

# L'Action BigData4Astro



André Schaaff

*Centre de Données astronomiques de Strasbourg*

**Journées Annuelles OV France**

**22 et 23 mars 2021 en virtuel**



# □ Contexte, le GDR

- Extrait du site du GDR: *la création du GDR MaDICS s'inscrit dans un monde en réseau qui génère un déluge de données de plus en plus difficile à appréhender par les humains. Ce déluge nous a amené à «l'âge du zettaoctet» ( $10^{21}$  octets), dans lequel les données sont désormais au cœur de nombreuses activités humaines et une mine d'or pour la découverte scientifique et la réalisation de services innovants.*
- Des Ateliers (1 an) et des Actions (2 ans renouvelables)
- Règle: Au moins 1 membre INS2i + 1 membre autre Institut

## Action BigData4Astro (2021-2022)

*Big Data for Astronomy*

### Responsables

- André SCHAAFF, CDS, INSU
- Mohand-Saïd HACID, LIRIS, INS2i
- Karine ZEITOUNI, DAVID, INS2i

Correspondant ComDIR : Sarah Cohen-Boulakia



MaDICS est un lieu d'échange interdisciplinaire



# □ Historique

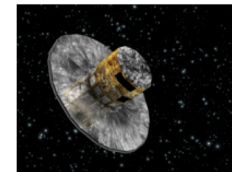
- Des **fondations** solides construites par l'Action MaDICS MAESTRO (MAsses de données En aSTRONomie et astrophysique portée par Christian Surace(LAM) et Sofiane Maabout (LaBRI) jusqu'en 2019.
- En 2020, une participation en tant qu'Atelier aux événements du GDR (Assemblée des Responsables d'Ateliers et d'actions, second Symposium) mais des ambitions limitées par la situation sanitaire.
- *En 2020, un second Atelier, OSIS "Outils Statistiques pour l'Imagerie hyperSpectrale du milieu interstellaire" porté par Emeric Bron et Antoine Roueff enrichissait également la participation de l'astronomie à MaDICS. L'Action BigData4Astro élargira sa thématique initiale pour donner une suite à cet Atelier.*

# □ Problématiques scientifiques

- **Gestion, analyse et exploration** des données scientifiques dans le domaine de la cosmologie et l'astrophysique
  - Grosses masses de données
  - Données complexes : images, données incertaines, multi-échelles...
  - Formats hétérogènes
- **Traitements variés et complexes** (analyse d'images, reconstruction de trajectoires, traitements ad-hoc...)
- **Verrous scientifiques**
  - Nettoyage de données
  - Passage à l'échelle
  - Intégration des données
  - Enrichissement de données
  - Analyse de données complexes
  - Visualisation de grandes masses de données
- Contextes applicatifs : LSST, SDSS, Gaia, Euclid, SKA ...



Vue d'artiste de SKA  
Crédit: SKA Organisation



Vue d'artiste de Gaia  
Crédit: ESA



Vue d'oiseau du Very Large Telescope  
Crédit: J.L. Dauvergne & G. Hädepohl (<http://www.eso.org>) / ESO



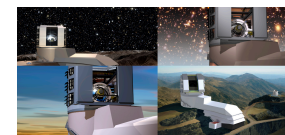
Hubble, le télescope spatial  
Crédit: NASA, 2002



XMM-Newton  
Crédit: Image courtesy of ESA



Sous le charme des Nuages de Magellan, ALMA  
Crédit: ESO/C. Malin



Vera C. Rubin  
Crédit: LSST  
Corporation

# □ La Mission de l'Action

- **Mobiliser** une communauté **interdisciplinaire** autour de ces **problématiques...**
- ... pour faire émerger des **concepts**, des **modèles** et des **méthodes** originales permettant une meilleure compréhension des traitements à réaliser sur les données astronomiques et une plus grande maîtrise des **logiciels** associés tant dans leur développement que dans leur maintenance et leur évolution.
- L'Action s'appuie sur cette **collaboration** entre **astrophysiciens** qui génèrent et exploitent des données et **informaticiens** dont l'expertise relève de la gestion de données massives et la visualisation.



# □ Feuille de route et perspectives

- **Elaboration** d'une **Cartographie** au niveau national des forces et compétences, et des moyens logiciels et techniques.
- **Morcellement** des **Centres de données**: encourager une mutualisation pour être à la hauteur des enjeux des grands projets (SKA, ...) au risque sinon de devenir des utilisateurs payants d'infrastructures de type Google ou Amazon.
- **Une Difficulté** de **recrutement** de doctorants sur des profils particuliers (ni complètement astro ni info): identifier les freins.
- Une **identification** régulière des besoins en **compétences** et en **formation**: une **participation** proactive à des actions de formation => ANF AstroInfo 2021
- Être une force de proposition => une Action Spécifique Numérique INSU
- Société: médiation scientifique envers les jeunes (le double attrait pour le numérique et l'astronomie)

# □ En 2020

- Nous avons participé au Second Symposium du GDR en animant une demi-journée

## Programme de l'Atelier BigData4Astro

Mercredi 8 juillet 9h00-12h30

9h00 *Introduction*

9h15 **Big Data astronomy with Spark**, *Stéphane Plaszczyński (IJCLab, Université Paris-Saclay)* 🇫🇷

9h50 **Deep Learning for the next generation of big-data surveys**, *Marc Huertas-Company (Observatoire de Paris/LERMA, Université Paris-Diderot)*

10h15 **Big FAIR Data in Astronomy**, *Mark Allen, Observatoire astronomique de Strasbourg*

10h50 *Pause*

11h00 **Spatial Query Optimization for Big Astronomical Surveys – ASTROIDE project**, *Karine Zeitouni (Université de Versailles-Saint-Quentin/DAVID)*

11h20 **Table Ronde : Perspectives et verrous scientifiques**, avec les invités et les animateurs de l'Atelier

12h15 **Conclusion**

Rendez-vous en juillet 2021  
(semaine du 5) pour le 3<sup>ème</sup>  
Symposium !

