



MI
Mission pour
l'Interdisciplinarité

DEFI MASTODONS

Quatrième Appel à projets

Les Big Data en recherche

Préambule

Différents programmes de recherche (ANR, PIA, H2020) ont été consacrés aux masses de données. Depuis 2012, le CNRS a également contribué à ces avancées en lançant le défi Mastodons (<http://www.cnrs.fr/mi/spip.php?article53>). Celui-ci propose un amorçage et un cadre dans lequel chercheurs et enseignants-chercheurs peuvent explorer de nouvelles approches interdisciplinaires pour appréhender les problèmes auxquels les grandes masses de données les confrontent. Le défi vise aussi à favoriser l'émergence d'une communauté scientifique interdisciplinaire autour des Big Data et de la Science des Données, capable de rivaliser avec les grands laboratoires internationaux et de produire des solutions originales sur le périmètre des données scientifiques.

De nombreux domaines scientifiques (ex : le séquençage haut débit, observatoire virtuel et grands instruments en astronomie, grands équipements de rayonnement synchrotron, de diffusion de neutrons et de physique des particules, simulation en physique et énergie, imagerie médicale, données environnementales biotiques comme abiotiques), économiques (ex : e-commerce, systèmes décisionnels), ou sociaux (ex : réseaux sociaux, bibliothèques numériques, patrimoines culturels) produisent et consomment des volumes de données considérables. La gestion, l'analyse, l'exploitation, le stockage, la sécurisation et la valorisation de ces données massives peuvent demander l'utilisation d'outils qui n'existent pas encore, ou qui proviennent d'autres domaines et qu'il faut adapter.

De surcroît le CNRS, à travers ses unités de recherche et ses grands instruments, concentre plusieurs centaines de bases de données et de corpus d'informations dont les volumes croissent de façon exponentielle. Lorsqu'on ajoute à cela l'ouverture des données (Open Data) et la corrélation entre sources de données (Linked Data), on se retrouve face à des opportunités uniques d'extraire de nouvelles connaissances, jusqu'alors inaccessibles. Toutefois, en plus des défis mentionnés ci-dessus, l'utilisation de ces données pose de nombreux problèmes d'hétérogénéité, de sémantique et de droits d'usage.

Le défi Mastodons a progressivement constitué une communauté scientifique interdisciplinaire autour des Big Data et de la Science des Données. Cette communauté s'est structurée au sein d'un Groupement de recherche créé en 2015, le GDR MaDICS (<http://www.madics.fr>).

Objet du 4^e Appel A Projet

Ce quatrième appel à projets du défi Mastodons a pour objectif d'impliquer de nouveaux domaines et de nouveaux acteurs dans l'exploitation de données massives, que celles-ci soient fraîchement acquises ou générées, ou qu'elles existent déjà mais font l'objet d'un traitement novateur. L'appel vise également à susciter des actions de recherche s'intéressant aux multiples challenges que représente une obligation, pour les chercheurs, à produire de la donnée ouverte.

Bien que portant sur tout le spectre de recherche des Big data, cet appel vise notamment à faire émerger des projets traitant de la qualité des données. En effet, quelle que soit leur origine, les données peuvent souffrir de multiples problèmes dus à leur hétérogénéité, leur sémantique ou leur transformation. Elles peuvent être incomplètes, incertaines et incohérentes. Les erreurs ou les imperfections peuvent être

d'origine technologique (survenant au niveau du capteur, de l'instrument de production, de la mémoire de stockage, du transfert sur le réseau...), d'origine humaine (générées lors des saisies, des annotations ou de l'interprétation des données...) ou, d'origine logicielle (erreurs de calcul, mauvaises transformations de formats, données tronquées, ...). Les solutions proposées doivent tenir compte de la dimension multicritère du problème et du caractère subjectif de l'appréciation de la qualité propre à chaque domaine d'application.

