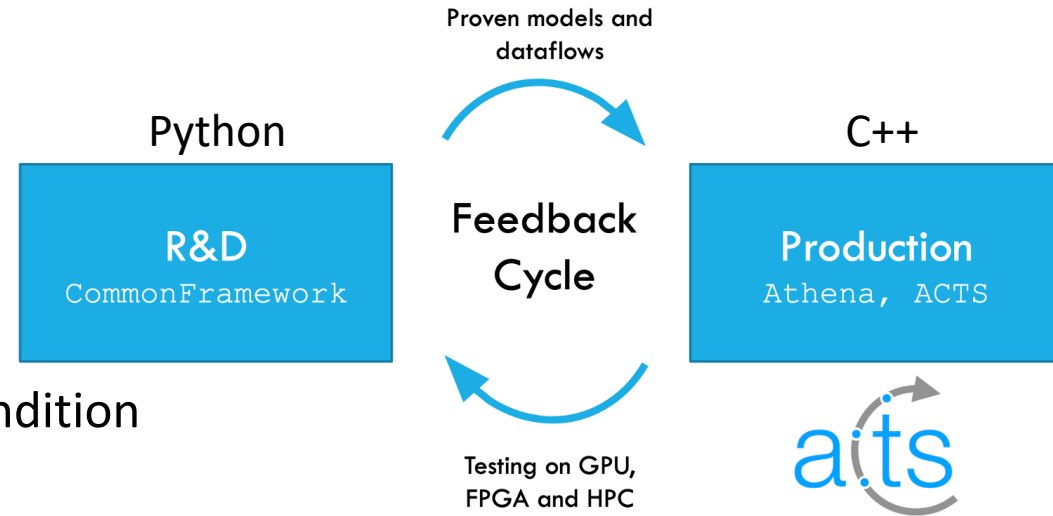
ML R&D in HEP take a LOT of time and (GPU) resources

- ML R&D: Models and hyperparameters exploration
- Data: Complexity of the simulated detector data in HL-LHC condition (Pythia + GEANT4 + ITk geometry)
- (International) Collaboration:
  - Do we speak the same language (share the same semantic) ?
  - Data format
  - Experimental conditions
  - Performance metrics definition & implementation
- Reproducibility: Lack of traceability of experimental conditions
- Software quality: Messy idea implemented, no documentation, no test, etc...

Improve the process => save some R&D time

- ⇒ Get faster for research
- ⇒ Use less (GPU) resources and do more sober R&D

**⇒ Need to develop of a common R&D framework**

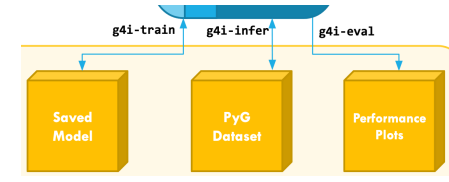


# gnn4ITk common framework (developped with D. Murnane, LBNL)

**PyG Data**

<p><b>Spacepoint features</b></p> <ul style="list-style-type: none"> <li>• <math>r, \phi, z, \dots</math> (required)</li> <li>• Hit ID (required)</li> <li>• Cell shape</li> <li>• Module</li> <li>• Region</li> <li>• Shared</li> <li>• Ghost</li> </ul>	<p><b>Track features</b></p> <ul style="list-style-type: none"> <li>• Truth graph (required)</li> <li>• Particle ID (required)</li> <li>• Primary</li> <li>• <math>p_T</math></li> <li>• Num hits</li> <li>• PDG ID</li> </ul> <p><b>Graph Features</b></p> <ul style="list-style-type: none"> <li>• Edge index (required)</li> <li>• Y (required)</li> </ul>
---	---

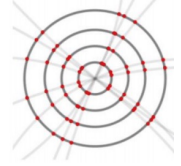
Modularity: Pipeline divided in stage  
 Each stage can include several models  
 It can be *train* or *infer* or *evaluate*  
 Share the same evaluation method



Also try to:

- Be Easy to use
- Be clean (keep messy dev & random ideas out of the repository)
- Be well documented
- To Include Test

Well defined common data format



Hits  
Events (csv)

Config (yaml)

