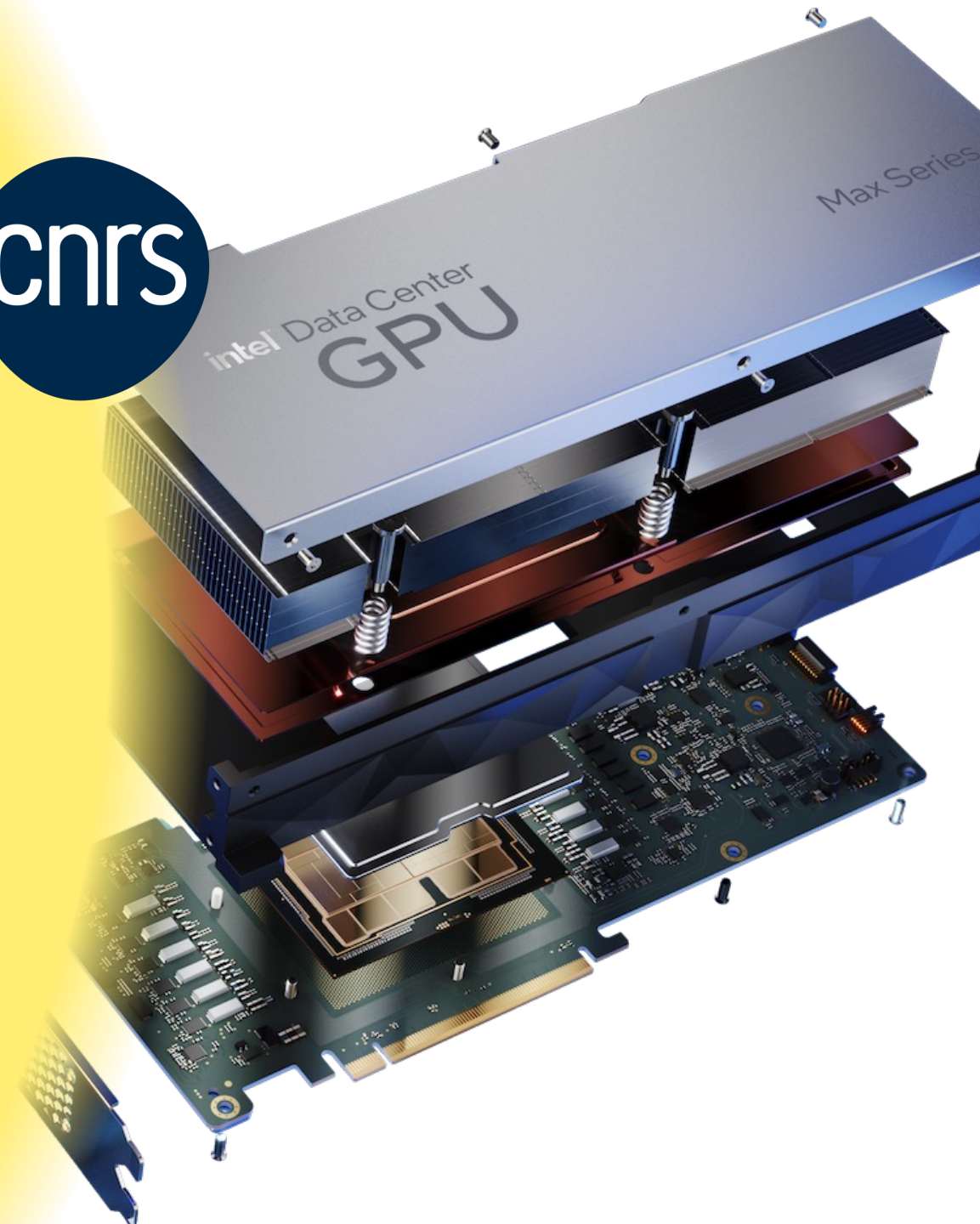


Atelier : Moyens de Calculs dans les Laboratoires

Matteo Cacciari (DS Calcul et données scientifiques -
Programmes européens)

Benoit Devincré (Physique de la matière condensée,
systèmes complexes, approches multi-échelles)

→ 21/06/2024



Sommaire

Enquête sur les usages et outils numériques de laboratoires de CNRS Physique

Besoins numériques au regard des spécificités de la recherche en physique

Organisation et exploitation des solutions numériques

01

Enquête sur les usages et outils numériques des laboratoires de CNRS Physique

Enquête diffusée via les DU

<https://framaforms.org/enquete-sur-les-outils-numeriques-de-cnrs-physique-1695287367>

<https://framaforms.org/survey-on-computational-tools-of-cnrs-physics-1707403489>

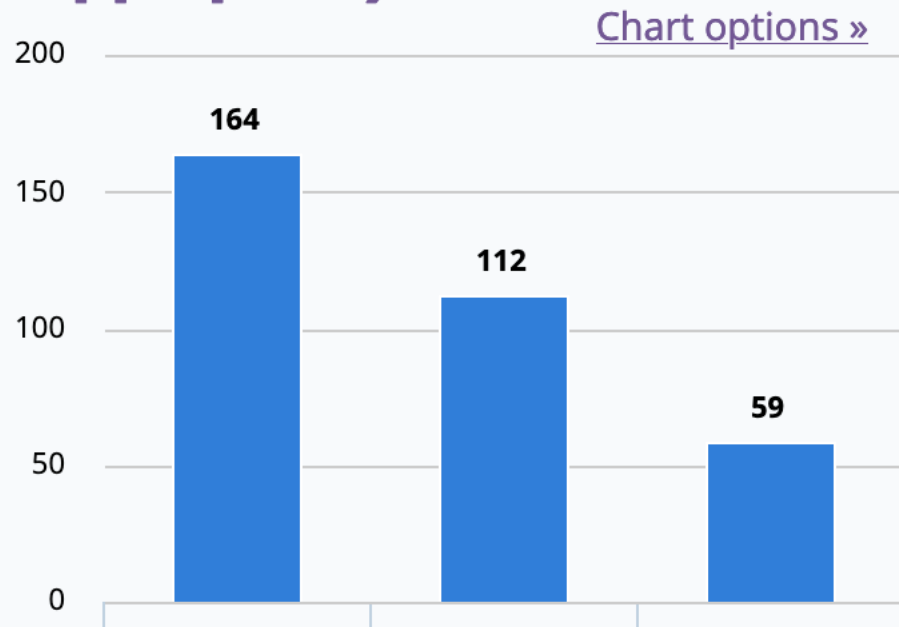
“Cette enquête vous concerne si vos recherches ont un volet numérique et/ou si vous avez, ou pensez pouvoir avoir, des besoins de stockage permanent de données”

- **Recenser les usages**
- **Comprendre les difficultés**
- **Identifier les besoins**

Enquête

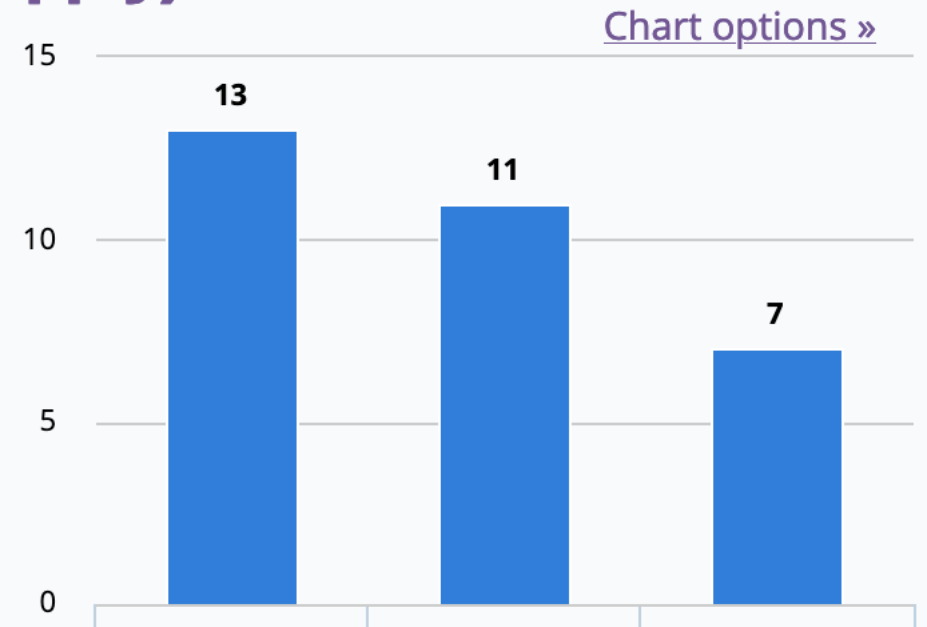
- **Trois parties : calcul, données, intelligence artificielle**
- **208 réponses (192 français, 16 anglais) provenant de 50 unités**
- **157 C/EC, 12 ingénieurs, 29 étudiants/postdocs**

