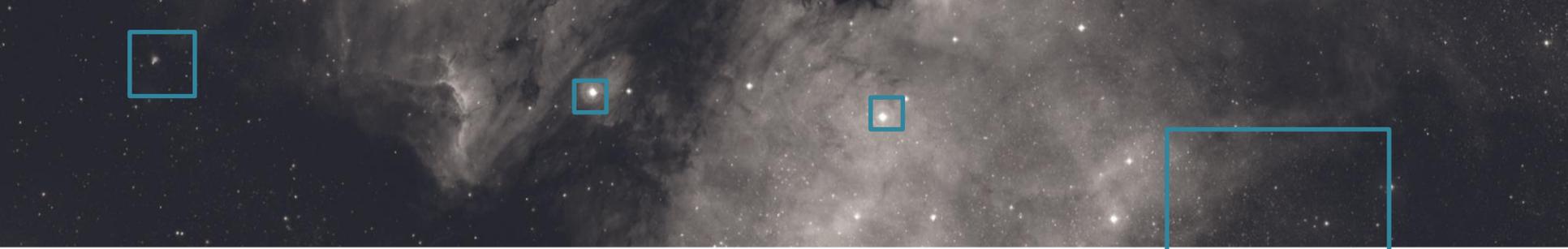


Extension à l'international - Le Centre de Données astronomiques de Strasbourg et la Research Data Alliance

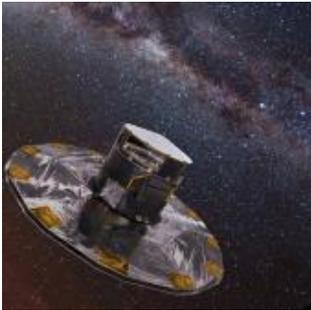
Françoise Genova, Observatoire
Astronomique de Strasbourg et RDA France





LE CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG (CDS)

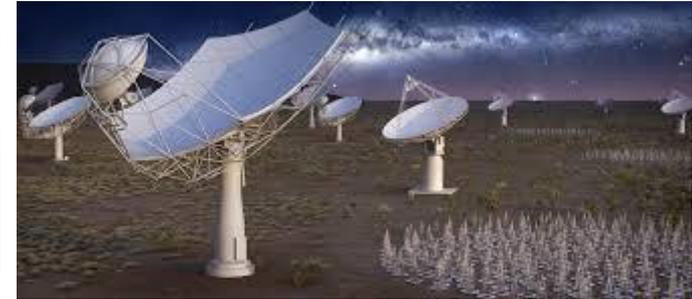
□ Les infrastructures de recherche en astronomie



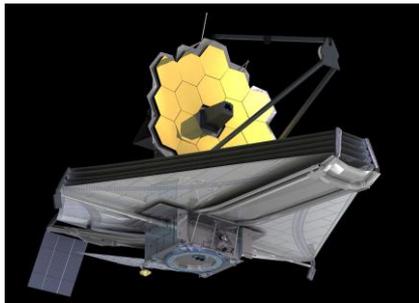
Gaia



ELT



SKA



JWST



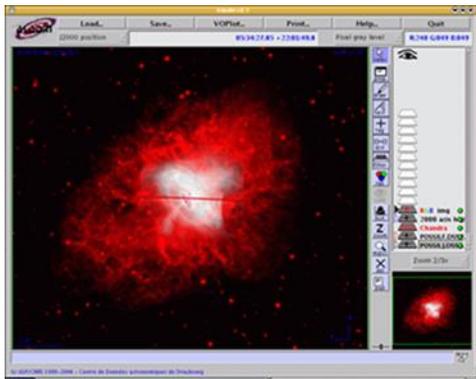
Planck

CTA



+ LES DONNEES!

□ Pourquoi partager les données en astronomie?



- Les données sont au coeur de l'approche scientifique de la discipline
 - Astronomie multi-longueur d'onde/multi-messenger (ondes gravitationnelles, etc.)
Combiner les observations de différents instruments pour comprendre les phénomènes physiques à l'oeuvre dans les objets
 - Variabilité sur des échelles de temps très différentes
 - Comparaison des modèles avec les observations
- Optimisation du retour scientifique d'instruments coûteux
- Une discipline pionnière du partage des données scientifiques, bien avant les préconisations de la Science Ouverte

□ La mission du CDS

- Créé **en 1972** par l'INAG (devenu l'INSU/CNRS) avec l'Université Louis Pasteur (devenue l'Université de Strasbourg)
- La mission du CDS 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-clés **depuis 1972**
 - **Données sous forme électronique**
 - **Curation et qualité des données**
 - **Rôle international**
 - **Objectif: Servir la recherche – Pilotage par les besoins scientifiques**
- Un des pionniers absolus des centres de données scientifiques, toute disciplines confondues
- Rôle précurseur en astronomie

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
- Dans le paysage français
 - Sur la Feuille de Route nationale depuis sa première version (2008)
 - Infrastructure de Service aux Données (CoSIN)
 - Centre de Référence Thématique (Recherche Data Gouv)



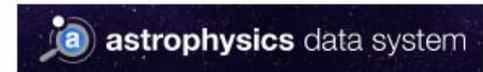
CDS - a part of the global astronomy data infrastructure

Conseil du CDS 2023, M. Allen

Connections to the Observatories and Space Agencies

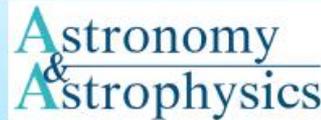


Collaboration with other Astronomy Data Centres



- Harvard Smithsonian ADS
- NASA Extragalactic Database

Astrophysics Journals



+ ...

