



— IN2P3 —



# Vers une stratégie calcul et données de l'IN2P3

Réunion développement logiciel

**Pierre-Étienne Macchi**  
**Sabine Crépe-Renaudin**

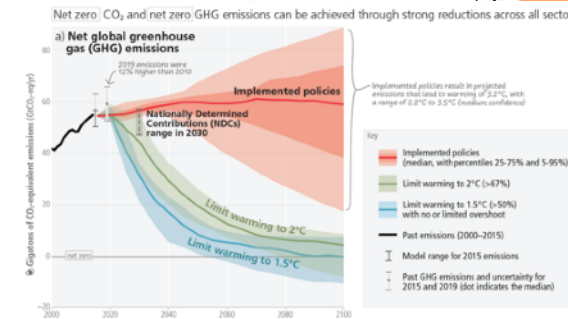
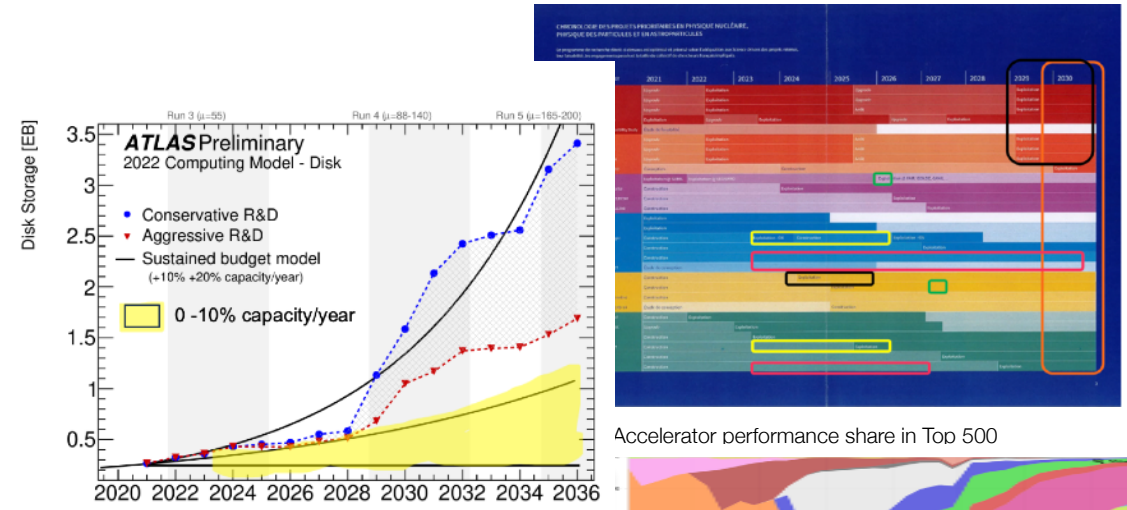
→ 30 mars 2026



# Quelques considérations

## Un moment charnière

- Des besoins considérables
  - chez nous arrivée de HL-LHC + LSST + ...
  - mais aussi dans les autres domaines scientifiques : ex SKA
  - et partout dans le monde : IA...
- Dans un environnement en forte mutation
  - coût du hardware devient volatile
  - mouvement d'une infrastructure ~homogène (CPU x86) vers une infrastructure hétérogène (FPGA, différents CPU, GPU)
  - Développement et impact fort de l'utilisation de l'IA
  - Science ouverte
- Et de plus en plus contraint
  - budget
  - RH
  - Minimisation de l'impact environnemental



➔ nécessaire de réfléchir à une stratégie globale pour les activités calcul et données à l'IN2P3

# Enjeux

- Dans un contexte en forte évolution (interne et externe) et face aux défis, il est important de préparer la réponse de l'institut en terme d'**infrastructure, compétences et organisation**
  - pour permettre le traitement et l'analyse des données de nos expériences
  - pour promouvoir nos développements et combler nos lacunes
  - pour construire ou faire évoluer les infrastructures et les compétences dont nous aurons besoin

## → Identifier, afficher et partager les priorités de l'institut

- en terme de logiciels données infrastructures...
- permettra aussi de mieux coordonner nos réponses aux appels à projet et de se préparer (Note : l'Europe s'adresse de plus en plus aux ESFRI), nos projets doivent connaître les points importants pour l'IN2P3

## → Adapter nos infrastructures

- CC-IN2P3
- plateformes IN2P3
- accès aux autres plateformes (HPC)

## → Préparer les RH définir des priorités de recrutement et de formation

- évolution des besoins en compétences, formation, carrière

## → Adapter notre organisation

- prise en compte des besoins en C&D dans les arbitrages
- meilleure coordination expériences/infra
- prise en compte de la SO et de l'IA
- coordination CC-labo-plateformes

# Vers une stratégie calcul et données

## Prise en compte

- des besoins liés à la Science Ouverte => Cf Projet DOP2I (Antoine Lemasson)
- des développements de l'IA => Cf stratégie IA (Julien Donini)
- des besoins des expériences => Cf groupe de travail C&D, présentation Éric Fede
- des besoins des services
- des évolutions des infrastructures => Réunion plateformes
- des besoins en développements logiciel => cette réunion
- de l'évolution des RH et des compétences => lien avec prospectives techniques (DAT) et GPEC (Gestion Prévisionnel Emploi et Compétence CNRS)