Les codes que vous développez sont utilisés par (cochez toutes les cases qui s'appliquent) :



vous même	164
votre groupe/laboratoire	112
une communauté à l'échelle nationale ou internationale	59

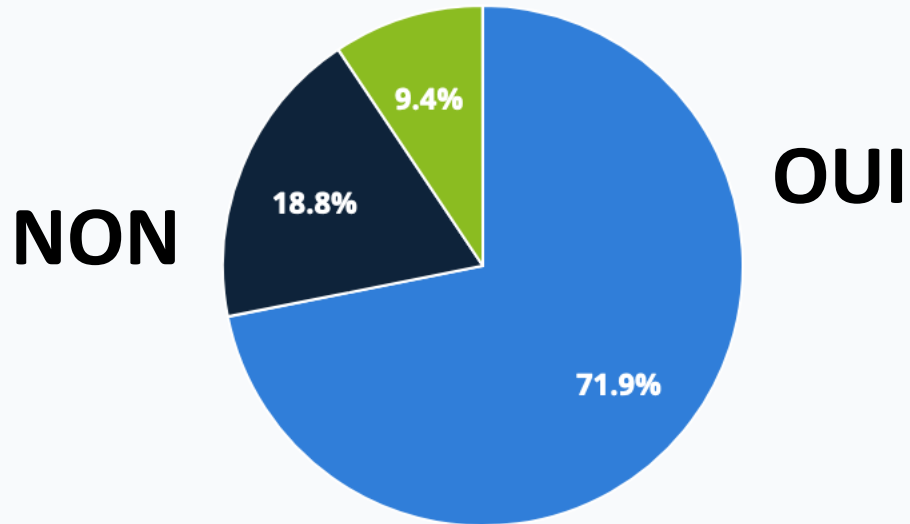
The codes that you develop are used by (select all cases that apply):



yourself	13
your team/laboratory	11
a community at the national or international level	7

Votre laboratoire/votre équipe possède-t-il des ressources de calcul partagées (un cluster) ?

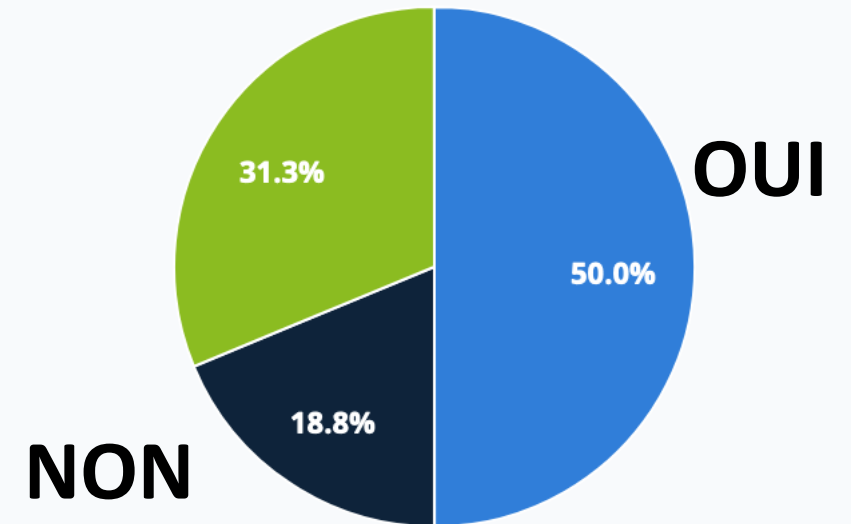
[Chart options »](#)



oui	138
non	36
je ne sais pas	18

Does your laboratory/team have shared computing resources (a cluster)?

[Chart options »](#)

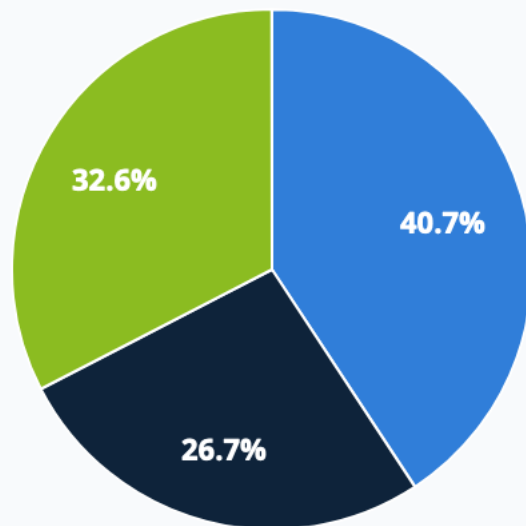


yes	8
no	3
I don't know	5

Utilisez vous ces ressources ?

[Chart options »](#)

NON

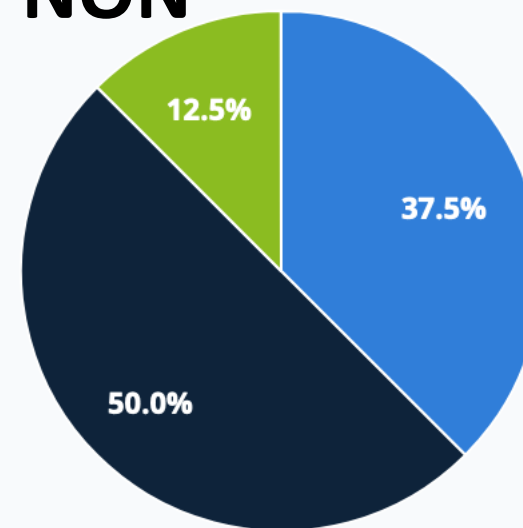


oui, régulièrement	55
oui, de temps en temps	36
non	44

Do you use these shared resources?

[Chart options »](#)

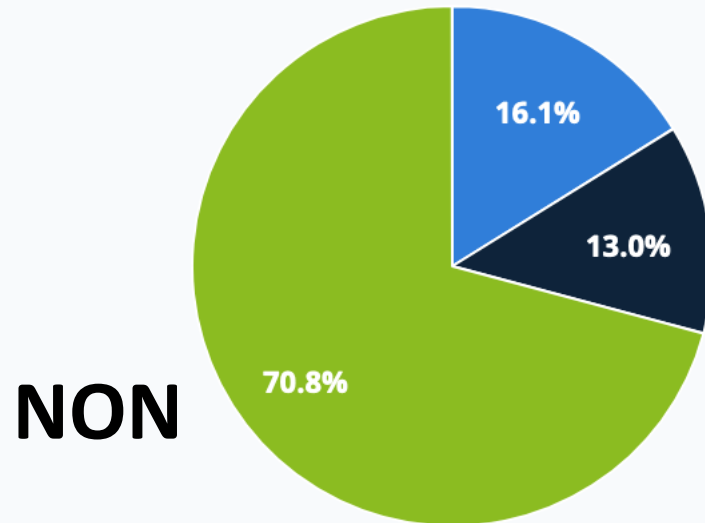
NON



yes, regularly	3
yes, occasionally	4
no	1

Utilisez vous un mésocentre de calcul de site (https://calcul.math.cnrs.fr/pages/mesocentres_en_france.html) ?

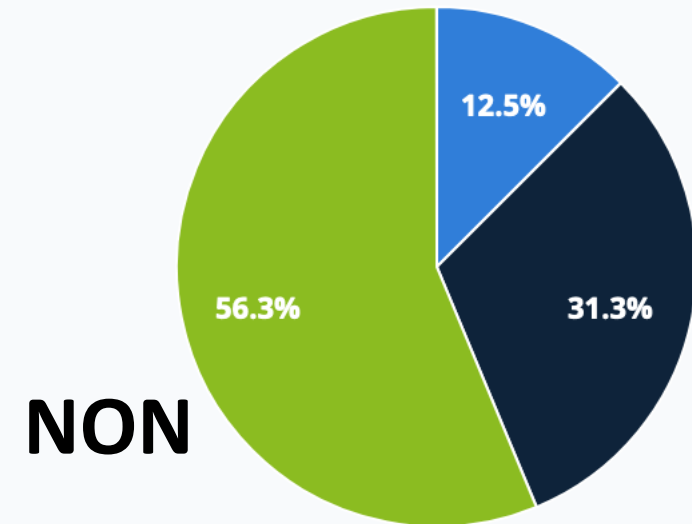
[Chart options »](#)



oui, régulièrement	31
oui, de temps en temps	25
non	136

Do you use a computing mesocentre (https://calcul.math.cnrs.fr/pages/mesocentres_en_france.html) ?