Dans cette perspective, le CNRS souhaite soutenir une quatrième vague de projets visionnaires, capitalisant des connaissances de plusieurs disciplines, et permettant de franchir un pas significatif dans le domaine applicatif concerné et/ou dans le domaine même de la Science des Données. Ces recherches peuvent impliquer, mais de façon non exclusive, les défis suivants, liés aux grandes masses de données :

Les défis visés

- **Défis liés à la gestion des données**
 - o Stockage, indexation et interopérabilité des données, archivage, ouverture et préservation des données ;
 - o Calcul intensif sur des grands volumes de données, parallélisme dirigé par les données ;
 - o intégration de données hétérogènes, de différentes natures et à différentes résolutions ;
 - o Problèmes liés à la préparation de données massives pour l'Open Data.
- **Défis liés à l'extraction de connaissances**
 - o Recherche d'information et requêtes complexes sur les grandes données ;
 - o Visualisation de grandes masses de données, exploration et analyse visuelle ;
 - o Datamining et textmining, apprentissage statistique et apprentissage profond.
- **Défis liés à la qualité des données**
 - o Perception de la qualité selon les domaines et les usages, identification des facteurs de qualité et des métriques associées;
 - o Modèles formels de qualité (fidélité de la représentation, estimation, prédiction, cohérence, optimisation, techniques de compensation) ;
 - o Benchmarking et procédures de mesure et de test, analyse d'impact sur les usages et la décision, résistance des données à des modèles approximatifs ou multi-échelles.
- **Défis liés à des problématiques sociales**
 - o Problèmes de propriété, de droit d'usage, droit à l'oubli, sécurité, confidentialité, aussi en lien avec les politiques d'Open Data ;
 - o Gains et risques sociaux de la combinaison de données détaillées sur les comportements humains, géo-localisées et résolues temporellement ;

Critères et modalités de soumission

Les consortiums intéressés peuvent déposer un projet scientifique (3 à 5 pages maximum) comportant les éléments suivants :

- Vision scientifique de l'équipe/consortium sur les thèmes du défi ;
- Les verrous scientifiques et les axes de recherche à moyen terme, avec un focus particulier sur la première année ;
- Les acquis scientifiques dans le domaine ou dans un domaine connexe susceptible de contribuer aux problèmes scientifiques ou sociétaux posés (publications significatives, projets passés ou en cours, applications réalisées, logiciels, brevets...) ;
- Les différentes disciplines impliquées et leurs contributions respectives au projet ;
- Une liste de 3 à 5 chercheurs seniors impliqués de façon significative dans le projet.
- L'originalité du projet par rapport aux recherches actuellement en cours dans les équipes impliquées.

L'Appel A Projets 2017 est ouvert aux chercheur(e)s et enseignant(e)s-chercheurs relevant ou non d'une unité CNRS, mais le porteur du projet doit appartenir à une unité CNRS (UPR, UMR, USR...). Bien que les projets soumis soient pluriannuels, leur financement se fait de façon annuelle en tenant compte des résultats scientifiques, de la synergie créée dans le projet et de l'évolution du budget de la Mission pour l'interdisciplinarité du CNRS.

Les projets sont évalués par un comité d'experts qui déterminera les montants alloués selon la nature du projet, son ambition, les résultats visés et le nombre de partenaires. Les expertises ne seront pas communiquées. Le porteur sera informé par courriel de la décision de ce comité.

Le budget alloué est exclusivement destiné aux frais de fonctionnement du projet. Les frais de personnels (CDD, salaires doctorants, post doctorants...) ne sont pas éligibles. A titre exceptionnel, quelques gratifications de stages pourront être autorisées mais leur nombre est limité pour l'ensemble du défi. Le budget alloué sera versé à l'unité CNRS de rattachement du ou de la porteur-e du projet, qui sera en charge de son utilisation selon les besoins scientifiques du projet.

Chaque porteur-e de projet s'engage à rédiger un rapport d'activité et un bilan financier et à présenter ses résultats au colloque de restitution annuel du défi. L'analyse de ce rapport et de cette restitution servira de base à la décision de prolongation ou d'arrêt du projet.

=====
Calendrier du déroulement de l'Appel A Projets 2017:

novembre 2016 : lancement de l'appel

décembre 2016 : clôture de l'appel

Janvier 2017 : expertise des candidatures

Février 2017 : réunion du comité d'évaluation

Février 2017: informations aux lauréats.

Pour candidater :

Il vous faudra remplir le formulaire de candidature disponible à l'URL :

<http://www.cnrs.fr/mi/spip.php?article985>

Le formulaire complété doit être obligatoirement déposé par le porteur du projet sur l'application SIGAP: <https://sigap.cnrs.fr/sigap/web/connexion.php>

(Le dossier complet ne doit pas dépasser 6 pages et être en format pdf ou word).

Date limite de dépôt des projets : 13 décembre 2016 à midi.

Contacts :

Mission pour l'interdisciplinarité du CNRS (mi.contact@cnrs.fr)