**EventReading**

Input testing

Truth and weight building

Athena Reader

ACTS Reader

PyG Data

CSV Data

Config (yaml)

**Graph Construction**

Input testing

Truth and weight building

Module Map

PyLightning

Metric Learning

Graph

PyG Data

Config (yaml)

**Edge Labeling**

Input testing

Truth and weight building

PyLightning

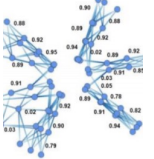
Filter

PyLightning

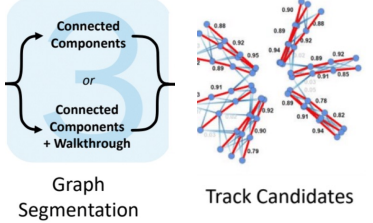
HomoGNN

PyLightning

HeteroGNN



Edge Scores



Traceability of experimental conditions

Saved Model

Saved Model

Save each stage output

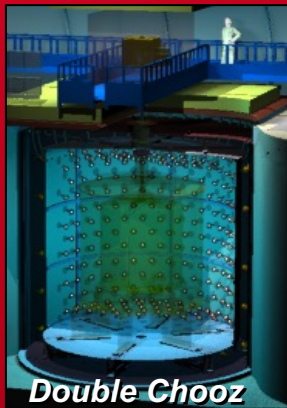
# Next steps...

- Reproduce performance on approval dataset
- Use the framework (and no more home code) for the next studies
- Publish it with open data demo
- Towards a common platform / tools for Geometric ML R&D in HEP ?

DE LA RECHERCHE À L'INDUSTRIE



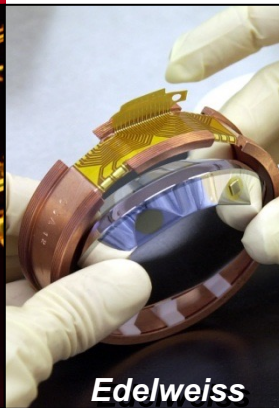
# Frugalité dans un contexte HPC



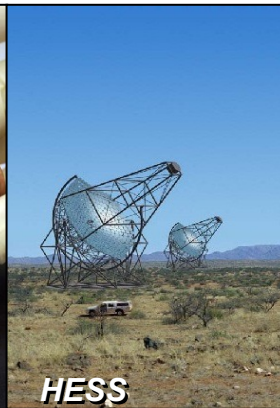
*Double Chooz*



*ALICE*



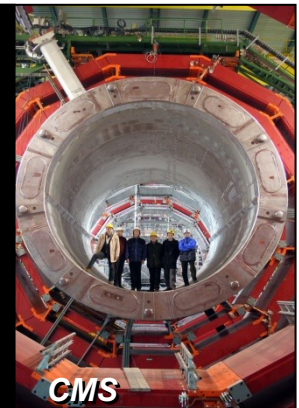
*Edelweiss*



*HESS*



*Herschel*



*CMS*

*Déchiffrer les rayons de l'Univers*



**P-F Honoré**

Mercredi 16 novembre 2022



Nos salles : Grif+HPC+interactif

- PLUS GROS CONSOMMATEUR DE L'IRFU, ET DU CENTRE

Réduction des coûts = diminution de la consommation

Économies : combien et quand ?

Relevé des compteurs TGBT :

- cartographie de la distribution électrique
- initialement pour calculer PUE

IPMI permet accès Sensor Data Repository :

- puissance consommée, courant PDU
- vérification avec pince ampèremétrique

Très limité pour un cluster HPC :