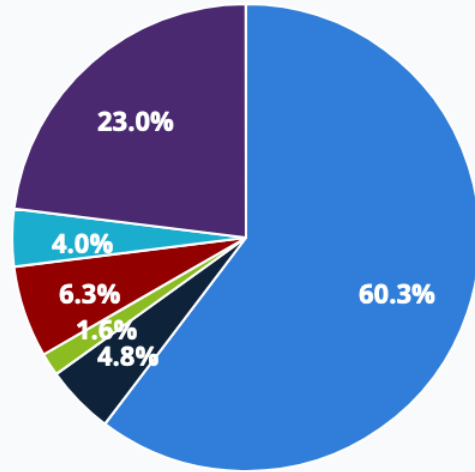
[Chart options »](#)



yes, regularly	2
yes, occasionally	5
no	9

Pourquoi vous ne l'utilisez pas ?

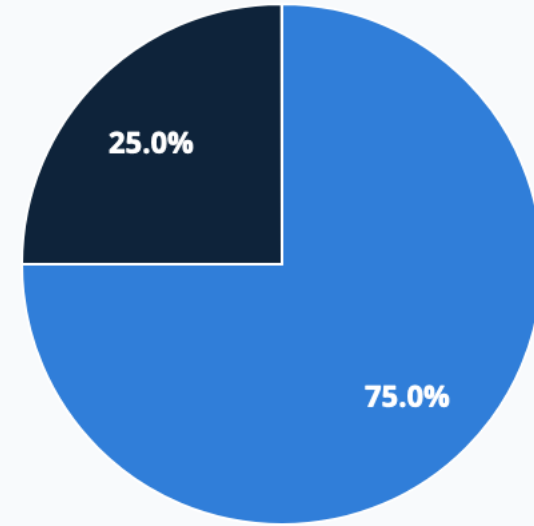
[Chart options »](#)



Il n'est pas nécessaire pour mes besoins de calcul	76
Il est insuffisant pour mes besoins, je calcule ailleurs	6
Pas disponible sur mon site	2
Je n'ai pas accès	8
Accès administrativement trop compliqué	5
Autre raison	29

Why don't you use it?

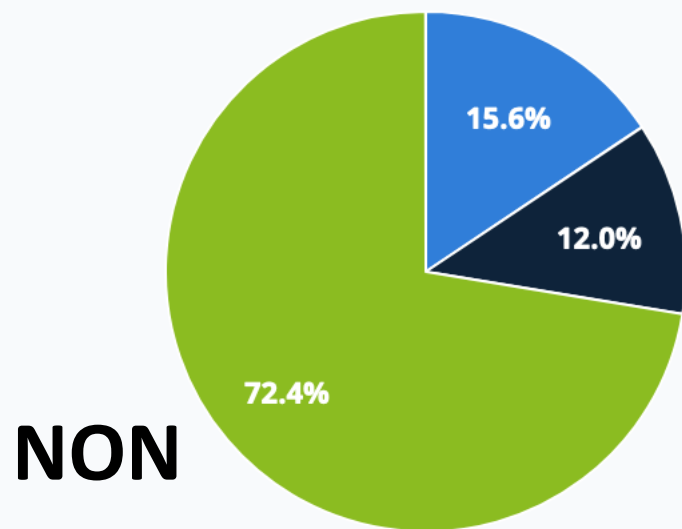
[Chart options »](#)



It is not necessary for my computing needs	6
I do not have access	2

Utilisez vous les grandes infrastructures de calcul intensif nationales (<https://www.science-ouverte.cnrs.fr/calcul-intensif/>) ?

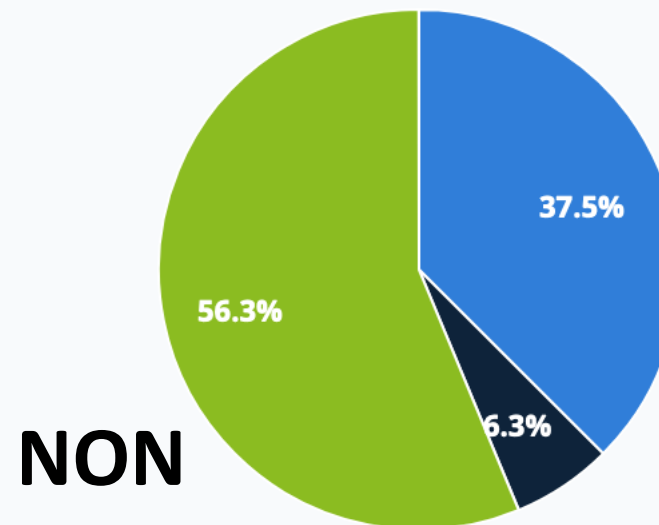
[Chart options »](#)



oui, régulièrement	30
oui, de temps en temps	23
non	139

Do you use the national HPC infrastructures (<https://www.science-ouverte.cnrs.fr/calcul-intensif/>)?

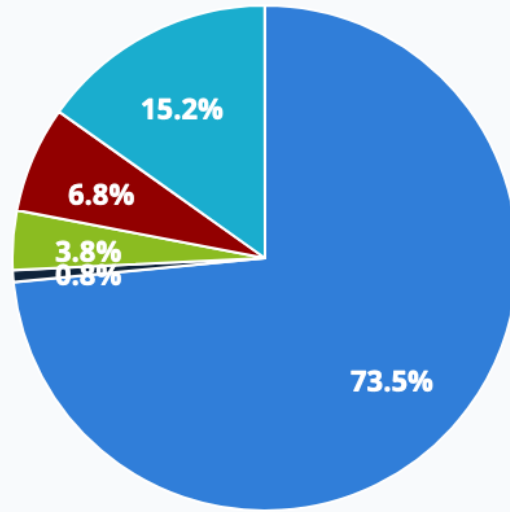
[Chart options »](#)



yes, regularly	6
yes, occasionally	1
no	9

Pourquoi vous ne les utilisez pas ?

[Chart options »](#)



Elles ne sont pas nécessaires pour mes besoins de calcul 97

Elles sont insuffisantes pour me besoins, je calcule ailleurs 1

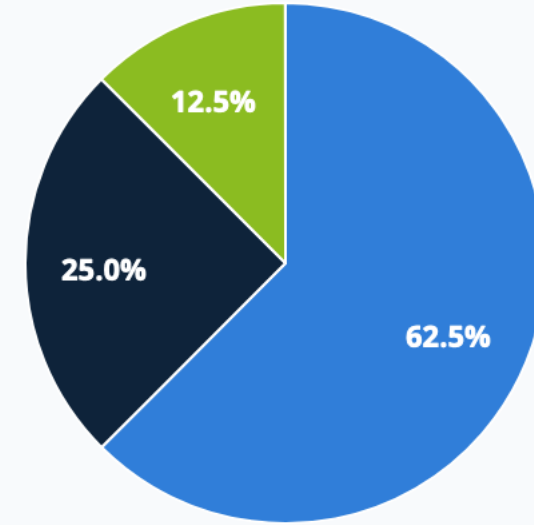
Je n'ai pas accès 5

Accès administrativement trop compliqué 9

Autre raison 20

Why don't you use them?