## Méthode

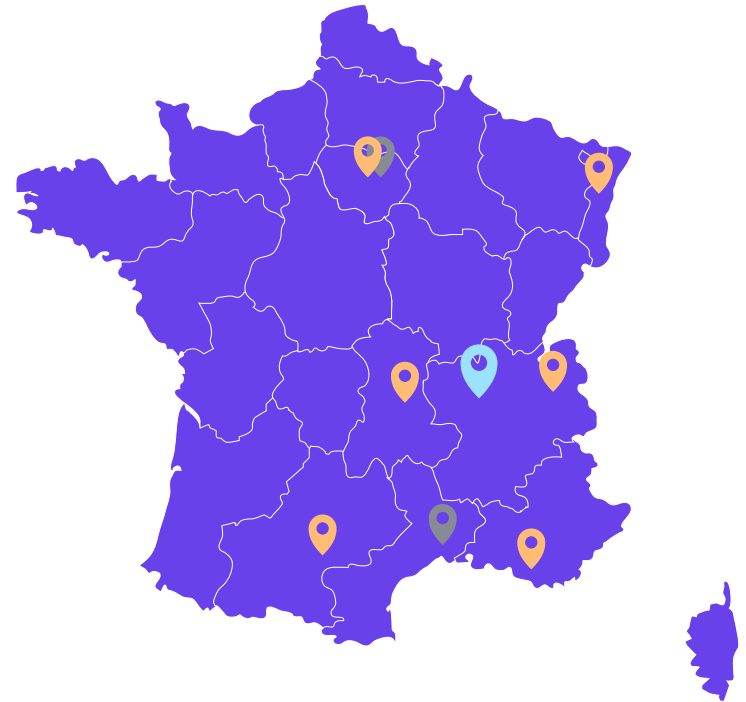
- recenser nos activités et compétences
- recenser les besoins actuels et futurs (~5 ans) de nos expériences
- définir nos objectifs et nos priorités
  - à partir de nos ressources RH et de nos infrastructures actuelles
  - en tenant compte de notre environnement
- faire évoluer nos activités, compétences et nos infrastructures en conséquence

# Infrastructures IN2P3

# IN2P3 Computing infrastructures

## Main IN2P3 data and computing centres

- **National centre:** CC-IN2P3
  - support almost all our experiments
  - our WLCG Tier1
- **Regional centres**
  - in general linked to universities
  - provide grid and cloud ressources and storage
  - 8 Tier2s (7 IN2P3 1 CEA)
  - linked to France-Grilles initiatives (-was- French NGI for EGI)
- **Local small HPC ressources**
  - mainly for developments (astro-cosmo field)



# CC-IN2P3

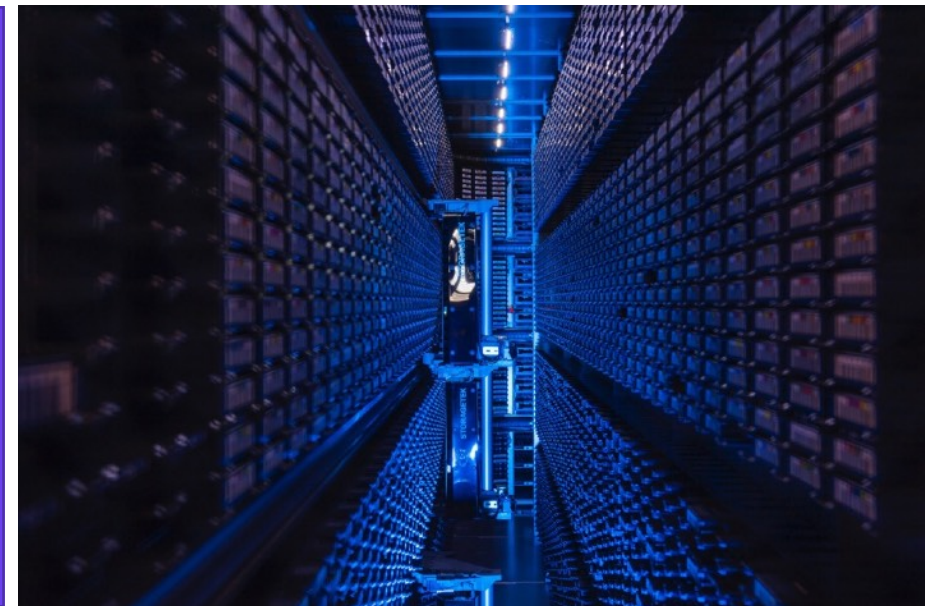
## Description and missions

- National research infrastructure for IT resources dedicated to our research fields
  - main experiments LHC/HL-LHC T1, LSST, Belle II, CTA, KM3NET, DUNE,...)
- Provides storage (disk+tape) and computing resources with the more appropriate architecture
  - Mainly HTC but increasing part of GPU and small HPC resources included
- Provides IT related services
  - IN2P3 sites connexion in relation with Renater or through access networks,
  - Tools for software developments, set of collaborative tools



### CC-IN2P3 Datacenter

- 2 computer rooms: 1700 m<sup>2</sup> up to 390 racks
- 1 064 kHS23 ~ 62 016 cores
- 95 PB disks + 210 PB tapes
  
- 80 staffs
  
- Users :
  - CPU 104 teams/930 users
  - STO 221 teams/4596 users



# CC-IN2P3 FITS

## Objectifs

- améliorations des infrastructures
  - Extension CC
  - Upgrade IDRIS
- Déploiement d'un stockage partagé IDRIS-CC et portail pour un accès plus transparent aux ressources de calcul pour les IR du CNRS

## Extension CC

- dans les temps
- salle livrée, inauguration prochaine

## Accès aux ressources

- RH IDRIS absentes pénalisent l'avancement du déploiement du stockage partagé
  - pas de soucis côté CC
- Utilisateurs : Soleil, Huma-Num, HAL
- Ouverture prochaine du site web



# Plateformes IN2P3 labellisées



DANCE : <https://dance.in2p3.fr>  
Data Analysis Center CPPM

- Calcul et Stockage HTC : grille (T2) et cloud
  - grille 5000 coeurs, 3 Po stockage
- Outils pour l'informatique distribuée : intergiciel de distribution des tâches et des données DIRAC
- Animation, support et formation



