



# MAESTRO : MAsses de données En ASTROphysique

<http://www.madics.fr/actions/actions-en-cours/#maestro>

C. Surace, S. Maabout et le consortium MAESTRO  
LAM : Laboratoire d'Astrophysique de Marseille  
LaBRI : Laboratoire Bordelais de Recherche en Informatique

## RESUME

MAESTRO est une action du GDR Madics pour entamer des collaborations profondes entre la communauté astrophysique et la communauté des sciences des données. L'objectif principal de l'action que nous proposons dans le cadre du GDR est l'organisation et la mise en réseau d'une communauté interdisciplinaire de chercheurs et d'ingénieurs qui s'intéressent à ces problèmes pour permettre une fertilisation croisée entre différents champs scientifiques et technologiques.

## Activités de l'Action :

Les actions que entreprises dans le cadre de l'action se focalisent sur :

- la mise à disposition de données astrophysiques,
- la mise à disposition de requêtes types intéressant astrophysiciens,
- l'échange autour de résultats obtenus sur différents outils de gestion de données (BD classiques, Hadoop, Spark, SciDB, Vertica,...)
- la fouille de données en astrophysique : par exemple, en étudiant l'apport de l'échantillonnage et de l'algorithmique approchée ou en envisageant l'exploration (résultats partiels) versus interrogation des données (résultats complets).
- les méthodes de caractérisation des objets astronomiques
- la visualisations de données astronomiques mais aussi de résultats de fouille de données

L'action MAESTRO a pour but aussi de promouvoir les activités transverses qui concerne l'Astrophysique et le BigData. N'hésitez donc pas à proposer des activités qui pourraient être soutenues par l'Action et le GdR. Et puis surtout, n'hésitez pas à parler de l'action et du GdR MADICS autour de vous.

## Introduction

Dans le domaine de l'astronomie, l'évolution rapide des télescopes et autres instruments scientifiques ainsi que le recours intensif à la simulation informatique ont conduit, ces dernières années, à une production massive de données. L'exploitation de ce type de données pose alors de nombreux défis pour les scientifiques experts du domaine ainsi que pour la communauté.

## Laboratoires impliqués :

Plusieurs laboratoires issus de 3 instituts différents du CNRS sont directement impliqués dans cette action :

- LaBRI (Bordeaux)
- LAM (Marseille)
- LIF (Marseille)
- LIMOS (Clermont)
- LIRMM (Montpellier)
- LPC (Clermont)

Le LPC et le LAM participent aux grands projets de surveys astrophysiques (LSST, EUCLID) et sont particulièrement impliqués dans les problématiques de traitement de ces données.

Le LaBRI, le LIF, le LIMOS et le LIRMM sont très impliqués dans les recherches d'optimisation pour l'analyse, le stockage de très grandes masses de données d'images ou d'information, distribuées.

## MADICS

<http://www.madics.fr>



La création du Groupe de Recherche GdR MaDICS (Masses de Données, Informations et Connaissances en Sciences) s'inscrit dans un monde en réseau qui génère un déluge de données de plus en plus difficile à appréhender par les humains.

MaDICS a pour objectif de mettre en place un écosystème afin de faire avancer les connaissances scientifiques et technologiques sur la collecte de grandes masses de données scientifiques, leur organisation, leur stockage et préservation, l'extraction de connaissances, la recherche et l'analyse statistique, l'interprétation et la visualisation pour la prise de décision et la production de valeur.

MaDICS fournit donc un cadre ouvert et flexible permettant à des scientifiques de rejoindre/proposer des activités interdisciplinaires dont le point de départ sont les masses de données en Sciences.

## Données Astrophysiques

Les instruments, sols ou spatiaux, sont de plus en plus perfectionnés et sensibles ; le volume et la qualité des données recueillies imposent de revoir les méthodes de réduction, d'analyse et d'interrogation de ces données.

En effet, les données collectées sont volumineuses et complexes (grandes images, données incertaines, multi-échelle, incomplètes, imprécises, bruitées, incohérentes,...). De plus elles sont entachées d'erreurs à chaque étape de leur création, depuis l'acquisition des données (détecteur), du traitement (image), de l'analyse et de l'exploitation.

Des grands projets en cours tels que EUCLID (<http://euclid-ec.org>) ou GAIA (<http://sci.esa.int/gaia/>) ou LSST (Large Synoptic Survey Telescope : <http://www.lsst.org/>) vont fournir une quantité énorme de données du ciel et posent avec eux des problèmes statistiques importants tels que les validations de modèles, les problèmes inverses mal contraints, réduction de dimensionnalité, classifications, fouille de données, ...).

La mission spatiale Gaia prévoit de gérer un volume de données de l'ordre du Petabyte relatif à plus d'un milliard d'objets, et dont la collecte a débuté depuis mi 2014. Dans le cadre du projet LSST, qui vise la construction d'un télescope de nouvelle génération, il est prévu la mise à la disposition des scientifiques et des utilisateurs finaux d'un volume de données total d'environ 140 Petabytes à la fin de cette décennie

Par ailleurs, dans le domaine des simulations cosmologiques, grâce à l'évolution des logiciels et de la puissance de calcul disponible, le facteur limitant devient actuellement la capacité de stocker, transférer et analyser les données produites par les simulations



60 Pbytes de données seront transmises par le LSST pendant 10 ans et GAIA observera 40 Millions d'objets tous les jours



## Inscrivez vous



Si vous êtes intéressé(e)s pour participer à cette action, apporter vos connaissances et jouer avec des données astrophysiques, c'est très simple :

- **Inscrivez vous sur le site du GDR MADICS (<http://www.madics.fr>)**
- **Complétez votre profil en notifiant l'Action MAESTRO.**

• Vous recevrez automatiquement les informations qui seront diffusées par le GdR et l'Action MAESTRO.

- Vous pourrez aussi participer aux campagnes d'analyse de données astrophysiques et participer à la découverte des secrets de l'Univers.

Les étudiants, doctorants et pot-doctorants sont fortement encouragés à s'inscrire. Encouragez les, ils pourront bénéficier de financement, d'actions spécifiques pour leur mobilité.

## ATTENTION :

prochain Workshop : Mardi 5 et Mercredi 6 Juillet 2016 - IRIT - Toulouse

INSCRIVEZ VOUS : <https://maestro-2016.sciencesconf.org/>