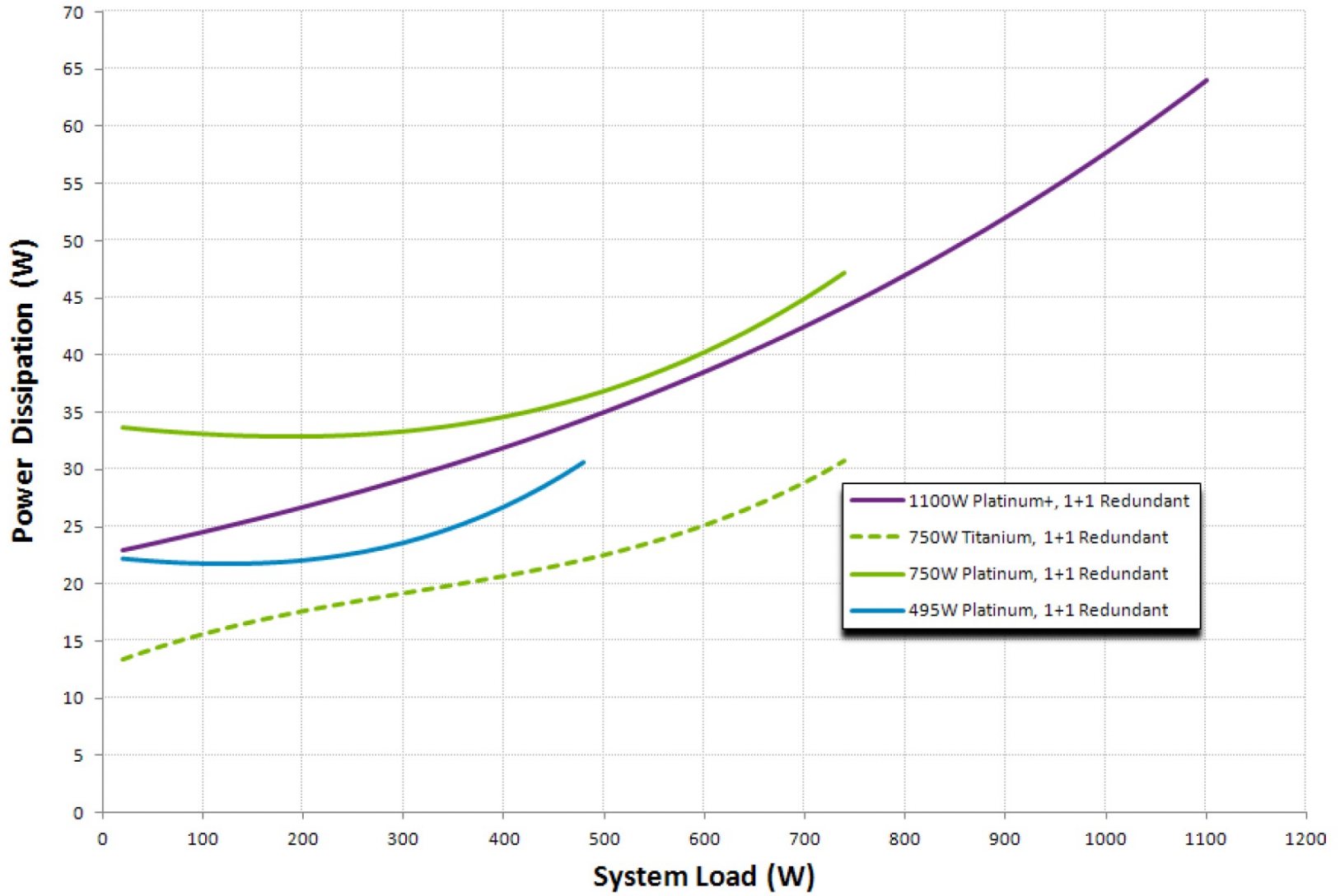
- C6525 architecturé pour produire SpecInt/SpecFP
- 2 AMD 32c + 16x16G + IB 100Gb/s > 400W
- Jusqu'à 2x 4A pour les PDUs

Faire les bons choix à l'achat :

- [Power Efficiency "How To" for Dell PowerEdge](#)

Optimisation possible mais dangereuse pour les jobs MPI multinœuds :

- Adaptation de la fréquence :
  - BIOS MODE PERFORMANCE -> PERF-PER-WATT
  - GOVERNOR **SCHEDUTIL**



Slurm permet d'activer un « SuspendTime » pour éteindre les nœuds inutilisés :

