



GENCI : Grand équipement national de calcul intensif

Jean-Philippe Proux, Delphine Théodorou, Pascal Voury¹

GENCI est une très grande infrastructure de recherche (TGIR), qui a pour mission de *mettre le calcul intensif au service de la recherche scientifique, de l'innovation et de la compétitivité des entreprises*. GENCI est un opérateur public de l'État, doté d'un budget de fonctionnement de 39 millions d'euros, dont les associés sont le MESRI (49 %), le CEA (20 %), le CNRS (20 %), les universités représentées par la Conférence des présidents d'université (10 %) et enfin Inria (1 %).

Les principales missions de GENCI sont :

- (1) *Définir et mettre en œuvre la stratégie nationale d'acquisition de moyens de calcul intensif et de stockage de données*. Ceci passe par des dialogues compétitifs, avec les compétences nécessaires à l'acquisition de calculateurs et de moyens de stockage dont les performances sont optimales pour les besoins des chercheurs. GENCI finance ces moyens de calcul et de stockage, coordonne leur exploitation et gère l'attribution des ressources mises ainsi à disposition sans coût pour les chercheurs.
- (2) *Promouvoir la simulation et le calcul intensif*. GENCI s'assure que tous les besoins en calcul intensif de la recherche académique, mais aussi de l'industrie, soient couverts avec une implication dans toute l'échelle des moyens disponibles, depuis les centres régionaux, en passant par les centres nationaux, jusqu'aux centres européens.

1. GENCI, <http://www.genci.fr>

(3) *Soutenir l'écosystème du calcul intensif au niveau européen.* GENCI représente la France dans l'infrastructure PRACE (*Partnership for Advanced Computing in Europe*) et dans l'initiative EuroHPC récente, dont l'objectif est de construire des machines d'une puissance de l'ordre de l'exaFLOPS (10^{18} *floating-point operations per second*) en Europe en 2022–2024 (avec du matériel européen, processeurs compris) comme l'a déjà fait la Chine en 2019 et comme le feront les États-Unis en 2021.

De plus, depuis 2019, GENCI est chargé du volet calcul de la mise en œuvre du plan *AI for Humanity*, suite à la remise du rapport Villani en intelligence artificielle. L'ensemble des missions de GENCI sont représentées graphiquement sur la figure 1.

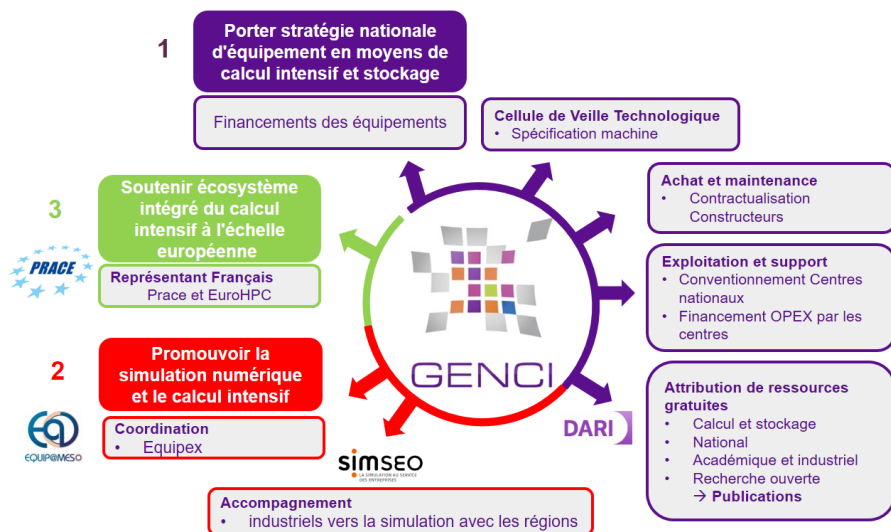


FIGURE 1. Les missions de GENCI.

À l'échelle européenne, les moyens sont répartis en *tiers* comme le montre la figure 2 (Tier2 pour le niveau régional jusqu'à Tier0 pour le niveau européen). Une particularité de GENCI est d'intervenir sur les trois niveaux à la fois. Actuellement, une machine européenne offre de l'ordre de 20 pétaFLOPS (10^{15}) avec plusieurs centaines de milliers de cœurs ; une machine nationale dispose de dizaines de milliers de cœurs pour une puissance de 3 à 15 pétaFLOPS. Nous allons balayer la structure de la pyramide du bas vers le haut : tout d'abord, l'écosystème régional, puis les structures à l'échelle nationale avec l'apport récent du plan *IA for Humanity* ; enfin, les centres européens dont l'évolution va modeler l'avenir de l'offre de moyens de calcul.