[Chart options »](#)



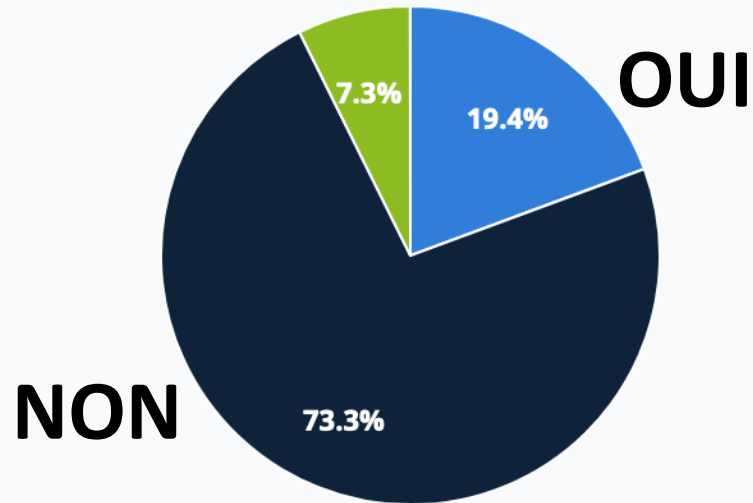
They are not necessary for my computing needs 5

I don't have access 2

Administratively too complicated to access 1

Avez vous accès à des ressources humaines pour l'aide au développement de codes (ingénieurs, personnels de maisons de la simulation, etc.) ?

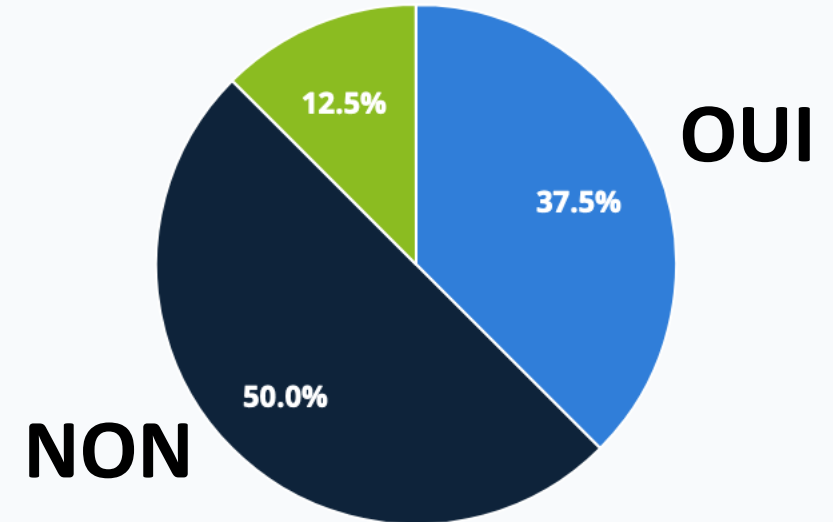
[Chart options »](#)



oui	37
non	140
je ne sais pas	14

Do you have access to human resources to help with code development (engineers, simulation centre staff, etc.)?

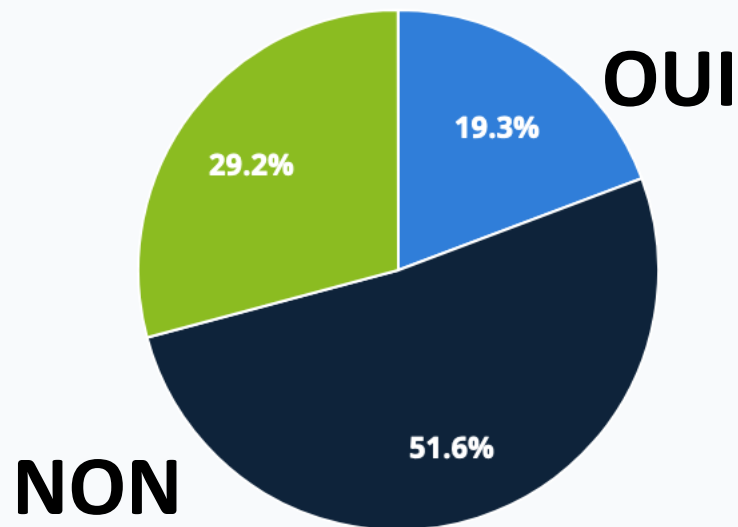
[Chart options »](#)



yes	6
no	8
I don't know	2

Votre thématique utilise-t-elle un (ou plusieurs) entrepôt de référence de données référencables (= avec DOI) ?

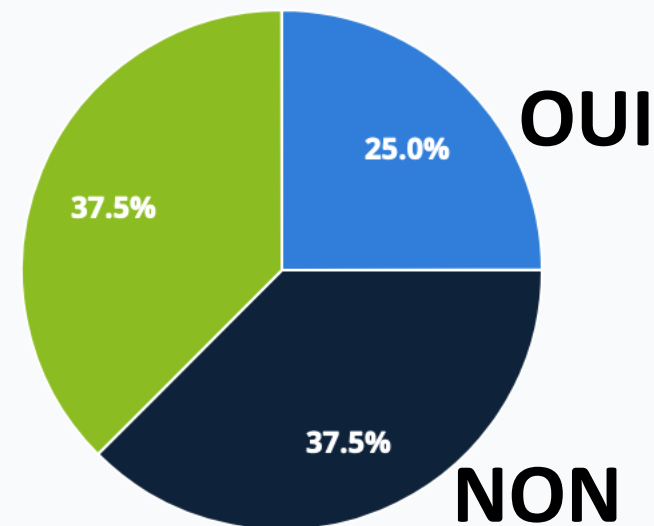
[Chart options »](#)



oui	37
non	99
je ne sais pas	56

Does your discipline use one (or more) reference repositories of citable data (= with DOI)?

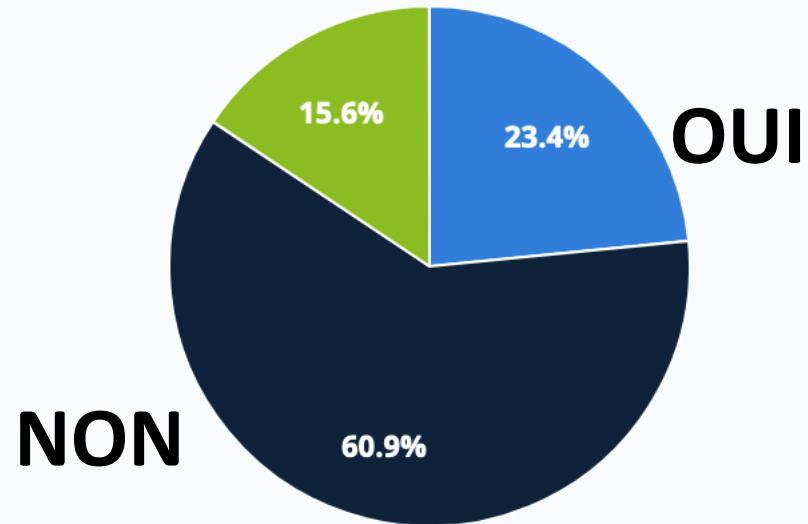
[Chart options »](#)



yes	4
no	6
I don't know	6

Avec vous, personnellement, des nécessités de stockage permanent de données référencables (= avec DOI) ?

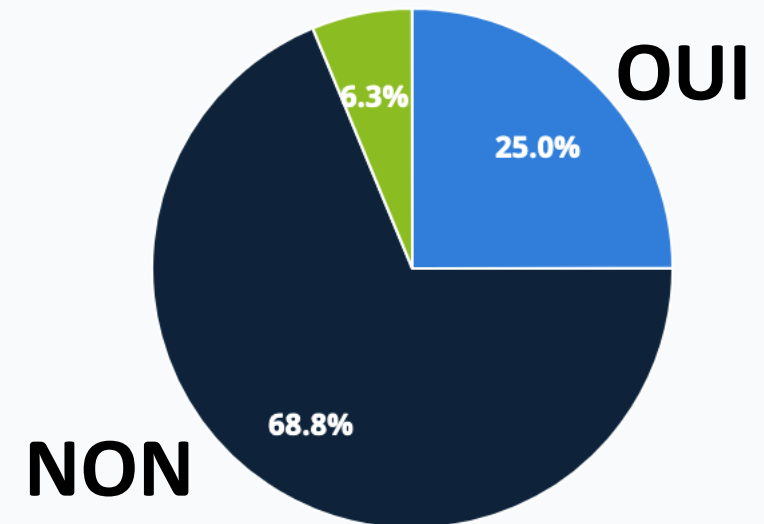
[Chart options »](#)



oui	45
non	117
je ne sais pas	30

Do you personally need to permanently store data that can be cited (= with DOI)?

