



**DATA**  
**TERRA**

# Ecole Thématique DATA SDUE

**Guide de Survie dans la jungle des données  
en Sciences de l'Univers et de l'Environnement (SDUE) :  
Comment gérer les données pour les valoriser?**

**« Un exemple en Astronomie/ Astrophysique :  
le CDS »**

Françoise Genova

# Le CDS

---

Françoise Genova, Mark Allen et l'équipe  
CDS



# □ La mission du CDS

- Créé **en 1972** par l'INAG (devenu l'INSU/CNRS) avec l'Université Louis Pasteur (devenue l'Université de Strasbourg)
- Depuis 1972
  - Collecter les données utiles sur les objets astronomiques, sous forme électronique
  - Les améliorer en les évaluant de façon critique et en les combinant
  - Distribuer les résultats à la communauté internationale
  - Conduire des recherches en utilisant les données
- Les mots-clé
  - **Données sous forme électronique**
  - **Curation et qualité des données**
  - **Rôle international**
  - **Objectif: Servir la recherche**



# Le CDS est une infrastructure de recherche



- Une gouvernance identifiée, effective
- Une instance de pilotage stratégique et scientifique, le Conseil Scientifique
  - International depuis l'origine
  - Représentants de l'ESA, de l'ESO et de la NASA
- Un des piliers de la pratique internationale de la science ouverte en astronomie
- Sur la Feuille de Route nationale depuis sa première version (2008)
- Infrastructure de Service aux Données (CoSIN)
- Centre de Référence Thématique (Recherche Data Gouv)



# □ Le CDS est et a été pionnier

- Un des pionniers absolus des centres de données scientifiques, toute disciplines confondues
  - ICPSR, Sciences Politiques et Sociales, 1962
- Pionnier du partage des données numériques en astronomie avec la base de données du satellite IUE (1985)
- L'astronomie est elle-même une discipline pionnière de la Science Ouverte – le partage et la réutilisation des données sont des besoins scientifiques essentiels
  - Des objets nombreux et divers -> statistiques
  - Utilisation conjointe de données obtenues par des instruments différents pour comprendre les phénomènes physiques à l'œuvre dans les objets – multi-longueur d'ondes/multi-messagers
  - Variabilités temporelles à différentes échelles



# □ Qui sommes-nous?

*Une équipe intégrée qui regroupe les différents profils des métiers de la donnée, en première ligne par rapport à l'évolution nécessaire de l'évaluation des chercheurs et des personnels techniques dans le contexte de la Science Ouverte*

- **Astronomes**
  - Missions : recherche, enseignement, tâches de services
  - Diversité d'expertise scientifique
  - Contributions au sein du CDS et de la communauté
- **Documentalistes**
  - Polyvalence : astronomie, informatique, information scientifique
  - Chaîne documentaire complexe -> données à forte valeur ajoutée
- **Informaticiens**
  - Découpage par services indépendants
  - Multi-casquettes : développeur, architecte software, bases de données, ingénieur système, support



# □ Les impératifs catégoriques

- Servir la recherche
  - Prendre en compte les besoins des scientifiques
    - Qualité et pertinence des services
    - Qualité et pertinence des données
  - Au meilleur niveau technologique
- Penser sur la longue durée dans un contexte qui évolue constamment
  - L'astronomie évolue – progrès scientifique, nouveaux instruments
  - La technologie évolue – la révolution du web! mais en fait évolution continue qui peut être très rapide
  - Le contexte 'politique' évolue – la science ouverte, FAIR, l'EOSC



From “Et Dieu Créa l’Internet”  
Christian Huitema (1995)



ensuite de commander la liaison à France Télécom et à la société américaine de communications MCI : le 28 juillet 1988, la France était raccordée à l’Internet.

Néanmoins, le 8 août, tout était prêt. Nous fîmes une belle démonstration, nos partenaires de la NASA purent accéder à la base de données qui, à Strasbourg, récapitulait des siècles d’observations astronomiques.

*40<sup>ème</sup> anniversaire du CDS*

<https://interstices.info/au-coeur-de-la-premiere-connexion-francaise-a-larpanet/>

# □ Continuer et progresser sur la (longue) durée

- Définir une stratégie explicite et évolutive qui prenne en compte les différentes facettes du contexte
- Créer, réunir et conserver les compétences et les faire évoluer
- Construire et tenir sa place dans le concert des nations
  - Fourniture de services essentiels
  - Réseau de collaboration avec les acteurs majeurs (agences, journaux, ADS)
  - Participation forte à la définition des standards
- Obtenir le soutien sur le long terme de la gouvernance