FIGURE 2. Répartition des moyens de GENCI.

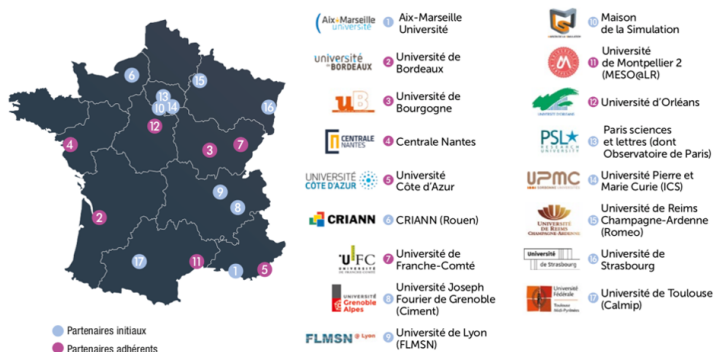


FIGURE 3. Centres de calcul régionaux.

Les centres de calcul régionaux

L'EQUIPEX Equip@meso a permis d'investir 11 millions d'euros dans les centres sur la période 2011-2019, en contribuant au financement d'environ 80 % de la puissance de calcul installée en régions (cf. figure 3). Ce réseau a permis de développer les interactions entre centres, mais aussi de mettre en place un soutien à la simulation numérique pour des TPE, PME et même ETI locales.

Les centres de calcul nationaux

Les centres de calcul nationaux utilisés par GENCI sont au nombre de trois :

- Le CINES (Centre informatique national de l'enseignement supérieur) situé à Montpellier. Il dispose de la machine Occigen, installée en 2015.
- L'IDRIS (Institut pour le développement et la recherche en informatique scientifique) du CNRS à Orsay. Il a installé mi-2019 la machine Jean Zay.

— Le TGCC (Très grand centre de calcul) du CEA, situé à Bruyères-le-Châtel, opère la machine Joliot-Curie, qui est la seule à être mise à disposition de l'infrastructure PRACE pour 60 % de sa capacité.

L'ensemble de ces machines permet, au début de l'année 2020, de mettre à disposition de la communauté recherche française un peu plus de 40 pétaFLOPS. Ces trois centres permettent de mettre à disposition de toute équipe de recherche académique française, ou de tout industriel ayant un projet de recherche ouverte (publiable), des supercalculateurs de technologies différentes mais complémentaires : on dispose ainsi de calculateurs généralistes, mais aussi de machines accélérées par des GPUs, ainsi que de toute l'infrastructure nécessaire pour stocker ou analyser efficacement les résultats des calculs effectués.

Globalement, tous les dix mois, GENCI fait évoluer ces architectures, avec pour résultat un doublement chaque année en moyenne de la puissance de calcul mise à disposition de la recherche.

GENCI coordonne une cellule de veille technologique pour anticiper sur l'architecture des futures machines : cela passe par l'animation d'un groupe de vingt experts issus des associés, mais aussi par l'acquisition et la mise à disposition de prototypes, comme actuellement le supercalculateur Inti au TGCC, constitué de 60 processeurs ARM (1920 cœurs). Deux anciens prototypes, basés l'un sur une architecture à grand nombre de cœurs Intel et l'autre sur une architecture accélérée GPU d'IBM, viennent d'être décommissonnés. Cette veille technologique continue permet d'identifier, de tester les dernières technologies et de s'assurer que l'écosystème mis en œuvre par un constructeur est suffisamment mature pour être efficace pour les chercheurs.

Le plan intelligence artificielle

D'après un bilan établi par le MESRI en 2018, la France est leader mondial pour la production d'articles en IA avec 5000 chercheurs répartis dans plus de 250 équipes. Par ailleurs, 300 *startups*, 80 PME et ETI ont une activité significative.

