Mission pour les initiatives transverses et interdisciplinaires

Données massives pour la découverte scientifique : production, sélection, curation et analyse

Appel à projets 2025

Le CNRS, à travers la Mission pour les initiatives transverses et interdisciplinaires (MITI) et la Direction des Données ouvertes de la recherche (DDOR), lance en 2025 l'appel à projets « Données massives pour la découverte scientifique : production, sélection, curation et analyse ».

On assiste depuis deux décennies à une **diversification** des données produites dans toutes les disciplines scientifiques. Que ce soit pour le suivi des populations, de leurs déplacements et de leur santé, pour l'analyse de l'évolution du climat, des écosystèmes, pour la reconstruction 3D de vestiges archéologiques et paléoanthropologiques, pour le réglage de dispositifs technologiques, de nombreux domaines scientifiques font aujourd'hui un usage intensif de la modélisation numérique en exploitant des masses de données, pour appuyer la compréhension et la découverte scientifique.

Les recherches reposant sur des phases d'observation permettant de générer des hypothèses suivies d'expérimentations pour les valider, sont ainsi confrontées à la question de la production et du traitement adéquat de données. Dans un mode idéal, il s'agit de sélectionner – ou de concevoir - des capteurs et instruments permettant de produire les données nécessaires et suffisantes pour répondre à une question scientifique prédéterminée avec des approches d'analyse et de modélisation adaptées. Cependant, ce schéma (ou cercle vertueux) s'avère de plus en plus complexe à implémenter dans le contexte croissant de massification des données ou pour adresser des problèmes pluriels et développer une approche holistique.

Dans les disciplines liées à l'observation, des sciences de l'univers aux sciences du vivant, la démocratisation d'instruments de mesure a conduit à la production massive « en routine » de jeux de données standardisés de très grande taille (imageries, spectres, génomes) sur un objet scientifique. La conservation de données brutes devenant impossible, il faut développer des stratégies de sélection ou de prétraitement des données « in situ » au moment de leur acquisition, en fonction des objectifs scientifiques. Le caractère sensible du traitement de la donnée réside ici dans le choix du processus de traitement réalisé avant leur effacement et la garantie apportée sur la fiabilité des résultats issus de ce traitement.

Une autre tendance récente consiste à générer des **données associées à des individus**, en particulier dans les disciplines en lien avec le suivi des socio-écosystèmes. Ces données permettent le suivi ou l'identification de personnes, d'espèces ou de zones protégées avec cependant un risque pour leur anonymat. Il faut trouver le juste équilibre entre la précision des données analysées et la qualité des analyses attendues. Il s'agit ainsi de développer des dispositifs expérimentaux garantissant l'anonymat « per se », des méthodes mathématiques ou informatiques qui opèrent sur des agrégats de données en préservant les données individuelles, mais aussi des considérations juridiques et éthiques associées à l'arbitrage entre exploitation de données à des fins scientifiques et protection des individus. Ces questions qui doivent garantir la participation des acteurs à la collecte de données sensibles trouvent actuellement peu de réponses.

Cet appel vise à soutenir des initiatives de recherches pluridisciplinaires pour lesquelles l'acquisition, le tri, l'agrégation, la curation, le croisement, l'analyse ou l'exploitation des données engendrent des questions scientifiques avancées. Les projets doivent se positionner clairement au-delà des questions opérationnelles de la manipulation et du stockage des données. L'AAP sollicite des propositions qui portent sur des questionnements de recherche autour de la massification des données stratégiques et/ou sensibles. La co-conception entre instrumentation (et/ou méthodes) et données fait partie intégrante des sujets attendus dans cet AAP.

Ces projets doivent développer des approches originales, allant des dispositifs expérimentaux ou participatifs pour la collecte des données à leur analyse via des modèles physiques ou mathématiques, visant à produire des données précises en réponse à des questions scientifiques. Les projets attendus concernent la production, l'expérimentation, la conservation ou la transmission de données en adéquation avec les objectifs scientifiques, ainsi que les méthodes de de pré-traitement (e.g., biologging), de traitement, d'analyse et d'intégration dans des modèles numériques. Le croisement de données hétérogènes dans le but de développer de nouvelles approches scientifiques peut s'inscrire dans la thématique. Un enjeu important consiste à sécuriser ou agréger les données tout en les protégeant. Les recherches juridiques et éthiques visant à équilibrer les besoins en données avec le respect de la vie privée, de leur caractère sensible et des enjeux environnementaux sont également pertinentes. Les approches low-tech ou d'optimisation des temps/besoins de calculs sont particulièrement encouragées.

Les projets devront associer des chercheurs et chercheuses de différentes disciplines, travaillant dans au moins deux unités rattachées à titre principal à deux instituts différents du CNRS.