Building the Data Sharing framework of Astronomy: *The Virtual Observatory*



Certified:



Networks:



recherche.data.gouv.fr



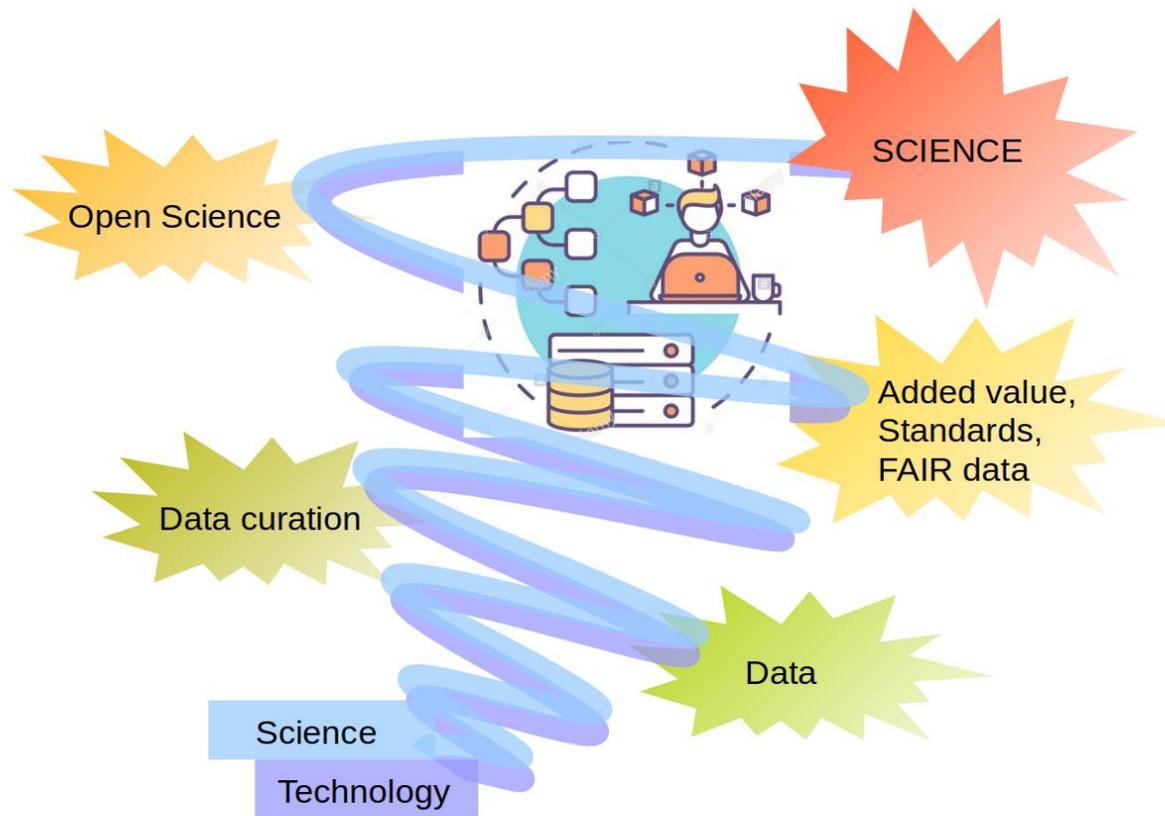
□ Les pionniers doivent innover sur la longue durée!

- Le CDS a inventé des services largement utilisés par la communauté astronomique internationale et ses infrastructures de recherche
 - 3 millions de requêtes/jour en 2023
- Il les fait évoluer pour répondre aux besoins scientifiques et aux évolutions technologiques
- Rôle significatif dans la définition des standards d'échange et d'interopérabilité des données
- Une équipe intégrée qui regroupe les différents profils des métiers de la donnée: chercheurs, documentalistes (data stewards/curateurs), ingénieurs informaticiens
 - 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

□ Qui sommes-nous?

- **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

Leur travail vu par les documentalistes du CDS



□ 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! – en fait évolution continue qui peut être très rapide
 - Le contexte ‘politique’ évolue – la science ouverte, FAIR, l’EOSC, Recherche Data Gouv

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^{ème} anniversaire du CDS

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

□ Le cœur de la mission du CDS

- Des services pour des données ***de référence*** en soutien à la Science Ouverte
- ***Stewardship*** des données liées à des publications référées
- ***Fourniture de données et de services FAIR*** via les standards internationaux pertinents
- ***Innovation*** pour la prochaine génération de services de référence