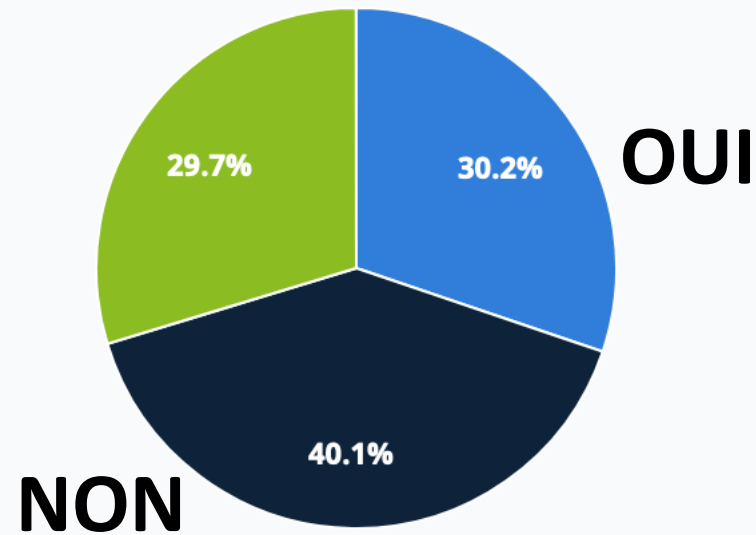
[Chart options »](#)



yes	4
no	11
I don't know	1

Vos activités de recherche ont-elles un volet Intelligence Artificielle/Machine Learning/Deep Neural Networks ?

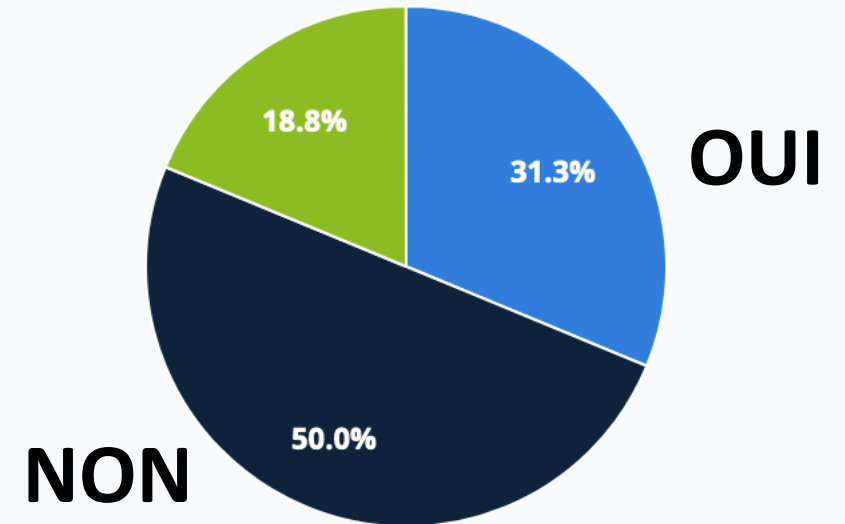
[Chart options »](#)



oui	58
non	77
pas encore	57

Do any of your research activities involve Artificial Intelligence/Machine Learning/Deep Neural Networks?

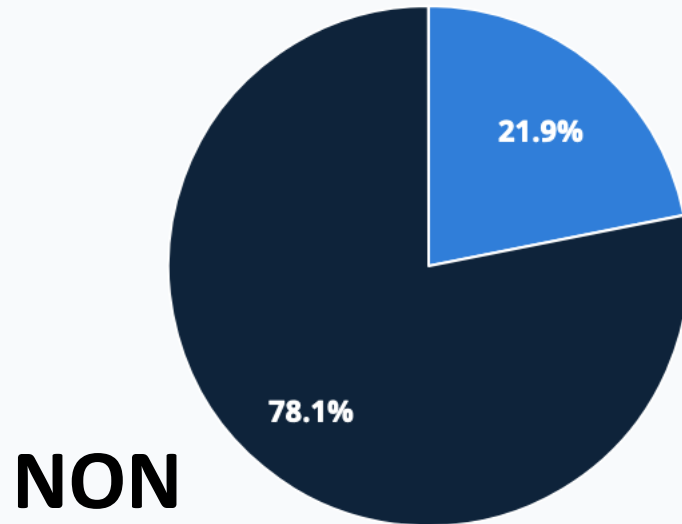
[Chart options »](#)



yes	5
no	8
not yet	3

Connaissez vous le portail Recherche Data Gouv (<https://recherche.data.gouv.fr/fr>) ?

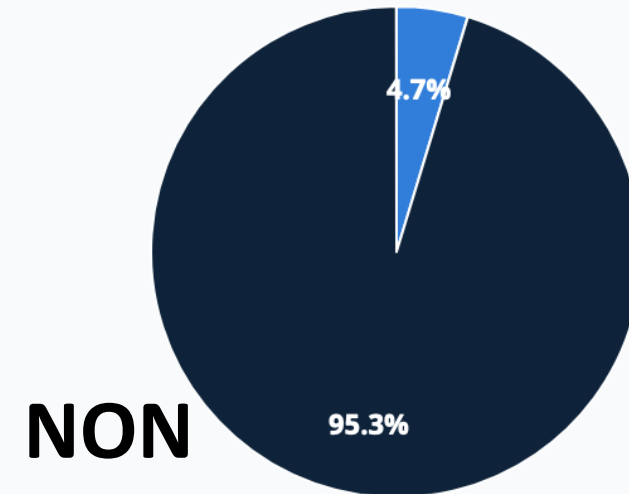
[Chart options »](#)



oui	42
non	150

Connaissez vous AISSAI, l'initiative du CNRS en soutien de la recherche autour de l'Intelligence Artificielle (<https://www.cnrs.fr/fr/le-centre-artificial-intelligence-science-science-artificial-intelligence-aissai>) ?

[Chart options »](#)



oui	9
non	183

Quelques considérations préliminaires

- Pas de besoins criants exprimés en termes de puissance de calcul
- Quelques difficultés exprimées pour l'accès aux grands centres de calcul
- Quelques demandes de support RH en développement
- FAIRisation des données encore méconnue et peu pratiquée
- Fraction non négligeable d'activité en IA
- Initiative du CNRS en faveur de l'IA très méconnue

02

Besoins numériques au regard des spécificités de la recherche en physique

Michele Casula (IMPMP)

Adil Khalifa (LKB)

Michele Casula

CNRS & Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, SU

Domaine de recherche : matériaux corrélés par méthodes *ab initio*
en particulier par *Monte Carlo quantique*

Responsable de l'équipe Théorie quantique des Matériaux (TQM)



Équipe TQM

8 membres permanents

4 PU/DR (1 émérite)

3 MCF/CR

1 ITA/IR



Marie-Anne Arrio
CR



Christian Brouder
DR émérite



Delphine Cabaret
PU



Michele Casula
DR



Maria Hellgren
CR



Benjamin Lenz
MCF



Lorenzo Paulatto
IR



Guillaume Radtke
DR

Projets HPC

- **H2020-INFRAEDI** pour la création du **centre d'excellence européen pour le calcul à l'exascale TREX** (début octobre 2020, fin mars 2024)



solution précise du problème électronique à grande échelle
grâce au Monte Carlo quantique
porté sur les futurs ordinateurs à l'exascale

- **HORIZON-EUROHPC-JU-2022-INCO-04 (EuroHPC international cooperation) HANAMI - HPC alliance for applications and supercomputing innovation: the Europe-Japan collaboration** (début février 2024 pour trois ans)



- Équipe TQM participe régulièrement aux appels **eDARI** pour l'allocation des ressources HPC du **GENCI**.



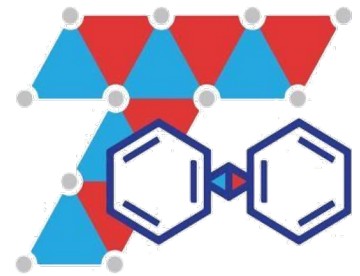
- Les membres de TQM ont accès également au mesocentre **MeSU (Plateforme SACADO)**



Développement de codes



Quantum Monte Carlo package



03

Organisation et exploitation des solutions numériques

Pierre-François Lavallée (IDRIS)



IDRIS

www.idris.fr



PHYSIQUE

Écosystème national du calcul intensif



Journées des DU et DA de CNRS Physique - 06/2024

Contexte

La simulation numérique au sens large (HPC & IA) est devenue, aux côtés de la théorie et de l'expérimentation, le troisième pilier sur lequel repose la recherche scientifique

- Développement modélisation / simulation / données
 - Avancées scientifiques significatives dans de nombreux domaines
 - Explosion des besoins de puissance de calcul
 - Explosion des volumes de données (simulations, grands instruments,...)