# □ R&D

- Le Centre de Données doit travailler sur la longue durée avec des contraintes opérationnelles fortes
- La technologie peut évoluer très vite dans le domaine
  - Des buzz sans lendemain
- Des solutions ‘suffisamment pérennes’ pour améliorer la manière de répondre aux besoins et aux attentes des utilisateurs
- Veille technologique et R&D à visées opérationnelles
  - Nombreux stages pour tester des solutions possibles
  - Décision explicite d’implémentation après évaluation



# Le CDS dans l'infrastructure globale des données en astronomie (1)

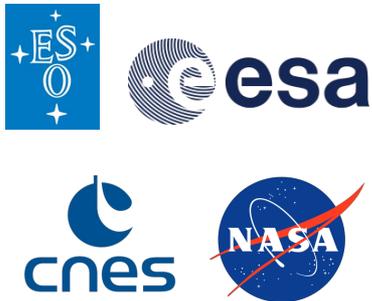
- Les données en astronomie
  - Sont FAIR - format FITS (**R**, 1981) et standards de l'Observatoire Virtuel (OV) Astronomique (**F, A, I**) définis par l'IVOA depuis 2002
  - **Vu de l'utilisateur: une seule infrastructure de données mondiale, multipolaire**
- Le rôle du CDS
  - Fourniture de services essentiels inclus dans l'Observatoire Virtuel
    - Plus de 2 millions de requêtes/jour
    - Augmente la visibilité et l'utilisabilité des services
  - Participation active à la définition des standards
    - 48 standards IVOA, 24 avec auteur(s) ObAS (dont éditeurs: 13)
    - Le CDS avait développé plusieurs précurseurs de l'OV
    - Pilotage du 1<sup>er</sup> Groupe de Travail international *Interoperability*, pré-IVOA, dans le cadre du Réseau OPTICON (2000-2002) ⇒ 1<sup>er</sup> standard de l'OV en 2002
  - En Europe, coordination de projets ou de WP
    - Soutien à l'OV dans des projets financés par la CE
    - L'OV est inclus dans les projets Cluster qui regroupent les ESFRI
    - Interfaçage de l'OV avec l'EOSC
  - En France
    - Coordination de l'Action Spécifique OV France de 2004 à 2020
    - Participation active à la diffusion des connaissances et au partage d'expérience



# Le CDS dans l'infrastructure globale des données en astronomie (2)



Ground and Space  
Observatories, Instruments  
and missions



Journals



+ ...

Astronomy Data Centres



Virtual Observatory



ASOV

H2020 projects:



Global data initiatives



# □ Les services du CDS – données de référence



**La base de référence pour les objets astronomiques:**  
*~13 millions d'objets, ~52 millions identifiants, ~28 millions citations, ~15000 acronymes*



**Le service de référence des catalogues astronomiques:**  
*>22000 catalogues, 48000+ tables publiées, ~51 milliards de lignes*  
**> Données associées aux publications**



**Visualisation et base de données d'images:**  
*1000+ relevés HiPS: images et cubes (>408 TB)*  
*Client Aladin, Aladin Lite – dans les applications web*