□ Continuer et progresser sur la (longue) durée

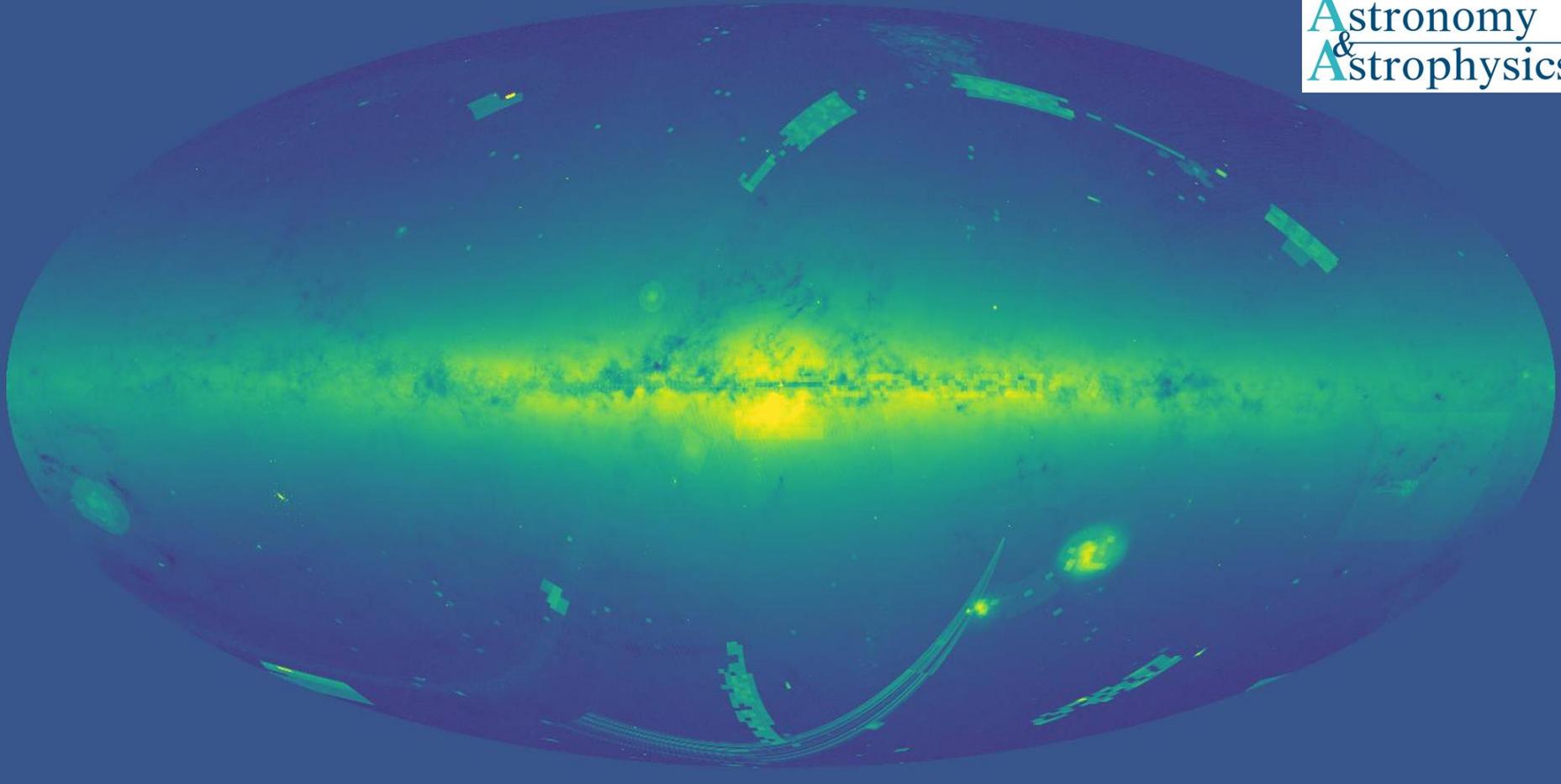
- Définir une stratégie explicite et évolutive qui prenne en compte les différentes facettes du contexte
- Créer, réunir et conserver les compétences et les faire évoluer
- Construire et tenir sa place dans le concert des nations
 - Fourniture de services essentiels
 - Réseau de collaboration avec les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

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



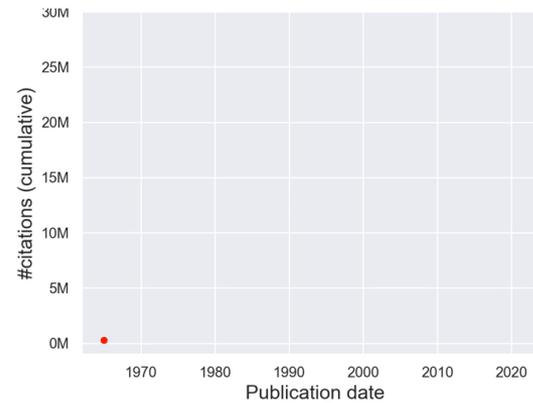
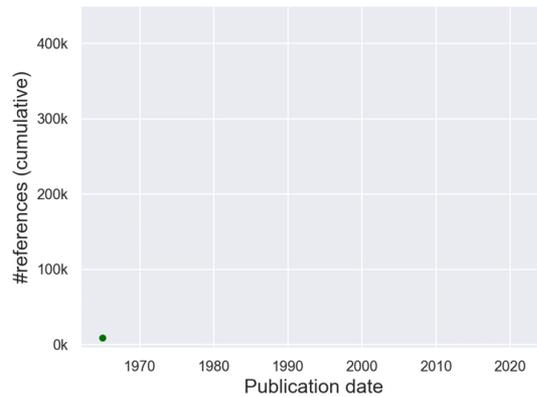
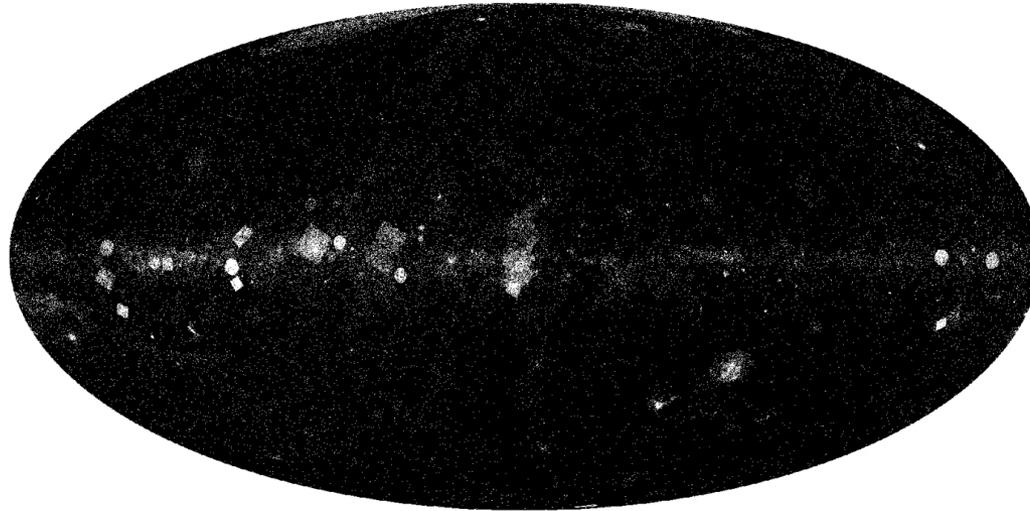
Ex: Tables attachées aux publications dans une carte de densité