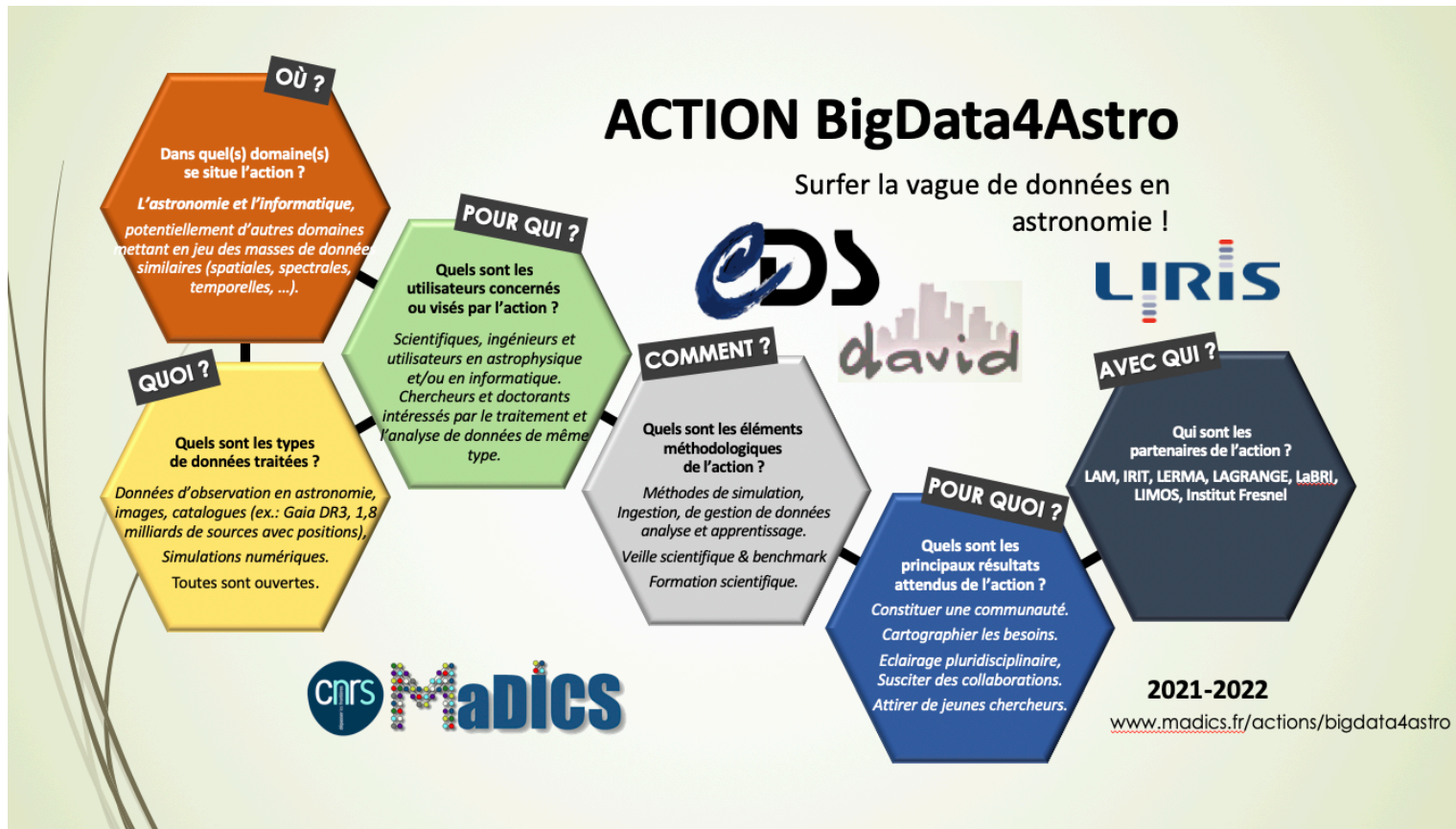
# □ Les évènements en 2021

- Participation à l'Assemblée des Responsables d'Ateliers et d'Actions le 12 mars
- Animation d'une demi-journée du 3<sup>ème</sup> **Symposium MaDICS** la semaine du 5 juillet
- **Un workshop de 2 jours autour de l'été + webinaires**
- Participation active à l'école **AstroInfo 2021**
- Participation en fin d'année à la **prospectivité du GDR**
- Collaboration éventuelle avec d'autres Actions



# □ Communication

BigData4Astro est sur Slack, nouvelles sur les listes déjà existantes comme celle d'OV France



Le GDR fera des actions de communication et de promotion