SCIGNE : <https://scigne.fr/>  
Scientific Cloud Infrastructure in Grand Est

- Calcul HTC, cloud IaaS / conteneur as a service et GPU à la demande
- Gestion et archivage de données
- 4000 coeurs, 3 Po, réseau 20Gb/s
- formations
- France Grilles, IFB, EGI et WLCG



MUST : <https://www.must-datacentre.fr/>  
Mésocentre de calcul et de stockage de l'USMB

- HTC 7500 coeurs (150TFlops), GPU 21 cartes (300Tflops), 8,5 Po, 250To SSD, réseau 40Gb/s ext 100Gb/s interne
- LHC du CERN (WLCG), CTA, LSST, HESS, ...
- accompagnement projet IDEFICS



VirtualData : <https://virtualdata.fr/>

- HTC grille site GRIF
- Cloud Virtual Data : 10000 coeurs, 500 TB
- service Spark et JupyterHub => FINK, enseignement
- service multimessagerie



# LCG-France

## SITES LCG-FRANCE



- Tier 1 CC-IN2P3
- T2 fédéré : GRIF
- Tier 2 au CPPM, IPHC, LAPP
  - test-bed fédéré : ALPAMED
  - => fermeture du T2 LPC annoncé
- Tier 3 à l'IP2I



RENATER (NREN)



# LCG-France

## Nouvelles LCG-France

- Arrêt du T2 de Clermont
- Diminution de moitié du nombre de T2 par rapport au démarrage du LHC
  - de 8 à 4 T2s (ou si on sépare les sites GRIF de 11 à 7)
- Principales raisons
  - RH : pyramide des âges et arbitrage des laboratoires ds un cadre RH tendu
  - Diminution des opportunités de financement local
  - Faible soutien des physiciens
- Impact à court terme sur les manip = diminution des ressources
  - en particulier ALICE



- impact à moyen terme sur les labos
  - perte de ressources locales pour l'analyse et la R&D
  - perte progressive de compétences techniques ou basculement vers des expertises plus tournée vers les logiciels ou le support type devops/ML/datascience
  - lien encore plus distendu entre les physiciens et ces activités

# HPC

## Vers plus d'utilisation du HPC ?

- Investissement massif par les états ds les ressources HPC
- Diversification des besoins et du hardware
- rapprochement des infrastructures HPC stratégique, complémentaire des ressources HTC
- utilisation actuelle : simulation (SMILEI, astro-cosmo, physique nucléaire, QCD sur réseau, IA)

## Discussion avec NUMPEX

- réunion "Stratégie pour l'interopérabilité des infrastructures numériques scientifiques » le 30/12 ds le cadre du PEPR NUMPEX
- note aucune activités IN2P3 retenues lors de la mise en place de NUMPEX
- Accord pour mettre en place des POC au TGCC
  - CMS (Adriano Di Florio, CC-IN2P3), ATLAS (L. Duflot, IJCLab), CC-IN2P3 (D Bouvet, Vanessa Hamar)

## Proposition SHAKE : Center of Excellence EuroHPC pour la physique des particules

- Enjeux : être une communauté identifiée par EuroHPC
- projet européen, financement 50% EuroHPC 50% national (budget dédié au ministère en France) : 4-8M€, dépôt mi-janvier
- financement pour porter les codes sur HPC => structuration WP1, WP2 code, WP3 interface avec HPC, WP4 codesign architecture
- IN2P3 :
  - WP3 : sous tâche coordonnée par A. Di Florio test des solutions sur les centres HPC
    - collaboration NUMPEX GENCI CEA+ IN2P3 (POC TGCC),
  - WP2 coordination Sabine CMS MLFL (A. Di Florio, F. Beaudette LLR), LHCb-ATLAS (coordination V. Gligorov, D. Rousseau... LPNHE-IJCLab), participation LQCD (B. Blossier IJCLab)

# Infrastructures de recherche nationales

## Maintien du CC sur la FdR

- publication prévue fin 2026

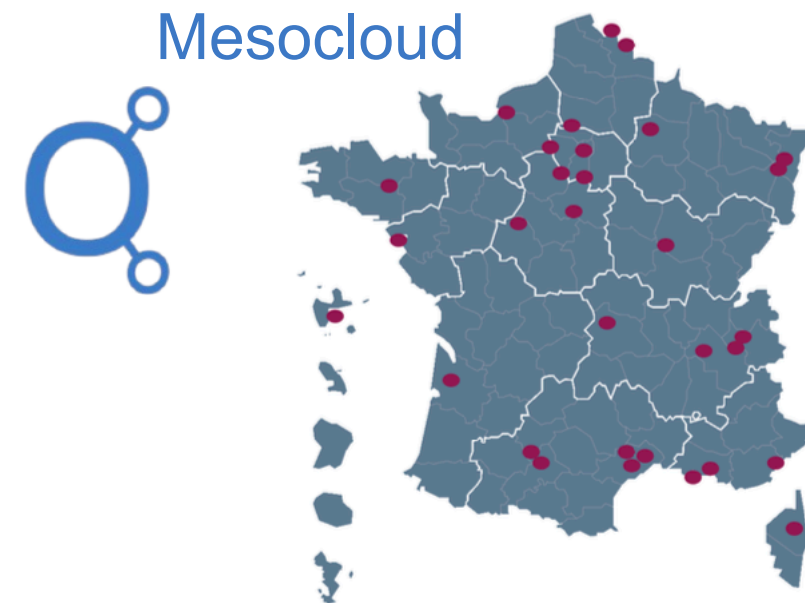
## Fin de France-Grilles

## Projet CRD (fond d'amorçage DGRI)

- Cloud Recherche & Données => portail de services basé sur prestation pour garantir la qualité des services
- peu d'intérêts à l'IN2P3 côté infrastructures