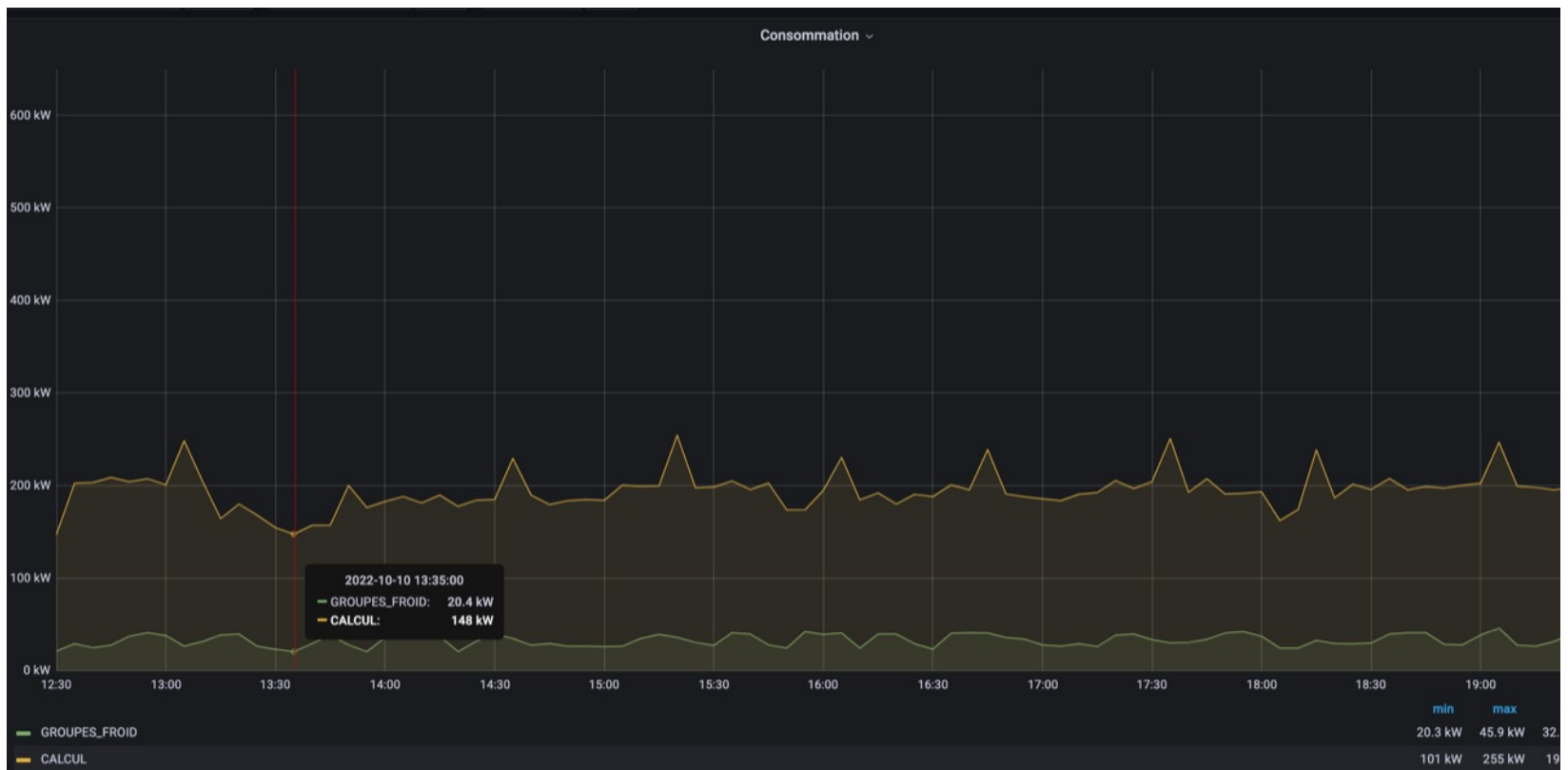
- Efficacité dépend de la charge
- Souple :
  - **SUSPENDPROGRAM**
  - **RESUMEPGRAM**

Éteindre ou  
réduire la consommation en IDLE ?