**+ Services de cross-identification, CDS Portal, etc**  
**Interopérables grâce aux standards (IVOA, CDS, +)**

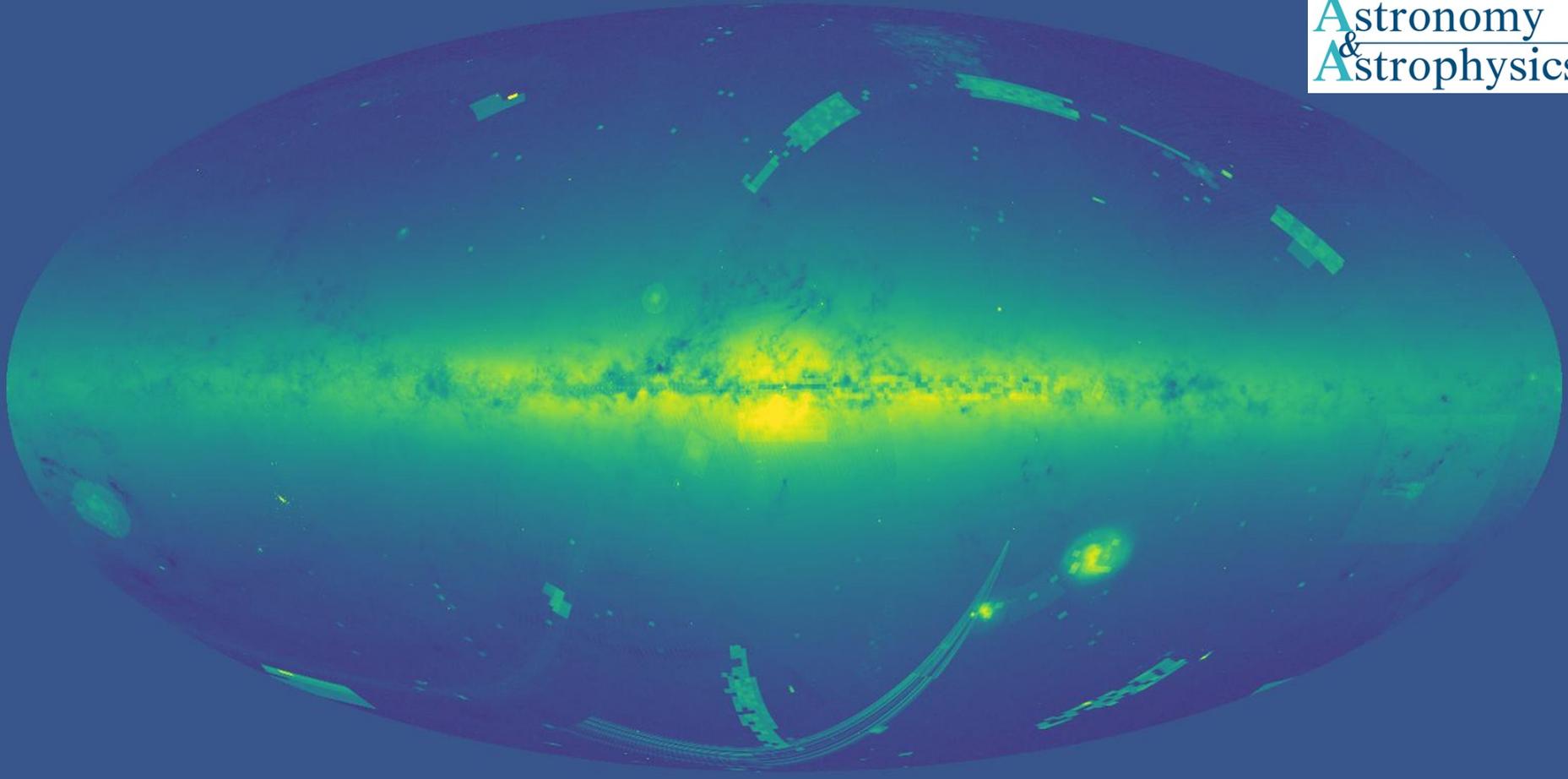


Valeur ajoutée: des données hétérogènes de qualité dans un système homogène

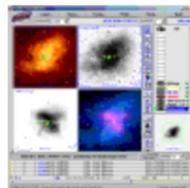
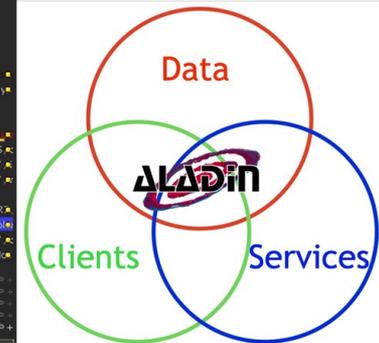
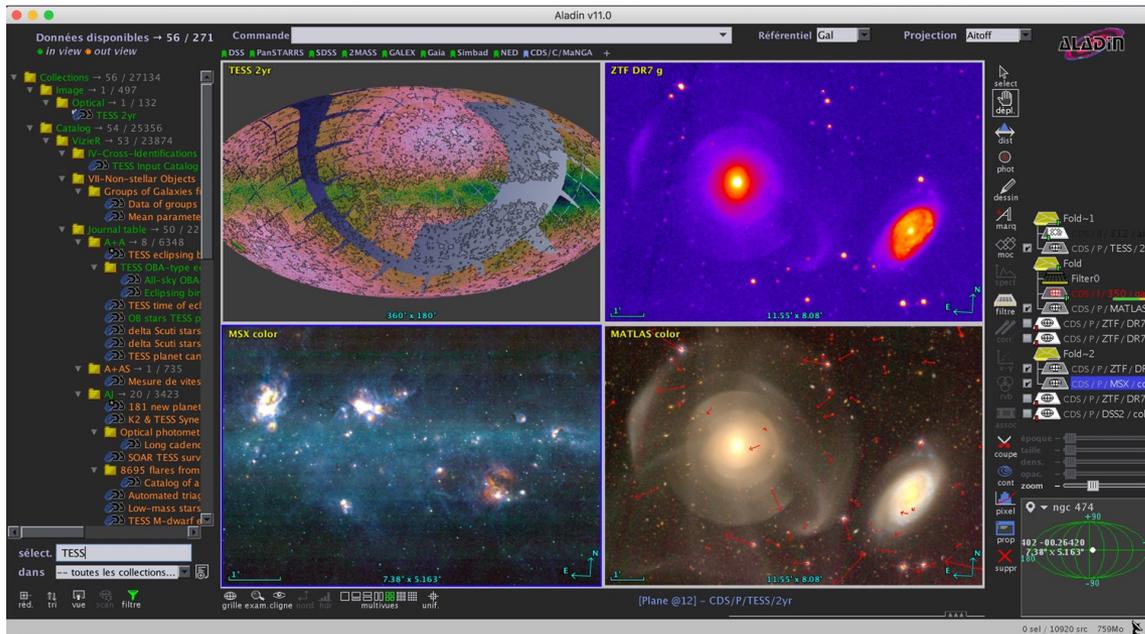