Critères d'éligibilité

- Le projet doit être porté par **un personnel scientifique titulaire, appartenant à une unité du CNRS** (CR, DR, CPJ, MCF, PR, IR).
- Le consortium doit réunir **au moins deux unités rattachées principalement à deux instituts différents du CNRS**. Le consortium peut impliquer des partenaires d'autres organismes de recherche que le CNRS, des partenaires étrangers ou des entreprises, mais ceux-ci ne peuvent pas être financés.
- La **demande budgétaire doit être clairement exposée, sur deux ans, et ne doit pas excéder 30 k€ par an**. Elle ne peut concerner que des dépenses de mission, de fonctionnement (hors équipement informatique et bureautique standard), d'équipement et de prestation de service en lien avec le projet. Aucun financement de ressources humaines ne peut être accordé. Le recours aux équipements informatique et bureautique (ordinateurs, nœuds ou heures de calcul, etc.) des laboratoires et/ou des moyens nationaux est recommandé. Exceptionnellement, un financement de ces équipements peut être accordé pour des besoins strictement liés à la conduite du projet, sur argumentaire.
- La **signature de la directrice ou du directeur d'unité** du porteur ou de la porteuse du projet est requise.

Évaluation des projets

- L'interdisciplinarité, la prise de risque, la rupture et le caractère exploratoire du projet seront les critères clefs pris en compte dans la sélection des projets.
- Les projets sont évalués par un comité scientifique constitué par la MITI et représentant les dix instituts du CNRS. Les lauréats et les lauréates sont individuellement contactés à l'issue de l'arbitrage, soit environ 2 mois après la date limite de dépôt des candidatures.

Modalités administratives et financières

- Les crédits sont versés à l'unité de rattachement CNRS du porteur ou de la porteuse du projet, qui les engage pour l'ensemble des partenaires. Ils sont de type Subvention d'État, ce qui implique qu'ils doivent être entièrement consommés avant le 31 décembre de l'année d'attribution, et qu'aucun frais de gestion ne peut être prélevé.
- À titre exceptionnel, une gratification de stage (de 3 à 6 mois) peut être accordée. Le financement sera directement notifié à la Délégation Régionale, qui établira la convention de stage. Le ou la stagiaire doit être recruté(e) par une unité CNRS. Le montant de la gratification de stage ne doit pas figurer dans le budget global du projet scientifique.
- Un rapport scientifique et financier sera demandé au porteur ou à la porteuse du projet à la fin de chacune des deux années de financement. Il ou elle s'engage par ailleurs à présenter les résultats de ses recherches lors de la journée de restitution intermédiaire qui se déroulera dans les locaux du CNRS, sur le site Paris Michel Ange. Les frais de mission sont à prendre en compte dans le budget.
- La seconde année du projet est financée sous réserve d'une évaluation positive du rapport scientifique et financier et de la présentation des résultats lors de la journée de restitution, et dans la mesure où le budget de la MITI le permet.

Il est demandé aux lauréates et lauréats de mentionner le financement obtenu (Ce projet a obtenu le soutien financier du CNRS à travers les programmes interdisciplinaires de la MITI ; This project has received financial support from the CNRS through the MITI interdisciplinary programs) dans toute production scientifique.

Dans le cadre de sa politique de science ouverte, le CNRS encourage ses scientifiques à ne pas payer de frais de publications (APC) et de déposer systématiquement les productions scientifiques issues des projets soutenus dans une archive ouverte, de préférence HAL.

Calendrier

- Ouverture de l'appel à projets : 16 décembre 2024
- Clôture des candidatures : 05 mars 2025 à midi (heure de Paris)
- Arbitrage et informations aux candidats : Première semaine d'avril 2025
- Allocation des crédits : Avril 2025

Dépôt de candidature

Pour déposer une candidature, veuillez-vous connecter à l'application NOA : https://noa.cnrs.fr/. Vous serez invité(e) à vous connecter via la « fédération d'identité »

RENATER (codes JANUS pour les agents des unités CNRS) et vous pourrez choisir l'appel « Données massives » dans la liste des appels. Nous vous incitons vivement à utiliser votre adresse institutionnelle afin de bénéficier du préremplissage de certains champs (informations vous concernant ou concernant votre laboratoire ainsi que ceux de vos partenaires).

Date limite de dépôt des candidatures : mercredi 05 mars 2025 à midi (heure de Paris)

Contacts

Suivi du projet : Arthur Panier (MITI ; miti.contact@cnrs.fr)

Responsables scientifiques de l'appel :

- Anne Siegel, CNRS Sciences informatiques
- Pascale Garcia, CNRS Ecologie & Environnement