Astronomy
&
Astrophysics



La construction du contenu: SIMBAD

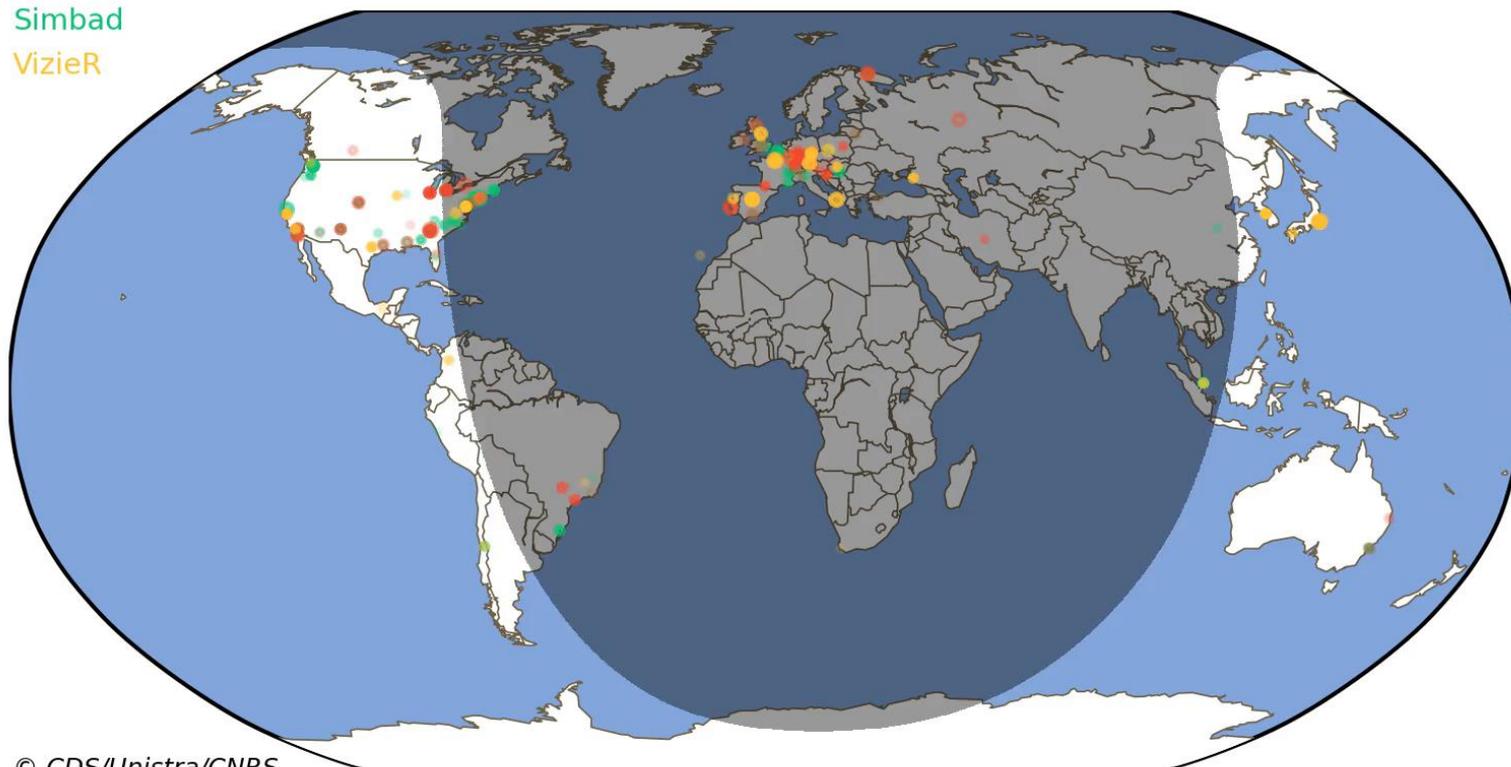
#references until 1965



□ Un jour normal au CDS

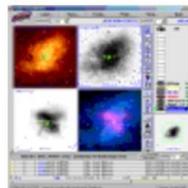
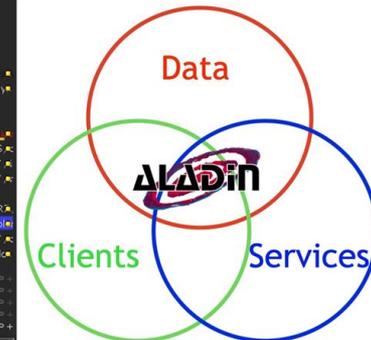
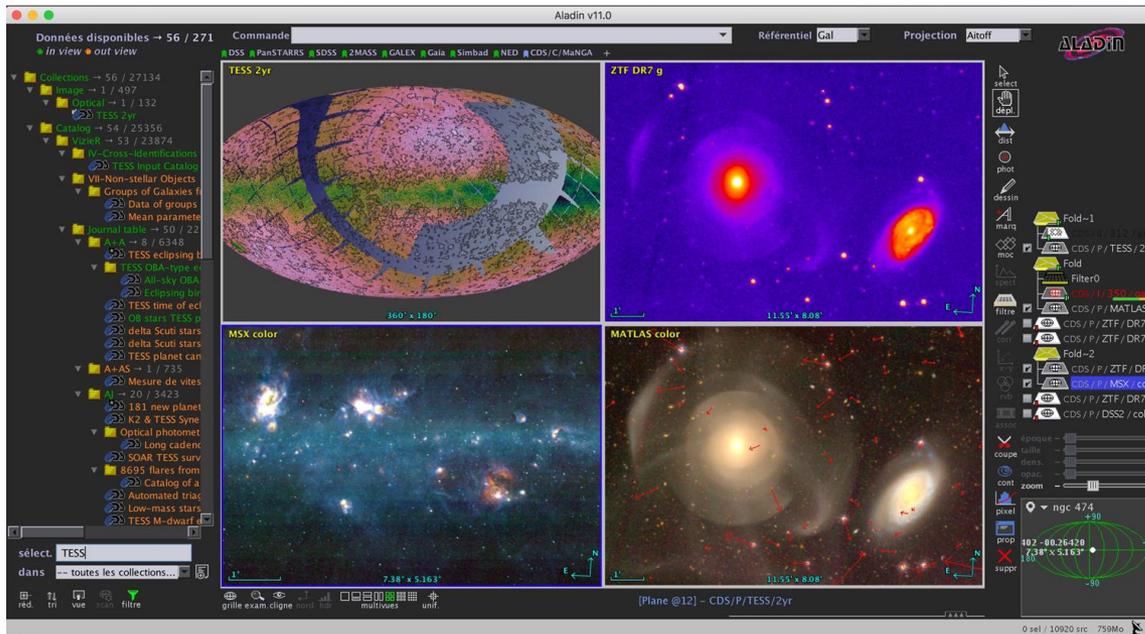
Aladin
Simbad
VizieR