## Candidature Mesocloud

- dossier déposé + audition ministère => retour positif
- une vingtaine de sites se sont déclarés intéressés (IN2P3 : IJCLab, LPNHE, APC, IPHC, CPPM et LPCA via leur université, MUST en discussion)
- travail en cours sur
  - l'accord de consortium et la mise en place de l'UAR
  - l'intégration de EGI
  - mise en place des services communs
- Plus d'info :
  - [JCAD 2025](#) (mesonet et mesocloud présentation)
  - Dossier [lien box](#)



- Fédération des mésocentres, complémentarité HTC/HPC
- Culture commune : support, partage de compétences et proximité
- Pérennisation des services et compétences de France-Grilles et MesoNET
- Mutualisation des ressources
- Accès simplifié pour les utilisateurs, taille critique pour financement nationaux et européens

# Nouvelles HPC/IA/QC

## Nouvelles machines quantiques au TGCC

- Ruby : calculateur analogique à atomes neutres, capable de contrôler 100 atomes de Rubidium, conçu par la société Pasqal
- Lucy : ordinateur quantique photonique (QPU photonique à 12 photons) couplé au supercalculateur Joliot-Curie



## Arrivée prochaine de Alice Recoque

- GPU AMD Instinct™ MI430X, processeurs AMD EPYC Venice, 432 GB de mémoire, bande passante mémoire de 19,6 To/s
- partition CPU processeurs SiPEARL Rhea2
- réseau haute performance Eviden BX1 v3 : débits de 800 Gbit/s entre GPU et 400 Gbit/s entre CPU
- 550M€ sur 5 ans



## AI factory France (AI2F)

- plateforme européenne pour la science, l'industrie et la société
- 30,7 millions d'euros sur 3 ans

# Activités logiciels

# Projets et activités logiciels

## Projets dédiés à un logiciel

- Dirac
  - gestion des données et tâches de calcul sur une infrastructure distribuée
- Geant4 et Geant4-DNA
  - MC interactions des particules dans la matière, modélisation des dosages biologiques

## Projets transverses

- Decalog : activités transverses autour des logiciels pour améliorer les performances des logiciels depuis le développement au déploiement
  - Reprises : performances des logiciels calcul intensif et hétérogène
  - ComputeOps : nouvelles technologies d'administration (virtualisation, conteneurs)
- ML (Compstat) coordination et animation autour de des activités IA

## Projets exploratoires/R&D

- QC2I Quantum Computing
  - veille technologique, formation, développements Quantum Machine Learning
- THINK
  - IA embarqué

## Activités dans les projets scientifiques et techniques

- Dernier recensement : CS IN2P3 2022
  - SMILEI
  - NPTools
  - ACTS
  - AGATA software
  - KaliVeda Heavy-Ion Analysis Toolkit
  - Gammapy
  - AMI
  - GATE
  - RTA LHCb
  - ...
- Développements spécifiques pour les expériences
- Online

The background of the image consists of numerous thin, wavy, light blue lines that create a sense of motion and depth. These lines are set against a dark blue background and curve across the entire frame.

**RH**

# État des lieux RH

## BAP E titulaires ou stagiaires (janvier 2026)

- l'effectif total est de 251 personnes
- l'âge moyen est de **49,7** ans
- 59,4 % de l'effectif ont plus de 50 ans, 39% plus de 55 et 17% plus de 60 ans
- déficit de jeunes sauf en calcul scientifique

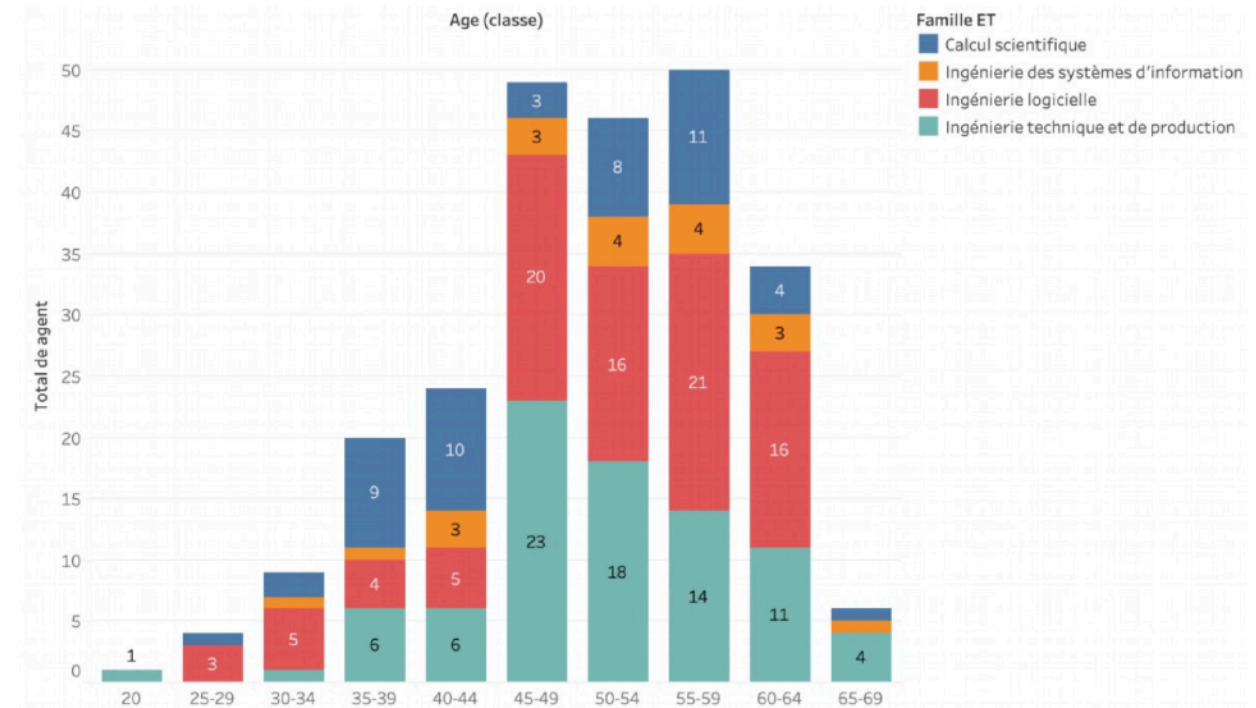
## Répartition des ETP (données NSIP) total 216 ETP