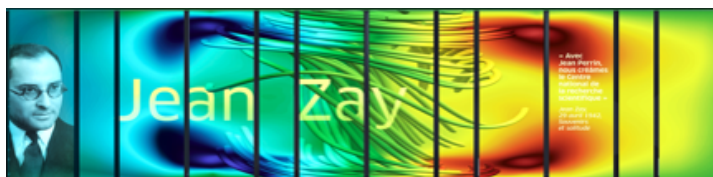
L'objectif de ce plan est de maintenir la France dans le domaine de l'IA en tête en Europe et dans le *top* cinq mondial, en doublant le nombre de doctorants et en permettant l'accès aux calculateurs de GENCI à 300 chercheurs au départ, puis à un millier par la suite.

Cet objectif a été rendu public fin mars 2018 par l'annonce présidentielle du plan *AI for Humanity*, basé sur le rapport Villani. Il est décliné au niveau du MESRI par un plan stratégique présenté par M^{me} Vidal en novembre 2018. Deux buts sont visés : apporter une puissance souveraine pour la recherche française en IA et aider au rapprochement des communautés HPC et IA.

Les conséquences concrètes de ce plan sont :

— *Un supercalculateur en partie dédié à l'intelligence artificielle.* Début 2019, GENCI a acquis l'un des plus puissants supercalculateurs en Europe, dénommé Jean Zay², installé au centre de calcul IDRIS du CNRS sur le plateau de Saclay. Cette machine, d'une puissance de calcul supérieure à 15 pétaFLOPS, étend les usages classiques du calcul à haute performance (HPC) à de nouveaux usages pour l'IA. Il comporte plus d'un millier de processeurs spécialisés, appelés GPU. Trois millions de dollars de dons sont apportés par Facebook, qui permettent d'augmenter encore la puissance de calcul disponible pour l'IA : en mai 2018, Mark Zuckerberg avait annoncé vouloir soutenir la recherche publique française en IA.

— *La facilitation de l'accès au calcul.* Afin de répondre aux besoins spécifiques des communautés de recherche en IA, une procédure particulière d'accès aux moyens de calcul a été définie par GENCI sur recommandation de l'alliance ALLISTENE (présidée par Inria). Cette nouvelle procédure appelée « Accès dynamique³ » a été mise en place en 2019 et permet notamment de disposer de ressources de calcul à la volée.



PRACE et les centres de calcul européens

Les machines disponibles permettent aux scientifiques européens familiers avec le HPC d'accéder à des puissances de calcul supérieures à celles mises à disposition au niveau national, grâce à deux appels à projets par an⁴. La machine Joliot-Curie, installée au TGCC, présente la particularité d'être disponible aussi bien grâce aux appels à projets nationaux qu'euroéens avec PRACE.

Les scientifiques français ont su s'appropriier ces ressources : à l'échelle européenne, ils font partie des premiers bénéficiaires en nombre de projets ou en nombre d'heures.

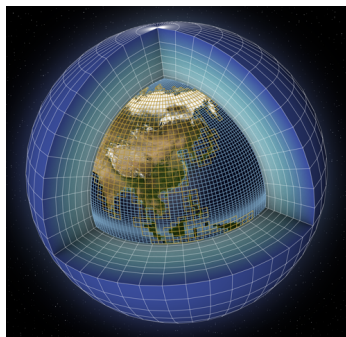
2. https://www.youtube.com/watch?time_continue=92&v=abtueZZzlrk

3. <https://www.youtube.com/watch?v=Marx-BSFMN0&t=1s>

4. <http://www.prace-ri.eu>

Exemples d'utilisation des moyens de GENCI

Recherche académique : modélisation du climat



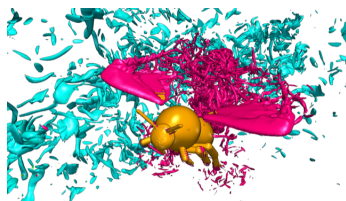
Terre grillée à très haute résolution.

©Animea, F. Durillon pour le
LSCE/CEA/IPSL.

aux négociations de la COP21. Le dernier exercice en date, CMIP6, a vu la mise en œuvre d'une nouvelle génération de modèles climatiques de plus haute résolution. Là encore, GENCI a mis à disposition une allocation dédiée de calcul et des espaces de stockage spécifiques aux équipes de l'IPSL sur la période 2015 à 2019, ce qui leur a permis d'être les premiers à mettre les résultats de leurs simulations à disposition de la communauté internationale.