Depuis juin 2022 et le supercalculateur Frontier, on est entré dans l'ère de l'Exaflop/s :

10^{18} opérations flottantes double précision par seconde

Les architectures accélérées sont devenues la référence pour les grosses configurations (forte contrainte consommation électrique)

Les capacités de stockage atteignent des centaines de Po

La donnée est au cœur des infrastructures de calcul qui deviennent « Data-centric »

Contexte

La simulation numérique au sens large (HPC & IA) est devenue, aux côtés de la théorie et de l'expérimentation, le troisième pilier sur lequel repose la recherche scientifique

- **« Discipline » transverse**
 - Par nature très inter-disciplinaire
 - Pas de réelle communauté « computational science »
 - ✓ Difficultés à tenir un discours commun cohérent
 - Pas de réelle « appropriation » par chaque communauté
 - ✓ qui compte un peu sur les autres pour défendre le calcul !
 - « Maturité » très différente selon les disciplines
- **Foisonnement d'initiatives plus ou moins coordonnées**
 - Multitude d'acteurs (internationaux, nationaux, régionaux, locaux)
 - Multiplicité des sources de financements (État, Régions, ...)
 - Multiplication des infrastructures
 - Coûts induits élevés

Volonté d'organisation / de coordination du CNRS et plus généralement de l'état (CoSIN)

Évolution du paysage

- **Volonté de l'Etat de rationaliser les infrastructures**
 - **Objectifs : réduire / optimiser le nombre d'infrastructures**
 - ✓ DINUM (Premier ministre, Dir. interministérielle du numérique)
 - ✓ CoSIN (MESR, Comité des services et infrastructures numériques)
 - Fédération des services et infrastructures numériques
 - « cloud ESR »
 - **Processus de labellisation de datacentres nationaux et régionaux**
 - ✓ Cahier des charges
 - Infrastructures (béton, électricité, climatisation,...)
 - Offre de services
 - ✓ 4 Centres nationaux (CC-IN2P3, CINES, IDRIS, TGCC)
 - Rôles datacentre national / régional
 - ✓ 13 centres régionaux
 - Quelques centres supplémentaires en cours d'instruction (Ile-de-France, Lyon,...)
 - ✓ Forte incitation à l'implantation du matériel dans ces datacentres
 - condition déterminante pour attribution de crédits publics
 - ✓ Datacentres non limités aux calculs et données scientifiques
 - systèmes d'information,...

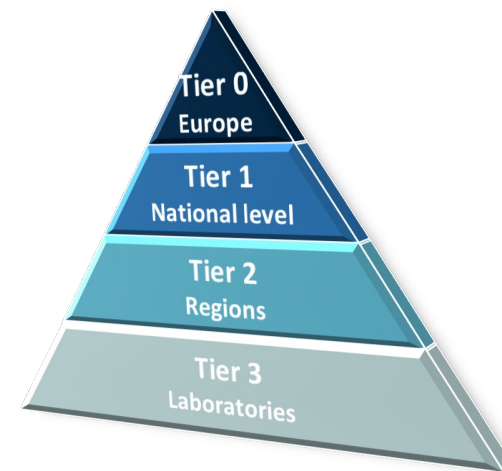
Le paysage : quatre niveaux principaux

• Moyens européen (Tier0)

- EuroHPC
 - ✓ Machines pré-exascales (Finlande, Espagne, Italie)
 - ✓ Machines Pétascale
 - ✓ Futures machines Exascale (Allemagne, France)

• Moyens nationaux (Tier1)

- Pour le CNRS : IDRIS (Orsay, Jean Zay) et CC-IN2P3 (Lyon)
- Pour France Universités : CINES (Montpellier, Adastra)
- Pour le CEA : TGCC (Bruyère-le-Châtel, Joliot-Curie)



• Moyens régionaux (Tier2)

- Méso-centres :
 - ✓ Promotion du calcul intensif
 - ✓ Répondre à des besoins particuliers
 - ✓ Étape préparatoire à l'utilisation
- Centres de compétences (IC)
- ✓ Pas directement opérateurs
- ✓ Souvent adossés à un méso-centre

• Moyens de laboratoires (Tier3)

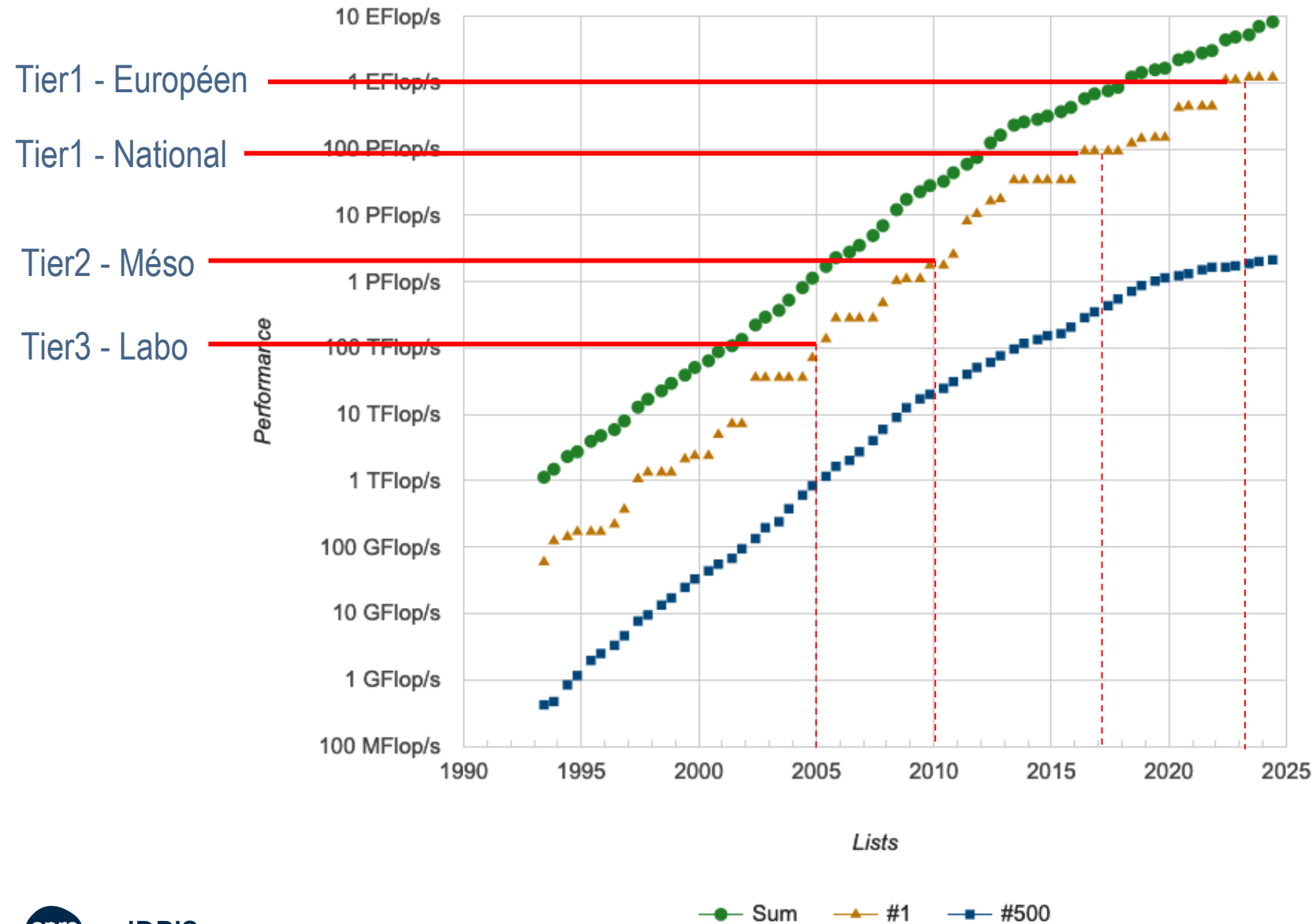
- Ressources de calcul et de stockage limitées mais nécessaires
- Tendances à la dispersion, coût totaux mal connus et mal maîtrisés