|                               | ETPT  | %    |
|-------------------------------|-------|------|
| CC-IN2P3                      | 34,1  | 15,8 |
| Plateformes hors CC           | 7,3   | 3,4  |
| EOSC+DOP2I + FGI + FITS + LCG | 14,0  | 6,5  |
| Master projets C&D            | 8,1   | 3,7  |
| Masters Projets non C&D       | 77,8  | 36,0 |
| Support- informatique         | 54,6  | 25,3 |
| Management                    | 14,0  | 6,5  |
| Autres                        | 6,2   | 2,9  |
| Total                         | 216,1 |      |

Plateforme : 26%

Projets : 40%

Support : 32%



Décembre 2025

Pierre-Étienne Macchi

# Répartition thématique

Répartition des ETP  
(données NSIP 2025)  
total 216 ETP

|                               | ETPT  | %    |
|-------------------------------|-------|------|
| CC-IN2P3                      | 34,1  | 15,8 |
| Plateformes hors CC           | 7,3   | 3,4  |
| EOSC+DOP2I + FGI + FITS + LCG | 14,0  | 6,5  |
| Master projets C&D            | 8,1   | 3,7  |
| Masters Projets non C&D       | 77,8  | 36,0 |
| Support- informatique         | 54,6  | 25,3 |
| Management                    | 14,0  | 6,5  |
| Autres                        | 6,2   | 2,9  |
| Total                         | 216,1 |      |