Recherche industrielle : le bourdon, source d'inspiration pour les micro-véhicules aériens

Le vol du bourdon se révèle être une mine d'informations pour concevoir de nouvelles générations de micro-véhicules aériens (drones par exemple). La mise au point d'une soufflerie numérique a permis de faire voler un modèle de bourdon à une vitesse de 2,5 m/s (9 km/h). Les simulations ont montré que les mécanismes de vol du bourdon pour créer de la portance sont robustes même en fortes turbulences, car il n'y a pas d'énergie supplémentaire requise. Néanmoins le contrôle de la turbulence reste un défi pour le bourdon, mais pas d'un point de vue énergétique.

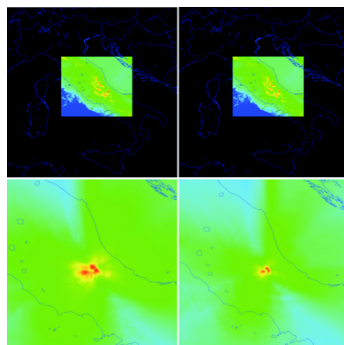


Soufflerie numérique montrant les tourbillons créés par les ailes battantes du bourdon. Kai Schneider, université Aix Marseille/CNRS.

Aide à la décision politique : Aquila, aider les autorités italiennes

La région de l'Aquila fut touchée le 6 avril 2009 par un séisme de magnitude 6,2 sur l'échelle de Richter. Le danger supplémentaire provenait des répliques très fortes qui se produisent durant les jours suivants. Il était donc crucial de pouvoir calculer rapidement et avec précision des scénarios de répliques possibles.

L'illustration ci-contre montre une carte de vitesse maximale du sol illustrant l'aléa sismique pour deux scénarios de répliques possibles : on voit que le premier scénario produit plus de dégâts que le deuxième.



Dimitri Komatitsch, université de Pau et des Pays de l'Adour, CNRS et INRIA.

Comment accéder aux ressources de GENCI, pour quelles ressources ?

Les conditions d'éligibilité pour les chercheurs académiques et industriels sont les suivantes :

- il est impératif qu'il s'agisse de travaux de recherche ouverte ; ce qui implique une obligation de publication ;
- le porteur de projet doit être membre permanent du laboratoire d'appartenance, avec un financement français (doctorants, post-doctorants, etc., ne peuvent pas être porteurs de projet, sauf dans le cas de la procédure « Accès dynamiques ») ;
- une sélection a lieu sur des critères d'excellence scientifique (pour les accès réguliers).

Ces conditions permettent un accès gratuit au calcul, au stockage, mais aussi aux équipes support aux utilisateurs des centres, ainsi qu'aux formations proposées par celles-ci. Il existe plusieurs types d'accès aux ressources de GENCI : accès réguliers (AR), accès préparatoires (AP) et accès dynamiques (AD). AR et AP sont les accès classiques pour le calcul HPC. AD a été mis en place en 2019 pour faciliter un accès rapide aux ressources de Jean Zay pour les chercheurs en IA : ils ne sont pas soumis à sélection par un comité. Demander un AR permet de candidater sur les trois centres nationaux en même temps : ils concernent environ 800 projets par an pour plus de 3500 utilisateurs.

Le tableau ci-dessous résume les particularités des différents accès aux ressources de GENCI.

Type d'accès	Accès Régulier	Accès Préparatoire	Accès Dynamique
But	Calcul important, mode production	Test, développement, étude d'extensibilité	Développement, d'algorithmes en IA
Quand postuler ?	Biannuel	Tout au long de l'année	Tout au long de l'année
Pour combien de temps ?	1 an	6 mois	1 an
Évaluation par un comité	Oui	Non	Non
Nombre moyen d'heures	5 Mheures-cœur	50 kheures-cœur	10 kheures-GPU

Plus de 5000 dossiers ont été expertisés depuis 2010. Parmi les projets qui se sont vus attribuer des heures sur les machines nationales, 60 % utilisent aussi un mésocentre régional ; 32 % sont soutenus par un projet ANR et près de 20 % par un industriel. Les demandes d'heures émanent de tous les domaines de recherche et sont répartis en cinq domaines comme le montre la figure 4.