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



© CDS/Unistra/CNRS

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



Download
Aladin Desktop
on your machine



Preview with
Aladin Lite
in your browser

□ Les standards jouent un rôle essentiel pour le partage des données

- Point de départ: le format FITS (1981)
 - Tout astronome peut utiliser les données de tout télescope
 - On peut construire des outils pour utiliser les données
 - FITS inclut les données et les métadonnées sur la provenance des données: les données sont **Réutilisables**
- 1989 – Accord entre services de données sur un identifiant pour les publications (bien avant la mise en œuvre du DOI!)
- A partir de 1993 – Mise en réseau des ressources sur le web
- Les standards pour trouver (**F**aciles à trouver/**F**ind) les données, y **A**ccéder et les faire **I**nteropérer sont développés et maintenus par l'International Virtual Observatory Alliance (IVOA) depuis 2002
 - Une fine couche au dessus des collections de données pour rendre possible F, A et I
 - Un répertoire (*registry*) des ressources disponibles
 - Les données et les outils qui permettent de les utiliser sont interopérables

INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE

The Virtual Observatory (VO) is the vision that astronomical datasets and other resources should work as a seamless whole. Many projects and data centres worldwide are working towards this goal. The International Virtual Observatory Alliance (IVOA) is an organisation that debates and agrees the technical standards that are needed to make the VO possible. It also acts as a focus for VO aspirations, a framework for discussing and sharing VO ideas and technology, and body for promoting and publicising the VO.

To learn more about the IVOA as an organisation, read the **"About"** section.

To learn more about the VO from a user's point of view, including how to find VO tools and services, read the **"Astronomers"** section. There is also a page about the **VO for students and the public**.

To learn how to publish VO services, or write VO-compatible software, start by reading the **"Deployers/Developers"** section.

Internal IVOA discussions are publicly viewable in the **"Members"** section.



IVOA NEWS

March 2022 Issue of the IVOA Newsletter

UPCOMING MEETINGS

IVOA May 2024 Interoperability Meeting
19-24 May 2024 Sydney (NSW, AUstralia)

For Astronomers



Getting Started / Using the VO
VO Glossary / VO Applications
IVOA newsletter / VO for Students
& Public



For Deployers/Developers



Intro to VO Concepts /
IVOA Standards / Guide to
Publishing in the VO / Technical
Glossary



For Members



IVOA Calendar / Working Groups/
Twiki / Documents in Progress /
Mailing Lists / IVOA Roadmap

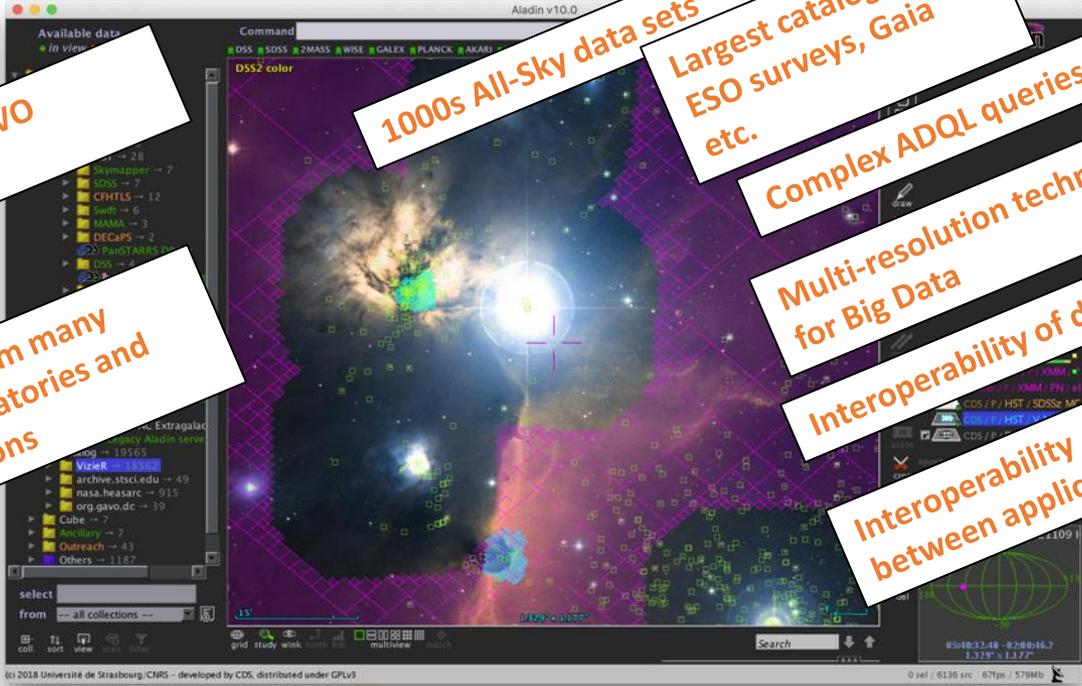


<https://www.ivoa.net/>

□ L'infrastructure des données en astronomie

- Vu par les utilisateurs: une infrastructure des données unique, mondiale, multipolaire
- L'infrastructure est ouverte: tout le monde peut partager des données ou développer un outil permettant d'accéder aux données
- L'infrastructure permet un accès transparent aux données et est invisible des utilisateurs
- Les données et les outils sont interopérables