Plateforme : 26%

Projets : 40%

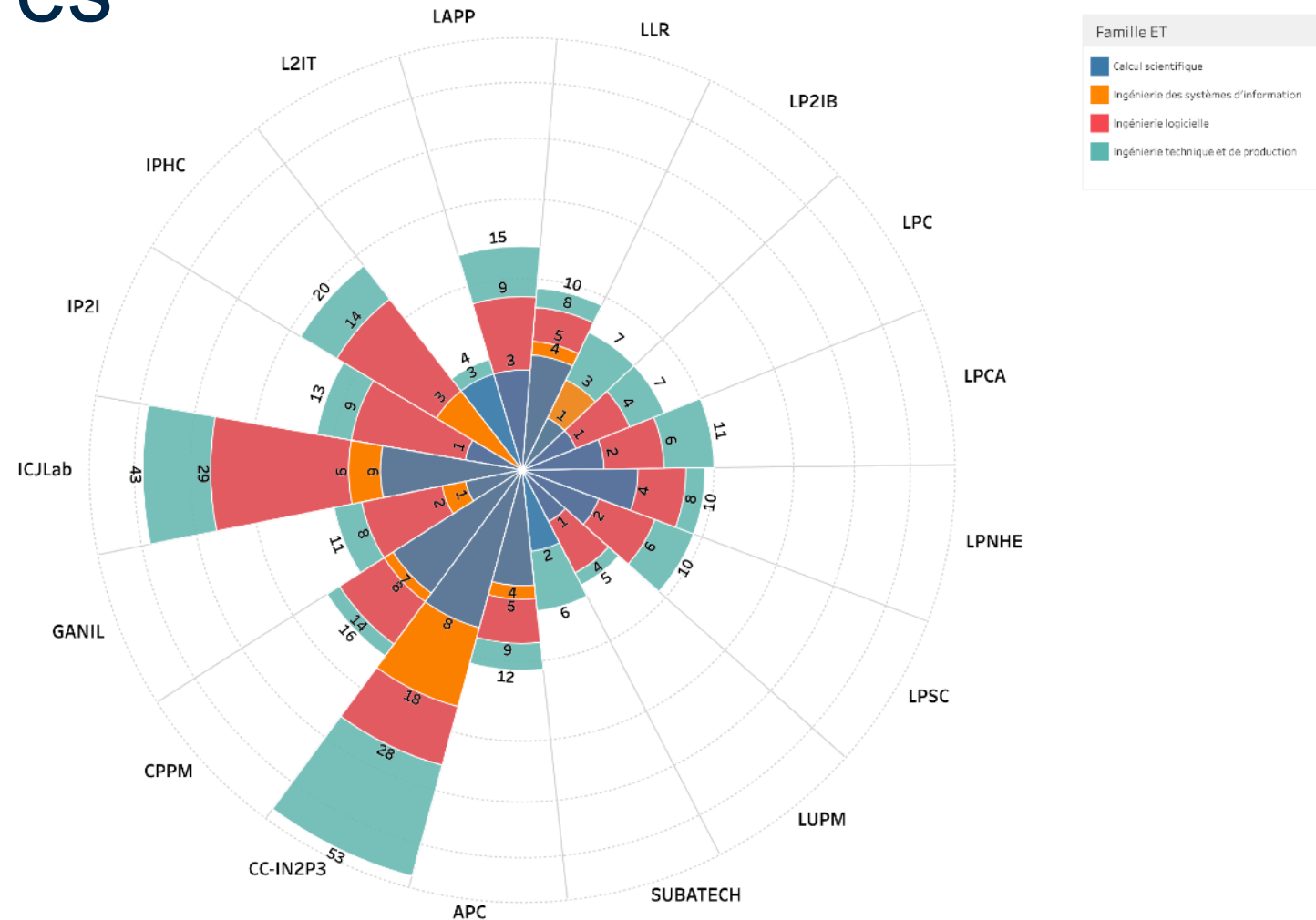
Support : 32%

| Domaine                       | Calcul scientifique | Ingénierie des systèmes d'Information | Ingénierie logicielle | Ingénierie technique et de production | TOTAL | %    |
|-------------------------------|---------------------|---------------------------------------|-----------------------|---------------------------------------|-------|------|
| Astroparticules et Cosmologie | 21,1                | 0,1                                   | 15,4                  | 2,7                                   | 39,2  | 46 % |
| Physique des Particules       | 5,9                 | 0,5                                   | 11,6                  | 0,7                                   | 18,6  | 22 % |
| Accélérateurs et Technologies | 1,1                 | 0,4                                   | 5,6                   | 0,3                                   | 7,4   | 9 %  |
| Nucléaire et Hadronique       | 0,7                 | 0,1                                   | 2,7                   | 0,9                                   | 4,4   | 5 %  |
| Nucléaire pour la société     | 0,2                 | 0,0                                   | 0,0                   | 0,1                                   | 0,3   | 0 %  |
| Calcul et Données             | 4,9                 | 0,2                                   | 1,3                   | 1,7                                   | 8,1   | 9 %  |
| Plateformes IN2P3             | 0,1                 | 0,1                                   | 7,1                   | 0,5                                   | 7,7   | 9 %  |
| Projets Laboratoire           | 0,0                 | 0,1                                   | 0,0                   | 0,1                                   | 0,1   | 0 %  |
| Total général                 | 33,9                | 1,4                                   | 43,8                  | 6,8                                   | 85,9  |      |

# Répartition Laboratoires

## Distribution selon les laboratoires

- des disparités selon les laboratoires
  - nombreux départs CPPM, IJCLab un peu plus tard IPHC
  - ingénierie infrastructure absente ds certains laboratoire



**Construire une stratégie**

# Stratégie calcul et données : science ouverte

Antoine Lemasson

## Projet DOP2I

- recensement des activités effectuées
- Discussions
  - ds DOP2I
  - ouverte à l'institut : [Journées Des Données Ouvertes de la Physique des 2 Infinis](#)

## → Recommandations en cours d'écriture

- 4 parties
  - Données Ouvrables (Qualité de la donnée/Cycle de vie/Pérennisation)
  - Logiciels / Analyses Reproductibles
  - Infrastructures et Services
  - Structuration et Accompagnement

## Membres DOP2I

- Antoine Lemasson
- 22 Participants :
  - APC Eric Chassande-Mottin, Bruno Khelifi (\*Astro/Cosmo)
  - CC-IN2P3 Frederic Azevedo, Yonni Cardenas, Sébastien Gadrat (RI3), Pierre-Etienne Macchi, Gino Marchetti, Jean-Yves Nief
  - CPPM Cristinel Diaconu (\*HEP)
  - GANIL Antoine Lemasson
  - IJCLAB Piera Ghia, Olivier Brand-Foissac
  - IP2I Michel Farizon, Olivier Stézowski
  - IN2P3 Mathieu Grives, Sabine Crépé-Renaudin
  - LAPP Ian Bird, Thomas Vuillaume (\* Software)
  - LLR Marc Verderi
  - LPC Caen Daniel Cussol, Adrien Matta (\* NP) + Phil Davies
  - LPC Clermont Louie Corpe
  - LPNHE Olivier Dadoun
  - LPSC Sabine Kraml (\* Theorie)