## Passer les pics de consommation

Bios permet « Power Capping » : conso globale -25%

clock / 6 => HS06 / 6 => consommation / 3



---

Commissariat à l'énergie atomique et aux énergies alternatives  
Centre de Saclay | 91191 Gif-sur-Yvette Cedex

Etablissement public à caractère industriel et commercial | R.C.S Paris B 775 685 019



Direction de la Recherche Fondamentale  
Institut de recherche  
sur les lois fondamentales de l'Univers  
Service



## RI3 : présentation

- Regroupe les informaticiens de l'IN2P3 et de l'IRFU (~280 inscrits sur la mailing liste)
- Partenaire des fédérations des réseaux métiers RESINFO et DevLOG
- Soutien financier IN2P3 et IRFU





## RI3 : présentation

- **Comité de Coordination du Réseau (CCRI)**
  - Une quinzaine de personnes (réunion mensuelle en visioconférence)
  - Représentants de la plupart des laboratoires
  - Organisation et suivi des actions
  
- **Interaction avec :**
  - Le Centre de Calcul
  - Les Chefs de Service des laboratoires
  - La formation permanente
  - La direction de l'IN2P3 (DAT, DAS « Calcul et Données » )



# RI3 : Vie du réseau

## ➤ Moyens de communication

- Liste de diffusion : [ri3-l@in2p3.fr](mailto:ri3-l@in2p3.fr)
- Site web : <https://ri3.in2p3.fr>
- Site TECH-NEWS : <https://tech-news.in2p3.fr>

## ➤ Actions récurrentes

- [Journées Informatique](#) (bisannuelles)
- [Écoles Informatique](#) (ANF annuelle)
- [Webinaires](#)
- [Groupes de travail](#)



## RI3 : Groupes de travail

- **Sécurité informatique IN2P3**

Animation : Guillaume Philippon (IJCLab), Benoit Delaunay (CCIN2P3)

- **Active Directory/Windows**

Animation : Cédric Muller (IPHC).

- **QUASAR (Qualité)**

Animation : Olivier Brand-Foissac (IJCLab)

- **CEPH**

Animation: Sébastien Geiger (IPHC)

- **Services collaboratifs : clôturé**

Animation : Jean-René Rouet (CCIN2P3)



## RI3 : Activités & événements 2022 (1)

- **École informatique : « Mettre en place et maintenir la SSI dans son unité »**
  - 5-9 septembre (La Vieille Perrotine, Centre CAES d'Oléron)
  - Porteur du projet : Guillaume Philippon
  - <https://indico.in2p3.fr/event/26597/>
  
- **Rencontre CCRI**
  - 5-6 octobre au L2IT
  - -> Techninaires
  
- **Journées informatique 2022**
  - 14-17 novembre (Domaine du Port aux Rocs, au Croisic)
  - <https://indico.in2p3.fr/event/27495/>



## RI3 : Activités & événements 2022 (2)

### ➤ Webinaires

- Trois sujets en 2022 :
  - Quantum Machine Learning (12 mai - <https://indico.in2p3.fr/event/27187/>)
  - Retour sur la conférence GTC de Nvidia (17 juin - <https://indico.in2p3.fr/event/27394/>)
  - THINK : Testing Hardware Instanciations of Neural network (28 octobre - <https://indico.in2p3.fr/event/27890/>)
- Webcast : <https://webcast.in2p3.fr/container/ri3>

### ➤ Site du RI3

- <https://ri3.in2p3.fr/>
- Finalisation du nouveau site
- Sélection et transfert des données de l'ancien site vers le nouveau



J12022

## Comité d'organisation

### **Khalil CHAWOSHI (Subatech)**

Jean-Michel Barbet (Subatech)

Merryl Flauss Decatoire (Subatech)

Tanja Pierret (Subatech)

Catherine Seznec (LP2I Bordeaux)

Patricia Warin-Charpentier (LPNHE)

## Comité de programme

### **Pierre Aubert (LAPP)**

Jean-Michel Barbet (Subatech)

Cécile Cavet (APC)

David Chamont (IJCLAB)

Jean Marc Colley (LPNHE)

Cécile Evesque (CC-IN2P3)

Édith Knoops (CPPM)

Eric Legay (IJCLAB)

Imed Magroune (Irfu)

Philippe Malbranque (APC)

Yannick Patois (IPHC)

François Talour (IJCLAB)

Bogdan Vulpescu (LPC)