One view of the VO from an application/portal :



Available data
in view

- SkyMapper → 7
- SDSS → 7
- CFHTLS → 12
- Swift → 6
- IRAS → 3
- DECaPS → 2
- PanSTARRS → 1
- DSS → 4

Command
DSS SDSS ZMASS WISE GALEX PLANCK AKARI
DSS2 color

select
from -- all collections --

© 2018 Université de Strasbourg/CNRS – developed by CDS, distributed under GPLv3

0 sel / 6136 src. 67pps / 578Mb

Built from VO Registry

Data from many observatories and missions

1000s All-Sky data sets

Largest catalogues: ESO surveys, Gaia etc.

Complex ADQL queries etc.

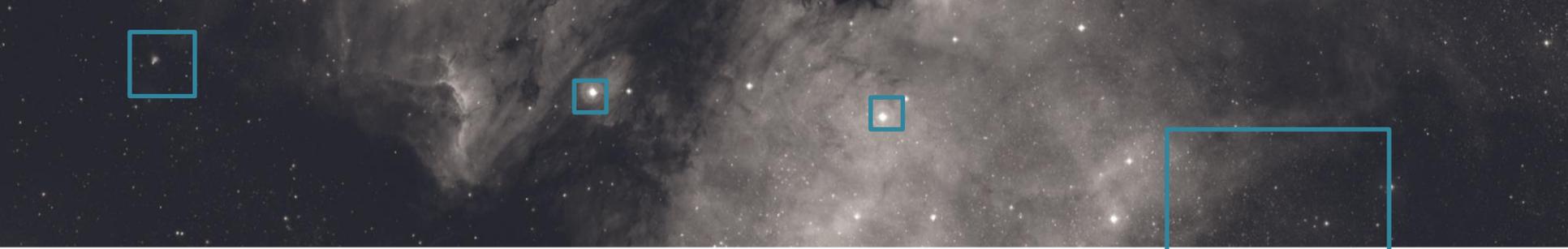
Multi-resolution techniques for Big Data

Interoperability of data

Interoperability between applications

Diapositive de Mark Allen, CDS





LA RESEARCH DATA ALLIANCE (RDA)

□ Le partage des données scientifiques

- Un élément essentiel de la Science Ouverte
- Certaines disciplines le pratiquent depuis longtemps (par ex. l'astronomie), d'autres commencent à y réfléchir, certaines sont en route
- De nombreux aspects, technologiques mais pas seulement!
- Des personnes avec des profils différents sont impliquées
- Il faut impliquer les disciplines scientifiques comme « piliers » de la science ouverte
- De nombreux aspects en commun

□ Panel interdisciplinaire à l'International Data Week 2016 (Denver)

- Astronomie/sciences de la terre/sciences humaines/ sciences des matériaux/cristallographie
- Les points communs
 - **Pilotage par les besoins scientifiques**
 - Définir la partie disciplinaire des standards d'interopérabilité est difficile mais indispensable
 - Il faut partager les données, mais aussi les outils qui permettent d'utiliser ces données
 - Incitations/motivations
 - Les aspects sociologiques sont souvent plus importants que les aspects techniques
 - Beaucoup des méthodes et des techniques peuvent être partagées
- Les systèmes de gouvernance sont très divers: ils dépendent de la discipline, de la culture de celle-ci et de la manière dont elle est organisée – chaque discipline a ses propres pratiques

[Genova et al. Data Science Journal 2017](#)

□ La Research Data Alliance



- Fondée en mars 2013 par le gouvernement australien, la Commission Européenne et la NSF (USA)
- 14 523 membres de 151 pays (le 6 avril)
- Travail « bottom-up » sur tous les aspects du partage des données scientifiques, technologiques et sociologiques
<http://rd-alliance.org>
- Forum de discussion neutre et international
- Recommandations et « produits »



Vision

Les acteurs de la **recherche** et de **l'innovation** partagent ouvertement les données **et les ré-utilisent**, quels que soient les technologies, les disciplines et les pays, afin de répondre aux grands défis de société.

Mission

La RDA construit **des passerelles à la fois sociales et techniques** pour permettre **le partage ouvert et la réutilisation des données**.

