



Écoles thématiques 2018 – Déclaration d'intention (en français)

Titre court (maximum 15 caractères) : ASTROInfo18

Titre complet de l'école (en majuscules) :

ECOLE FRANCAISE EN ASTRO-INFORMATIQUE

Responsable scientifique porteur du projet (un seul) :

Nom Prénom	Qualité	Code et intitulé du laboratoire	Adresse	Téléphone	Adresse électronique
SURACE Christian	IR1 CNRS	UMR 7326 Laboratoire d'Astrophysique de Marseille	38 rue Joliot Curie 13388 Marseille	0634410930	Christian.surace@lam .fr

Délégation régionale (DR) du CNRS organisatrice : DR12

Conseiller(e) formation de la DR en charge de l'école : Anne Marie Motte

Lieu pressenti : Laboratoire d'Astrophysique de Marseille

Date de début : 25/06/2018

Date de fin : 28/06/2018

Durée : 4 (jours)

S'agit-il d'une école récurrente ?

oui

non

Si oui, code de l'école de la session précédente :

Ou, à défaut : Titre de la session précédente :

Année de la session précédente :

Nombre de participants à l'école précédente : total :

dont CNRS :

PARTICIPANTS ATTENDUS :

Nombre total de participants prévus	Nombre de participants CNRS prévus
45	30

INSTITUT(S) SCIENTIFIQUE(S) DU CNRS CONCERNE(S) :

- Principal (1 seul) : INSU

- Secondaires (si nécessaire) : INS2I

SITUATION SCIENTIFIQUE (max une demi-page)

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. Par exemple, la mission spatiale Gaia (<http://gaia.obspm.fr>) 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.

Le projet LSST (Large Synoptic Survey Telescope <http://lsst.in2p3.fr/projet-LSST.html>), qui vise la construction d'un télescope de nouvelle génération propose la mise à la disposition d'un volume de données total d'environ 140 Petabytes à la fin de cette décennie. De manière générale, les instruments utilisés, 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. L'utilisation des masses de données implique de nouvelles méthodes de travail et de nouvelles techniques de traitement de données. La communauté française se structure avec des projets MASTODONS (PETASKY) une action nationale MAESTRO (masses de données en Astrophysique) (<http://www.madics.fr/actions/actions-en-cours/maestro/>) du GDR MADICS et une participation toujours plus grande en Europe (projet BigSkyEarth).

Cette école est dans la droite ligne de formation des étudiants ingénieurs et chercheurs dans les nouveaux concepts de traitement de grandes masses de données. Les responsables de ces actions et projets font partie du comité scientifique de l'école.

Cette école s'inscrit dans une politique de formation des chercheurs et ingénieurs pour une nouvelle méthodologie de traitement de données. Cette école a pour but de compléter efficacement les écoles comme Stat4Astro et ADA qui s'intéressent plus spécifiquement à l'aspect du traitement statistique des données

CONSÉQUENCES ATTENDUES (max une demi-page)

L'intérêt de cette école est de pouvoir initier, former et impliquer l'ensemble de la communauté INSU Astrophysique aux méthodes d'analyse dans les grandes masses de données. Les participants pourront être à même d'appréhender les nouvelles technologies concernant les grandes masses de données et les nouvelles méthodes pour les analyser. Ils pourront utiliser les différentes plateformes qui seront présentées. Le but est aussi de pouvoir définir, anticiper les nouveaux projets et monter des liens plus forts entre les communautés informatique et astrophysique afin d'assurer une meilleure cohésion interdisciplinaire. Nous pouvons lister comme objectifs attendus : l'utilisation de procédures d'analyse de données astrophysiques de grande masse, l'application à des plateformes de type Hadoop, l'utilisation de machine learning et deep learning.

GRANDS AXES DU PROGRAMME (max une demi-page)

Le programme est défini pour définir les meilleurs systèmes adaptés pour les différentes chaîne de traitement de données pour les grands projets futurs et les analyses de type machine learning. Ce sera une école dont le but est de mettre en avant la possibilité des données et toutes les analyses possibles de grandes masses de données. Les mots clés sont : Big Data, Machine Learning, Deep learning, Hadoop, plateforme GALACTICA

MODALITÉS PÉDAGOGIQUES (max une demi-page)

Nous avons choisi de mettre en place une méthodologie qui associe des cours d'introduction aux différentes méthodes, des cours de haut niveau et des moments d'échanges avec des Travaux pratiques s'appuyant sur des données présentes.

Les cours seront donnés par des professeurs qui resteront présents tout au long de la semaine afin d'optimiser les interactions avec les élèves.

Les journées « Cours » se dérouleront de la façon suivante : 1 heure d'initiation sur les concept de base du cours, 1h30 de cours, sur le sujet et 3 heures de Travaux pratiques.

L'école se déroulera sur 5 jours

1^{er} jour « Cours »

2^e jour « Cours »

3^e jour « Cours »

4^e jour Matin - ½ journée retours sur les principes et préparation Hackathon – présentation projets, faisabilité

4^e jour Après midi : ½ journée OFF –évènement social

5^e jour Matin Hackathon (3 projets) sur données spécifiques

5^e jour Après midi résultats Hackathon

L'ensemble de ces sessions seront filmées, broadcastées et mises à disposition pour un support de cours

JUSTIFICATION DU CHOIX DU FORMAT D'ECOLE THEMATIQUE (max une demi-page)

Le mode de l'école thématique est le cadre le mieux adapté pour communiquer et interagir. C'est en effet une occasion unique de conduire des chercheurs, étudiants, post-doc, ingénieurs vers des thématiques impliquant fortement l'interdisciplinarité. De plus, cette nouvelle discipline est fortement tournée vers l'expérimentation et par l'aspect « data driven ». Il est donc important de pouvoir manipuler les données et de pouvoir visualiser les résultats par des données. L'école thématique est idéal tant au point de vu formation qu'au point de vue de l'émergence de cette interdisciplinarité nécessaire et apporte une nouvelle vision des besoins et possibilités offertes par les machine learning, plateforme de calcul et d'analyse de grande masses de données pour les projets futur de la discipline

JUSTIFICATION DU LIEU PRESSENTI (max une demi-page)

Nous envisageons de réaliser cette école à Marseille. Plusieurs lieux sont envisagés. Le Laboratoire d'Astrophysique de Marseille est pour l'instant le lieu défini. Il est accessible facilement, possède un amphithéâtre de 100 places et un complément de plusieurs salles de réunions pour les splinter et préparation de hackathon. IL est situé dans un endroit ou il existe des possibilités d'accès à des chambres universitaires. Une grande cafétéria peut accueillir les buffets pour les déjeuner et les pauses café.

L'amphithéâtre du LAM est de plus équipé d'un système automatique de broadcast « live » et d'enregistrement de vidéos.

Document à renvoyer au Chargé de mission formation de l'institut principal et au conseiller formation de la délégation en charge de l'école