**Action structurante possible
Inciter à la mutualisation et la
rationalisation (infrastructures,
matériels, RH, ...)**

Accessibles gratuitement sur
appels à projets compétitifs

Le paysage : quatre niveaux principaux

Performance Development



ECOSYSTÈME DU CALCUL INTENSIF EN REGION

EquipeX MesoNet



❑ **Projet PIA3 via l'AMI « Équipements structurants pour la recherche »**

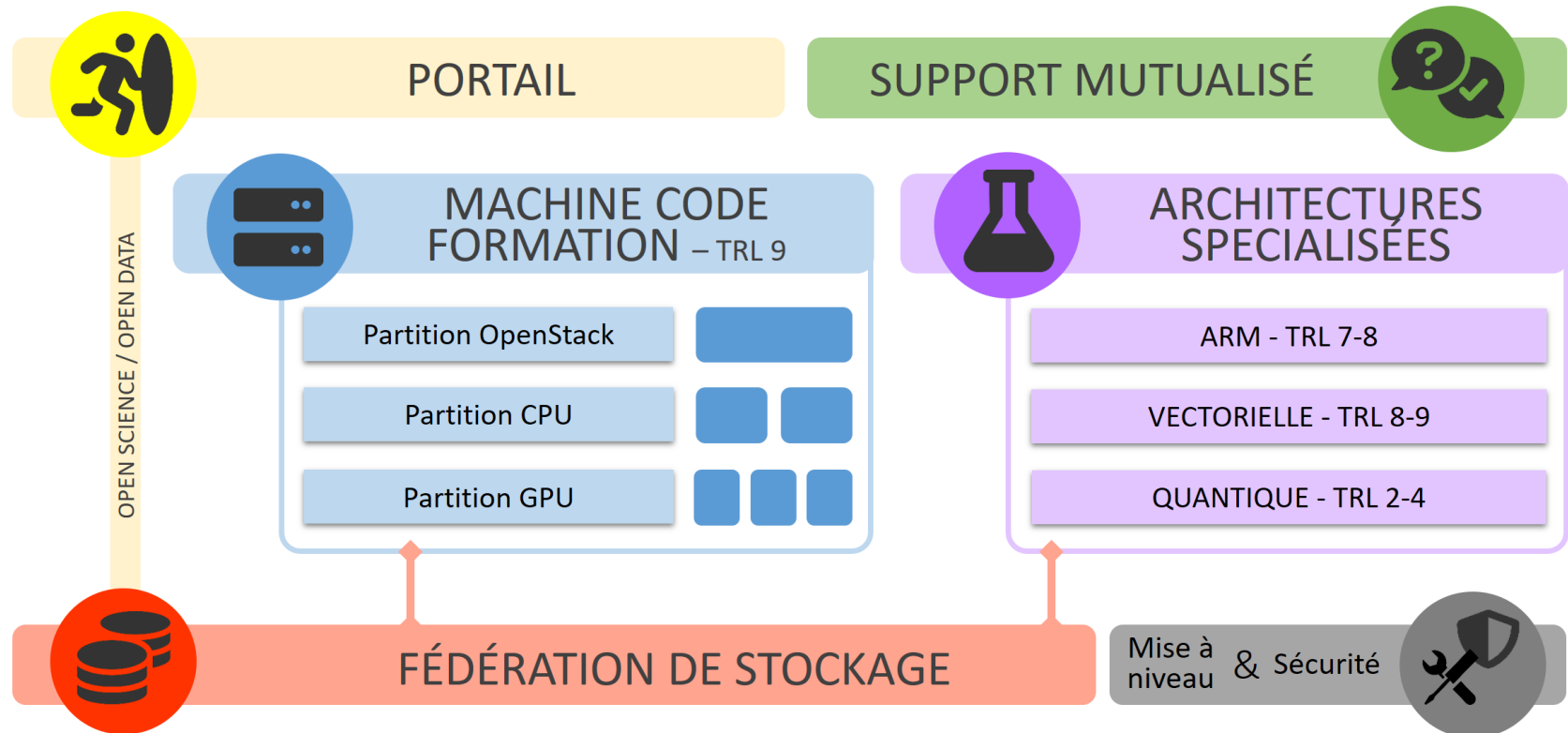
- 21 partenaires + GENCI coordinateur
- Début du projet : 01/10/2021
- Durée : 6 ans
- Budget : 30,4 M€
 - 14,2 M€ financés par le PIA3
 - 16,2 M€ par les partenaires

❑ **Objectifs**

- Renforcer la structuration de l'offre régionale
 - mettre en cohérence les offres régionales
 - mutualiser les expertises
- Disposer d'infrastructures calcul / IA au meilleur niveau technologique avec un bon niveau de sécurité
- Encourager les échanges Tiers1-Tiers2
- Fournir une infrastructure agile pour le développement des codes et la formation

ECOSYSTÈME DU CALCUL INTENSIF EN REGION

EquipeX MesoNET : le projet

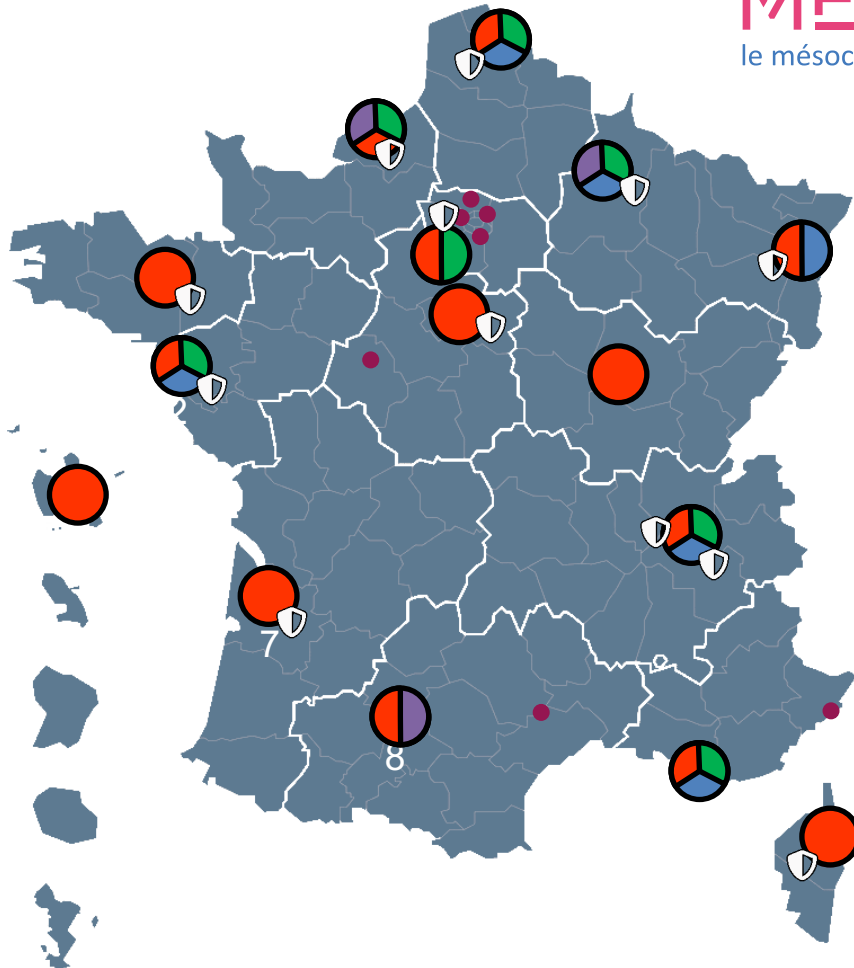


- ❑ 14 sites pour le stockage et 13 audits de sécurité
- ❑ 6 hébergeurs de machine de dev. / formation :
 - Grenoble, Reims, Strasbourg, Lille, Marseille, Nantes
- ❑ 3 prototypes :
 - ARM NVIDIA Occitanie
 - Vectoriel : Normandie
 - Quantique : Grand Est

Briques techno. en cours d'identification :
authentification, stockage, portail etc.

ECOSYSTÈME DU CALCUL INTENSIF EN REGION

EquipeX MesoNET : Infrastructure



- Fédération de stockage
- Personnel support
- Machine code-formation
- Architecture spécialisée
- Audit sécurité

Portail

□ Mise en place d'une gouvernance

- Assemblée Générale
- COPIL
- Conseil Scientifique
- GTs (portail/sécurité/stockage etc.)