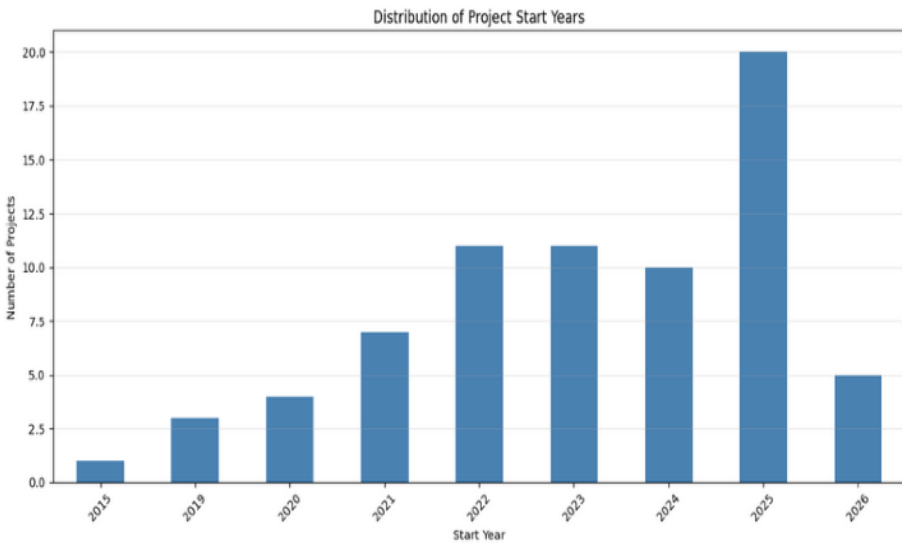
# Stratégie Calcul et Données : IA

Julien Donini

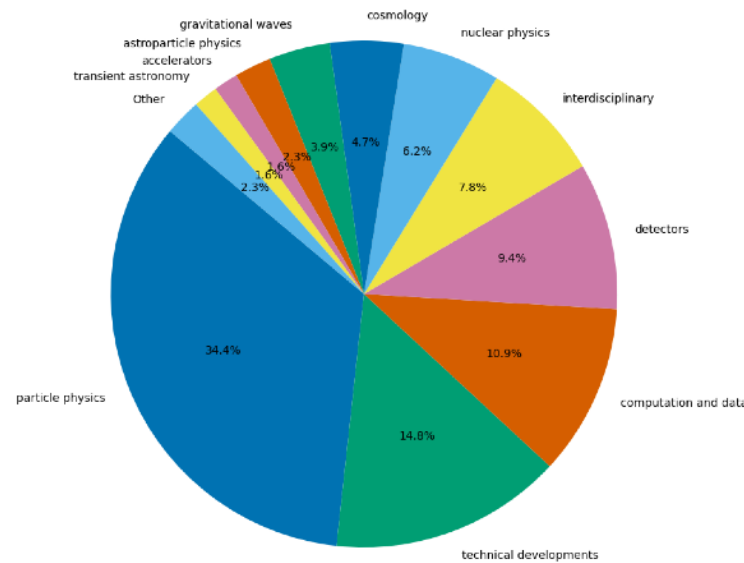
## État des lieux

→ Sondage (via aussi projet ML)

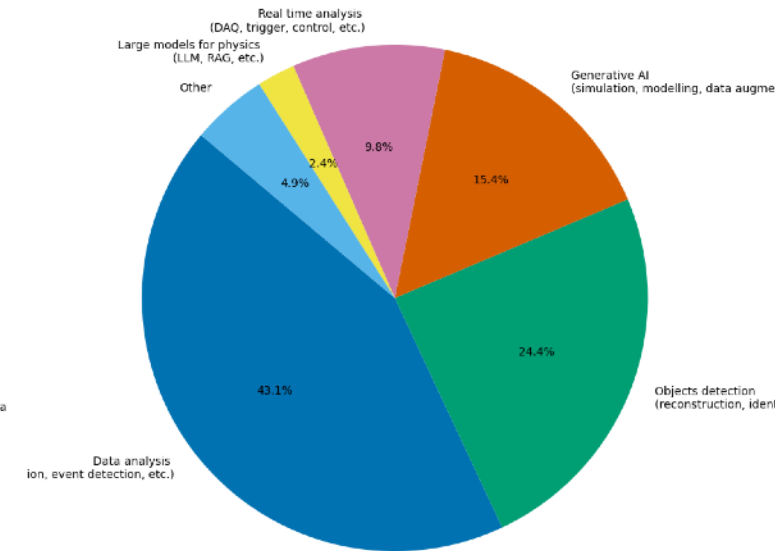
Nombre de projets ML démarrés ces dernières années  
(enquête récente menée à l'IN2P3)



Projects: Physics domains



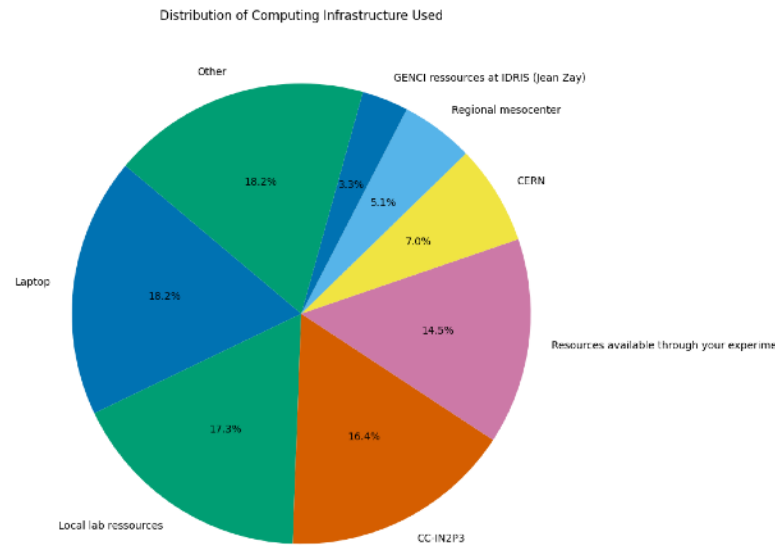
Projects: AI application domains



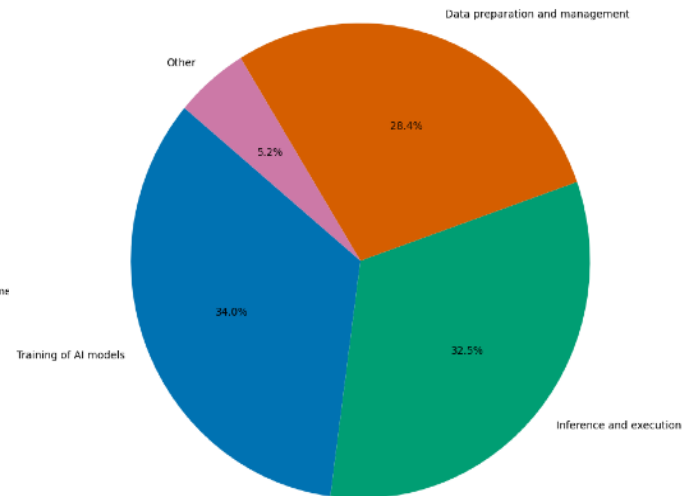
## Infrastructures utilisées

→ Sondage (via aussi projet ML)