LA RESEARCH DATA ALLIANCE

www.rd-alliance.org

Construire les passerelles sociales et techniques permettant le libre partage et la ré-utilisation des données

72 REALISATIONS PHARES

dont 8 Spécifications
Techniques ICT

200+ CAS D'UTILISATION

dans différents domaines,
organisations et pays

89 GROUPES QUI TRAVAILLENT SUR L'INTEROPERABILITE GLOBALE DES DONNEES

35 Groupes de Travail
53 Groupes d'Intérêt
1 Communauté de Pratique

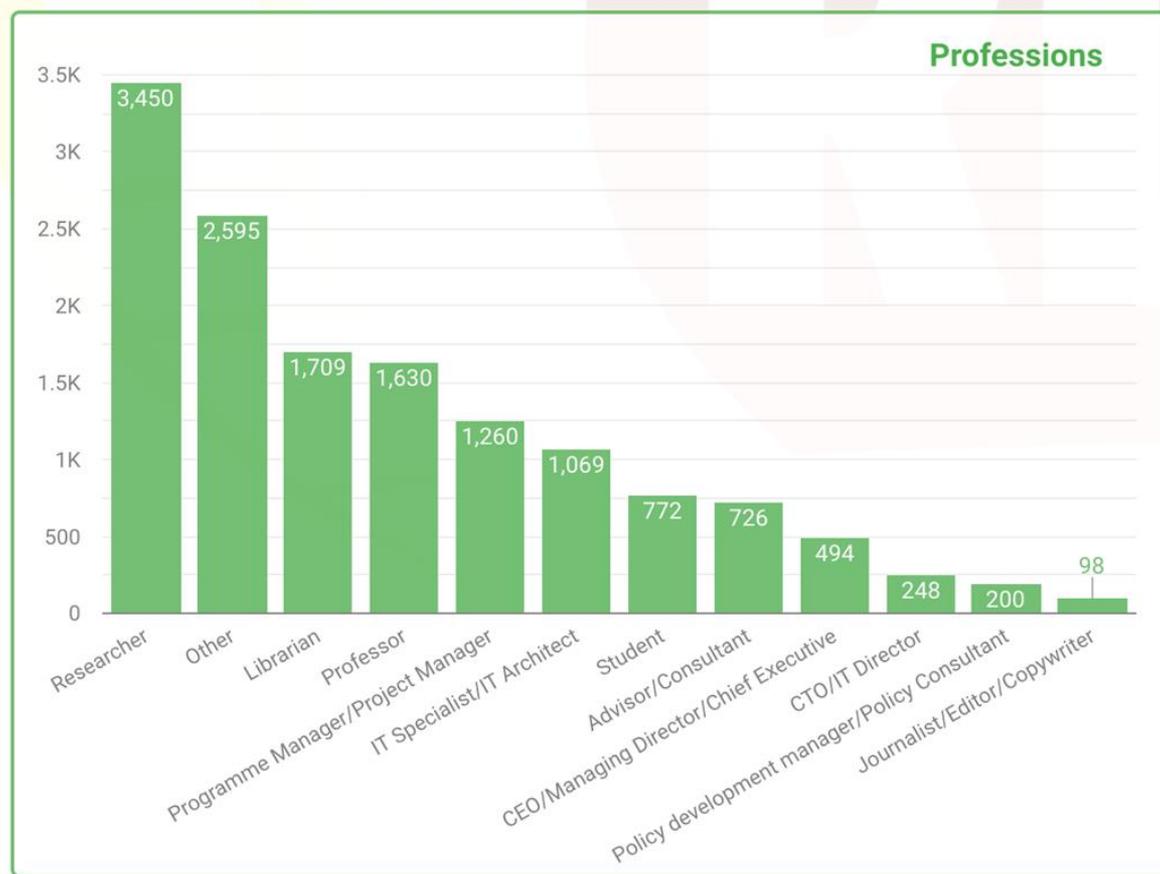
14 251 MEMBRES INDIVIDUELS DE 151 PAYS

70% recherche et milieu académique
14% administration publique
11% entreprise et industrie

67 MEMBRES ORGANISATIONNELS
13 MEMBRES AFFILIES



Mot clé: Diversité - Les profils des participants



Février 2024

□ Mot clé: Diversité - Les thèmes abordés

- Découverte des données
- Accès aux données et interopérabilité
- Préservation et “stewardship”/curation des données
- Provenance et réutilisation
- Passage à l'échelle
- Culture du partage des données
- Education et formation
- Gérer et maintenir les infrastructures de données
- Sécurité, aspects légaux
- etc

□ Exemples de recommandations

- Parmi les dernières endossées par le Conseil
 - FAIR Principles for Research Software
 - Metadata Schema for the Persistent Identification of Instruments
 - Guidelines for publishing structured metadata on the web
 - Interoperable Descriptions of Observable Property Terminologies (I-ADOPT) WG Outputs and Recommendations
 - 10 Things for Curating Reproducible and FAIR Research
 - Recommendations and Capacity Development Resource Kit
 - Guidelines for raising FAIR Adoption in Health Data and Health-related Research Performing Organisations (HRPOs)
 - National PID Strategies Guide and Checklist
- Parmi les recommandations plus anciennes
 - Repository Audit and Certification DSA–WDS Partnership WG Recommendations



□ RDA France

- 4 projets financés par la Commission Européenne entre 2012 et 2020 en soutien à la RDA
- Le CNRS a participé aux trois premiers projets
- Le 4^{ème} projet, RDA Europe 4.0, a mis en place des Nœuds Nationaux à partir du 1^{er} mars 2018
 - Nœud National RDA France
- RDA France: Françoise Genova (UMR 7550) & Francis André, puis Laurence El Khouri (DIST/DDOR)
- Support de la DIST puis de la DDOR du CNRS
- Une priorité depuis l'origine: la certification des entrepôts de données
 - Groupe de Travail commun avec le Collège Données du CoSO
<https://www.ouvrirlascience.fr/certification-des-entrepots-et-services-de-donnees/>

Le Plan National pour la Science Ouverte – PNSO1

Axe Structurer et Ouvrir les Données de la Recherche

Structurer

- Généraliser la mise en place de plans de gestion des données dans les appels à projets de recherche
- Développer des centres de données thématiques et disciplinaires.
- Développer un service générique d'accueil et de diffusion des données simples.
- Engager un processus de certification des infrastructures de données. 

Organiser

- Soutenir la *Research data alliance* (RDA) et créer le chapitre français de l'alliance (RDA France). 
- Soutenir *Software heritage*, la bibliothèque des codes sources

□ Le Plan National pour la Science Ouverte – PNSO2

Axe Structurer, partager et ouvrir les Données de la Recherche

La France continuera à apporter son soutien à la Research Data Alliance (RDA), réseau international définissant les bonnes pratiques dans le domaine des données de la recherche.