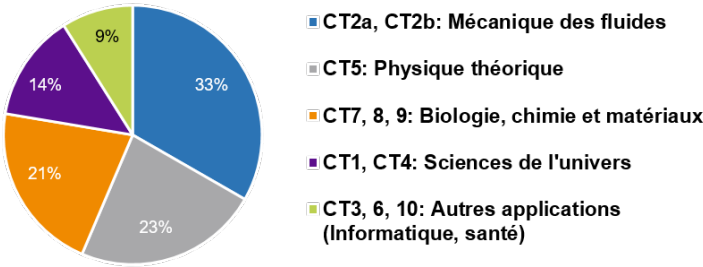


FIGURE 4. Répartition des heures demandées pour les principaux comités thématiques.

Depuis 10 ans, malgré l'augmentation régulière des moyens mis à disposition par GENCI, la demande reste plus élevée que l'offre comme le montre la figure 5. Il est donc important, pour les gros projets, de pouvoir participer aux appels d'offre européens PRACE afin de ne pas « encombrer » les demandes à l'échelle nationale. GENCI, avec son positionnement et son action à tous les niveaux (régional, national, européen), permet une gestion cohérente de ces moyens de calcul, quels que soient les besoins du projet candidat.

Le futur : l'initiative européenne pour l'Exascale

La création d'une entreprise commune, EuroHPC, signée en 2019 entre la Commission européenne et 24 pays, dont la France, acte la mise en place de budget d'un milliard d'euros et le démarrage de la mise en place d'une infrastructure de calcul et

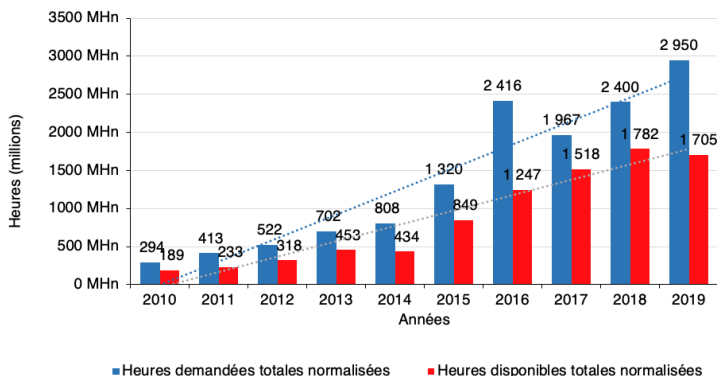


FIGURE 5. Évolution des demandes et disponibilités d’heures de calcul (en millions).

de données intégrées de classe mondiale. Le principe est d’acquérir des machines en copropriété entre l’Europe et les différents consortiums, chacun apportant la moitié du financement. Cette entreprise commune (*Joint Understanding – JU*) EuroHPC⁵ s’appuie sur les structures existantes : PRACE pour les machines (et le stockage) et GEANT pour la partie réseau. L’objectif de JU est de développer puis de mettre en œuvre des technologies et des logiciels exclusivement européens, en faisant un enjeu de souveraineté.

Dans une première phase, cinq machines multi-Petascale et trois machines pré-Exascale sont financées.

- multi-Petascale : 20 pétaFLOPS, financement à 35 %. Cinq consortiums portés par le Luxembourg, le Portugal et l’Espagne, la Slovénie, la République Tchèque, la Bulgarie.
- pré-Exascale : 150 pétaFLOPS avec des GPUs, financées à 50 %, trois consortiums :

- (1) BSC, Espagne avec le Portugal, l’Irlande, la Roumanie, la Croatie, la Turquie.
- (2) CSC, Finlande avec la Belgique, le Danemark, la Norvège, la Pologne, la Suède, la République Tchèque et la Suisse.
- (3) CINECA, Italie avec l’Autriche, la Slovaquie, la Slovénie et la Hongrie.

Dans une deuxième phase, un nouveau financement de 5 milliards d’euros sera débloqué, dont une partie servira à l’acquisition de deux machines Exascale.

5. <https://eurohpc-ju.europa.eu/>

Il reste d'ici là à inventer les modes de financement pour faire face non seulement à la moitié du coût d'acquisition, mais surtout aux coûts de fonctionnement pour les centres (consommation électrique propre, plus refroidissement). Néanmoins, la très forte capacité de calcul attendue en France, avec la machine Exascale, permettra de desserrer les contraintes, et ce d'autant plus que les besoins en capacité de calcul de certaines communautés scientifiques croissent à une vitesse soutenue alors même que d'autres commencent seulement à s'intéresser aux potentialités de la simulation numérique associée au traitement de données massives et à l'intelligence artificielle.