Ex: Tables attachées aux publications dans une carte de densité

Astronomy  
&  
Astrophysics



# L'interopérabilité: Aladin portail de l'Observatoire Virtuel astronomique



Download  
**Aladin Desktop**  
on your machine



Preview with  
**Aladin Lite**  
in your browser



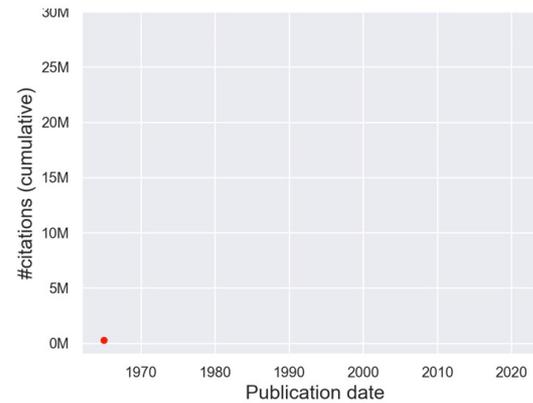
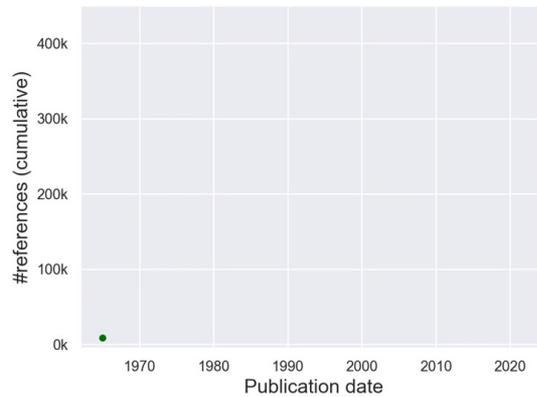
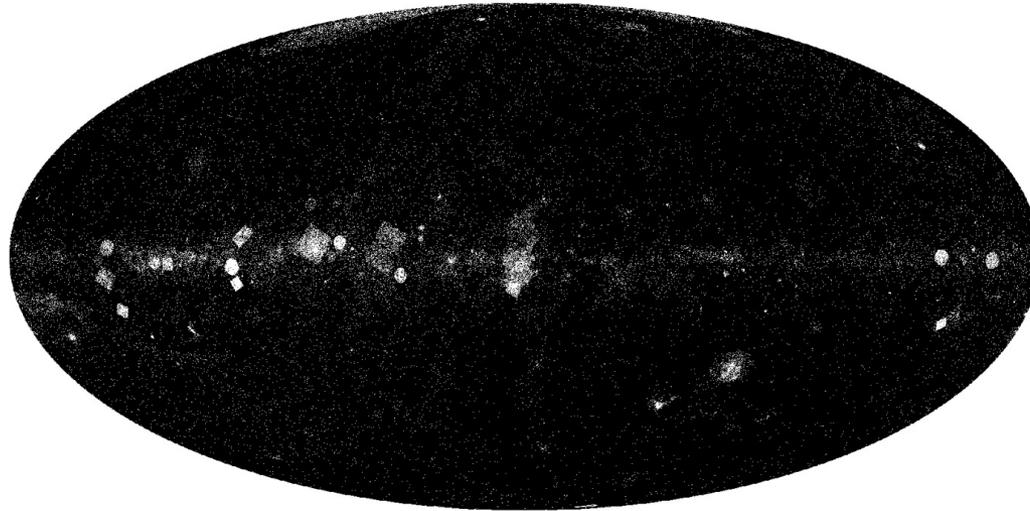
# □ Sur la durée

- Une charge de travail très lourde, qui continue à augmenter
- La confiance des utilisateurs
  - Pertinence des services par rapport aux besoins des utilisateurs
  - Souci constant de la qualité du contenu et des services
- La veille technologique et la R&D comme partie intégrante du travail du CDS
- Le réseau de collaboration internationale et le choix de construire sur la collaboration même s'il y a des risques (cf client-serveur, interopérabilité)
- Le soutien du CNRS/INSU et de l'Université
- Le risque que tout le monde pense que tout va bien au CDS alors qu'il relève constamment des défis complexes
- L'équipe CDS



# La construction du contenu: SIMBAD

#references until 1965



# □ Un jour normal au CDS

Aladin  
Simbad  
VizieR

2022-02-10 00:00:00 UTC+01