Axe Ouvrir et promouvoir les codes sources produits par la recherche

Afin de faciliter la coordination des communautés du logiciel libre à l'échelle nationale et internationale, un Collège des codes sources et des logiciels sera créé au sein du Comité pour la science ouverte, des liaisons seront établies avec la mission logiciels libres portée par la Direction interministérielle du numérique, avec l'European Open Science Cloud (EOSC), la Research Data Alliance (RDA), la Research Software Alliance et FORCE 11.

RDA FRANCE ACTIVITIES AND ITS RELATIONSHIP WITH THE NATIONAL PLAN FOR OPEN SCIENCE

March 2018

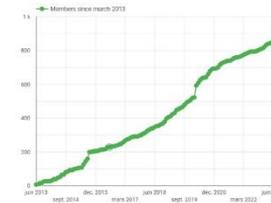
Creation of the French RDA Node

September 2019

Financial support from the French National Fund for Open Science



Figure: Evolution of RDA French members since March 2013



1,000+ members RDA France and RDA in France in October 2023

RDA FRANCE ACTIVITIES

LIAISON WITH RDA GLOBAL

-  Financial support to RDA Global
-  RDA Regional Advisory Board Co-chairpersonship
-  Support to RDA French community for participating in RDA Global activities
-  « RDA en Bref » Translation of "RDA" in a "Nutshell"

LEADING THE FRENCH MEMBER COMMUNITY

-  Organisation of annual meetings since 2018
-  Workshops, webinars, University courses
How to participate in the RDA Data Policies of Research Entities
How to bring researchers to engage in DMPs And more...
-  Promotion of CoreTrustSeal Certification
10+ workshops
-  RDA France FAIR newsletters
-  Working group on disciplinary DMPs

RDA FRANCE SUPPORTING THE FRENCH OPEN SCIENCE STRATEGY

RDA France **supports** its very diverse community **acculturate** to the data aspects of Open Science, fully in line with the objectives of the **National Plan for Open Science**



FIRST NATIONAL PLAN FOR OPEN SCIENCE 2018-2021

"Support the Research Data Alliance (RDA) and found the French chapter of the organisation (RDA France)"



SECOND FRENCH PLAN FOR OPEN SCIENCE 2021-2024

"France will continue to provide support for the Research Data Alliance (RDA), an international network that defines best practices in the domain of research data"

RDA FRANCE & THE FRENCH COMMITTEE FOR OPEN SCIENCE



Working group "**Certification of data repositories and services**"
Steering activities linked to certification for "trusted" data repositories and services at the national level

Internationalise topics addressed by the national Committee for Open Science by bringing them to the RDA.



Contacts
François Genova
Laurence El Khouri
contact-rdafrance@services.cnrs.fr

Poster RDA France
21^{ème} Plénière RDA

Parmi les ateliers et webinaires RDA France récents

- Amener les scientifiques vers la démarche d'élaboration des plans de gestion des données, 18 et 25 mars 2022
 - Billet de blog, [compilation des contributions des participants](#)
- Webinaire sur la Politique des Données des Entités de Recherche, 26 mai 2023
- Atelier sur la Certification des entrepôts de données, 11 mars 2024

Les ateliers et webinaires des Journées RDA France 2023

- Réunion plénière le 12 octobre
- Webinaire Projets nationaux sur les données de la recherche le 10 octobre, l'après-midi
- Une série d'ateliers:
 - Lundi 9 octobre après-midi, atelier sur la Politique des Données des entités de recherche, suite du webinaire organisé le 26 mai avec les collègues de l'ANR BRIDGE
 - Mardi 10 octobre matin, atelier "Plan de Gestion des Données: Bonnes pratiques", organisé avec l'INIST
 - Mercredi 11 octobre après-midi, atelier sur les Nouveautés de DMP OPIDoR, organisé avec l'INIST
 - Vendredi 13 octobre matin, atelier autour de la Certification destiné aux entrepôts de données, organisé par le Groupe de Travail commun au Collège Données du CoSO et à RDA France
 - Vendredi 13 octobre après-midi, atelier " Comment participer à la RDA "

<https://rdafrance2023.sciencesconf.org/>

En cours et à venir

- › Soutien à la participation à la vingt-deuxième réunion plénière de la RDA, 14-23 mai – date limite **10 avril à 12h**

<https://framaforms.org/rda-france-rda-p22-demande-de-prise-en-charge-1710666561>

- › Atelier Comment participer à la RDA et à RDA France

- › 11 avril 2024, 15h00

- › 23 avril 2024, 14h00

<https://framaforms.org/inscription-a-latelier-comment-participer-a-la-rda-et-a-rda-france-1707751434>

Parmi les sessions de la 22^{ème} Plénière

- IG Engaging Researchers with Data
- IG Evaluation of Research: RDA in the Evaluation of Research Landscape
- IG Professionalising Data Stewardship
- WG RDA & ReSA Policies in Research Organisations for Research Software
- IG Data Versioning: Translating the Data Versioning Principles into machine actionnable recommendations
- **BoF National Data Services: Do we want to meet in the RDA?**

Liens utiles

- › *Le site de la RDA est en cours de mise à jour*
- › Pour contacter RDA France
contact-rdafrance@services.cnrs.fr
- › Pour s'inscrire sur la liste rda-France
<https://listes.services.cnrs.fr/wws/subscribe/rda-france>
- › Listes thématiques RDA France
 - › RDA France FAIR
<https://listes.services.cnrs.fr/wws/subscribe/rda-france-fair>
 - › RDA France Certification
<https://listes.services.cnrs.fr/wws/subscribe/rda-france-certification>