Organisation au niveau national



1. GENCI

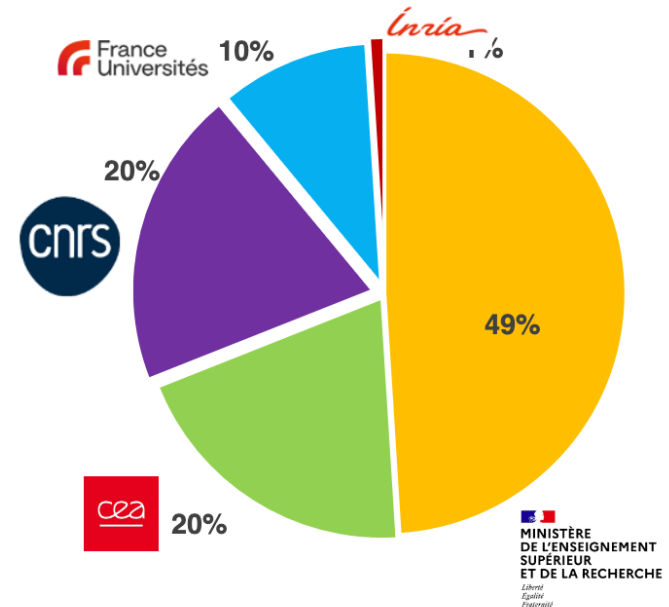
Grand équipement national de calcul intensif - point saillants

❖ Opérateur public de l'ESR ; Infrastructure de recherche ; Société civile créée en 2007

Maîtrise d'ouvrage nationale pour le **calcul intensif** et le **stockage de données** computationnelles associé à l'usage de **du HPC**, de **l'intelligence artificielle** et aux futures technologies quantiques.

❖ Accès gratuit aux heures de calcul et stockage

- Procédures d'appel à projets gérées par GENCI, sur critère d'**excellence scientifique**
- Ouvert aux chercheurs **académiques** et aux **industriels** avec publication des résultats
- **> 3 milliards d'heures disponibles sur les partitions scalaires**
- **> 60 millions d'heures disponibles sur les partitions accélérées**
- **3000 projets actifs par an dans tous domaines** (dont 10% soutien industriel)



❖ Renouvellement régulier des supercalculateurs des 3 centres de calcul nationaux (CINES, IDRIS et TGCC)

Puissance de calcul **> 230 Pflops**

GENCI - AVRIL 2024

Organisation au niveau national



3. COMMENT ACCÉDER AUX RESSOURCES GENCI ?

Les types d'accès aux ressources de GENCI



☐ 2 types d'accès pour tous les besoins

- **AR** : Accès Réguliers + demande complémentaire à mi-parcours
- **AD** : Accès Dynamique

+ Demande « au fil de l'eau » possible pour les AD ou AR et possible n'importe quand

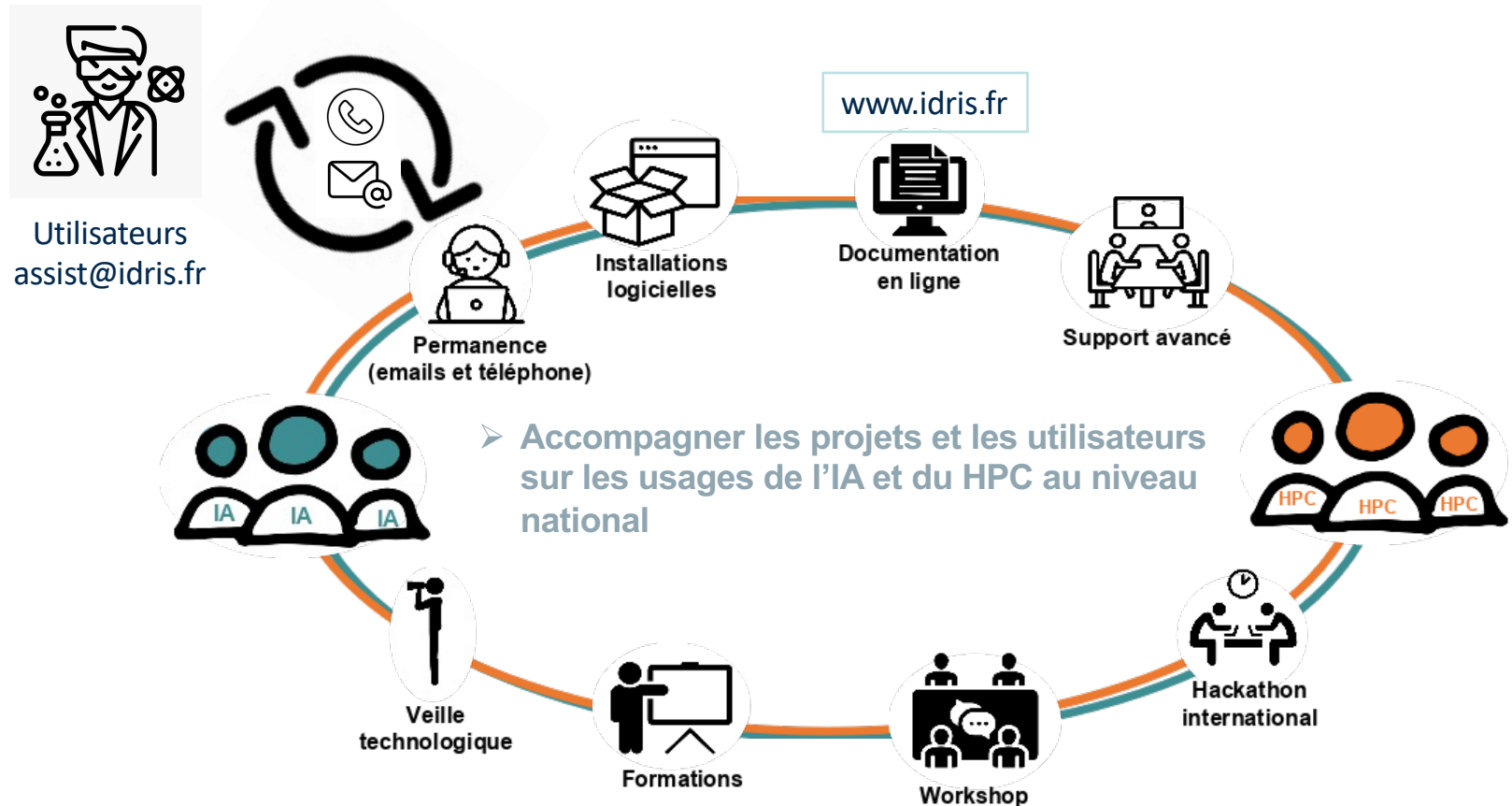
Type d'accès	Accès Réguliers	Accès Dynamiques
Seuil normalisé de la demande calcul	>500 kh cœur ou 50 kh GPU	< au seuil normalisé des AR
Quand postuler aux appels à projets ?	Semestriel	Tout au long de l'année
Pour combien de temps ?	1 an	1 an
Évaluation par un comité ?	Oui	Non
Nombre moyen d'heures demandées	5 Mh cœur	10 kh GPU
Qui peut demander des ressources	Permanent (CDD et Post-Doc)	Permanent (CDD et Post-Doc) + doctorant et stagiaire de M2 pour l'IA
Pour quoi faire	Simulation	Simulation + benchmark, développement
Qui valide la demande	Comité d'évaluation	Directeur du centre de calcul

GENCI - AVRIL 2024

6

De l'expertise en plus des moyens de calcul


- Missions des équipes Support aux Utilisateurs





CONTACTER GENCI & LES CENTRES DE CALCUL

- **GENCI** : www.genci.fr
 - Pour demande générale, communication : contact@genci.fr
 - Support technique : acces@genci.fr
- Inscription à la newsletter GENCI sur le site : [lien](#)

 **Restez informés !** Pour connaître toutes nos actualités relatives au calcul haute performance, à l'intelligence artificielle et au calcul quantique au service de la science, inscrivez-vous à la newsletter GENCI [S'ABONNER →](#)



- **CINES** : www.cines.fr
 - Pour tout support : svp@cines.fr
- **IDRIS** : www.idris.fr
 - Support administratif : gestutil@idris.fr
 - Support technique : assist@idris.fr
- **TGCC** : www-hpc.cea.fr/fr/TGCC.html
 - Pour tout support : hotline.tgcc@cea.fr