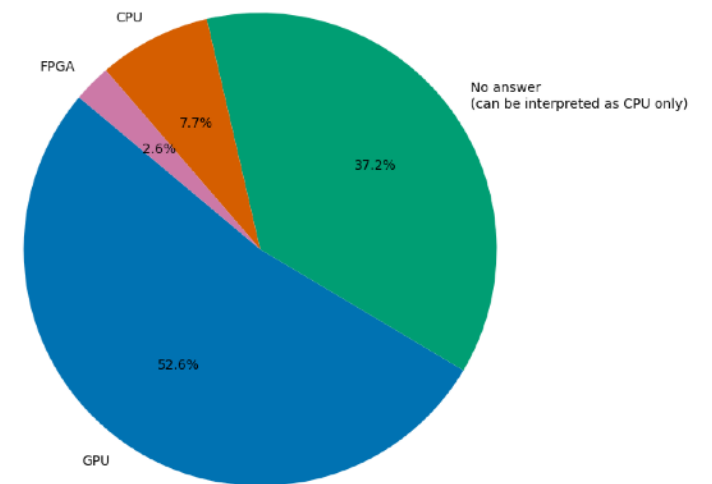
Computing: Infrastructure used



Computing: Usage of resources



Type de processeurs



## Réflexion sur le rôle de l'IA dans la recherche à l'Institut

- Quelle articulation en recherche fondamentale et **ingénierie** ?
- Comment accompagner la **passage à l'échelle** de projets ?
- Quelle structuration et **gouvernance** à l'Institut ?
- Nécessité d'une stratégie croisée **calcul et IA**

## Premiers éléments

- Panorama et état des lieux : compétences, projets, infrastructures, besoins formation
- Soutien des projets de recherche menés à l'IN2P3 par des ingénieurs avec une expertise en IA et aux compétences adaptées
- Propositions de projets pilotes et identification de cas d'usage pertinents (AISSAI et COMP CNRS)
- Données, benchmarks et science ouverte: Accès, mutualisation, catalogage des datasets ouverts "ML ready"
- Équipes IA : valoriser les équipes transverses déjà présentes dans les labos, avec possibilité d'un maillage au niveau IN2P3 (GDR ?)
- Formation en IA et recrutement des jeunes chercheurs

→ premier document collaboratif remis à la direction, déclinaison à venir

# Stratégie Calcul et Données : traitement des données

## Infrastructures, logiciels, du online à l'analyse

### Quelles priorités pour les outils à déployer et développer ?

- Partir des besoins des expériences
  - liste des outils utilisés développés aujourd'hui et des besoins pour les prochaines années
  - pour toute nos disciplines
- Mettre en face les compétences et ressources ds les services informatiques
- Groupe de travail qui couvre les différentes thématiques scientifiques et techniques mis en place

### Diffusion et Discussion

- Une fois l'état des lieux effectué, partager et discuter avec les projets, les services, la direction

=> Présentation Éric Fede

### Membres groupe de travail

- Antoine Lemasson
- David Chamont
- Eric Fede
- Adriano Di Florio
- Johan Bregeon
- Laurent Duflot
- Julien Donini
- Patrice Verdier
- Pierre-Etienne Macchi
- Vladimir Gligorov

# Questions

# Les questions

En tenant compte des contraintes => budget/RH => priorités ?

## Quelle organisation ?

- quelle coordination ?
  - pour les activités logiciels dans et hors les expériences
    - partage d'information (bonnes pratiques, exemple)
    - priorités sur les développements, mutualisation
  - lien avec la R&D
  - création d'un GDR calcul ?
- quelle lien avec les plateformes
  - CC
  - les plateformes des labos
  - HPC
  - répartition/recensement/partage du matériel/services
  - GPU, FPGA, processeurs spécifiques
- quels lien avec les experts : où sont ils ?

## Quelles activités prioritaires ?

- Cf recensement
- autres ?

## Quels supports prioritaires ?

- plateforme de développement, Gitlab
- plateforme de R&D
- quel support pour les labos sans plateforme ?

## Formation et compétences

- des compétences manquantes ?

## Rôles des plateformes ?

- rôles possibles
  - R&D
  - production (T2+ ?)
  - analyse, développement
  - support local
  - données et SO
- CC vs plateformes régionales
  - quelle répartition des rôles
  - quid des labos sans plateformes

**Bonus**

# Rappel : canaux d'informations

## *Développements logiciels*

- projet Decalog
- Liste de diffusion : [DEV-LOGICIELS-L@IN2P3.FR](mailto:DEV-LOGICIELS-L@IN2P3.FR)
- en cours : valorisation logiciel libre Open

## *Infra et projets européens*

- [euro-computing-L@IN2P3.FR](mailto:euro-computing-L@IN2P3.FR)
- plateformes, LCG, France Grille, EGI, EOSC
- prochaine JCAD : <https://jcad2024.sciencesconf.org/>

## *HPC*

- projet FITS, exascale (=> Alice Recoque), JENA HPC WG
- Liste de diffusion : [HPC-INFO-L@IN2P3.FR](mailto:HPC-INFO-L@IN2P3.FR)

## *RI3 <https://ri3.in2p3.fr>*

- JI
- Webinaires et écoles
- Webcasts : <https://webcast.in2p3.fr/container/ri3>

## *ML/IA Cf Julien Donini*

- Projet ML : <https://machine-learning.pages.in2p3.fr/>
- Liste de diffusion : [MACHINE-LEARNING-L@IN2P3.FR](mailto:MACHINE-LEARNING-L@IN2P3.FR)
- Workshop ML IN2P3 (+IRFU): <https://indico.in2p3.fr/event/36597/>
- Centre AISSAI CNRS :
  - centre dédié à l'IA : <https://www.cnrs.fr/fr/le-centre-artificial-intelligence-science-science-artificial-intelligence-aissai>
  - semestre thématique pour l'IN2P3, appel pour accompagnement d'ingénieur

## *Informatique Quantique*

- Projet QC2I : <https://qc.pages.in2p3.fr/web/>

## *Publication*

- Contacts : [IST@IN2P3.fr](mailto:IST@IN2P3.fr) ou [Mathieu Grivès](#)