

Etude de cas: Comment la reproductibilité est assurée dans le domaine de l'astronomie



Françoise Genova

Observatoire Astronomique de Strasbourg



Merci à Caroline Bot pour
notre discussion sur la question
et à Mark Allen pour ses diapos



□ Reproductibilité?

- Le concept est peu ou pas évoqué en astronomie
- On va essayer de comprendre pourquoi



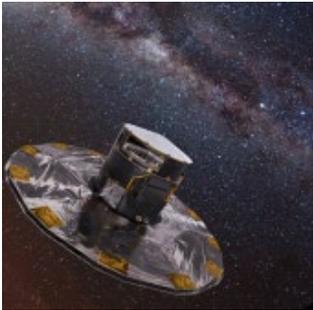
- Researchers and practitioners from any discipline are able to find, access and process the data they need. They can be confident in their ability to use and understand data, and they can evaluate the degree to which that data can be trusted.

<https://www.ouvrirlascience.fr/riding-the-wave-how-europe-can-gain-from-the-rising-tide-of-scientific-data-final-report-of-the-high-level-expert-group-on-scientific-data/>

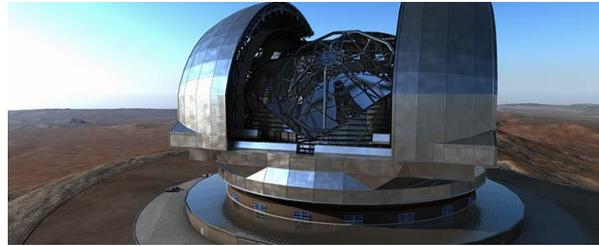


COMMENT LA COMMUNAUTÉ DES ASTRONOMES A CONSTRUIT LA CONFIANCE

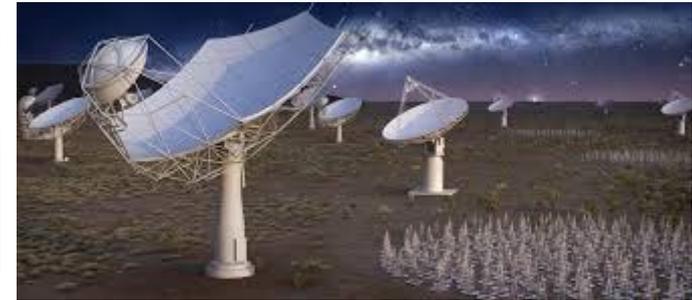
Les infrastructures de recherche en astronomie



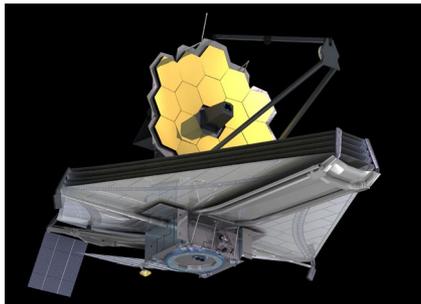
Gaia



ELT



SKA



JWST



Euclid



CTA

+ LES DONNEES!

□ Les données comme infrastructure de recherche

- Les astronomes utilisent des données qu'ils obtiennent en ligne dans leur travail de recherche
- L'infrastructure des données en astronomie a de nombreux composants
 - Les archives des observatoires sol et spatiaux
 - Les grands relevés du ciel
 - Les bases de données à valeur ajoutée – par exemple celles du Centre de Données astronomiques de Strasbourg (CDS)
 - Les journaux académiques, avec les données liées aux articles
 - Les résultats de modélisation

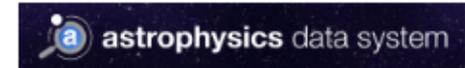
Les éléments du système sont en liaison

Conseil du CDS 2024, Mark 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*

OV FRANCE EUROVO



Certified:



Wider intl. networks:



European Projects:



ASTRO-CC

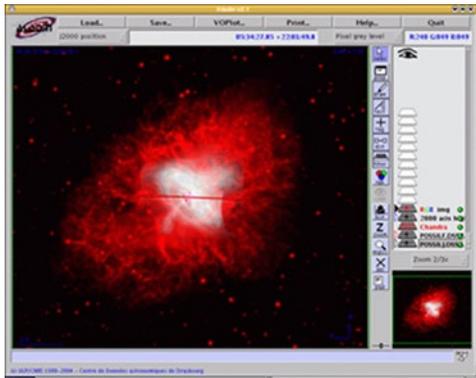
□ Comment s'établit la confiance

- La confiance n'est pas seulement liée à la qualité des données, mais aussi
 - à la « qualité » des différents éléments du système de production et de partage des données
 - Au fait que le système répond aux besoins des utilisateurs et des utilisatrices
- Les astronomes ont confiance
 - dans les producteurs de données
 - Long processus de validation des grands projets sol et spatiaux (pertinence scientifique, faisabilité, mise en compétition) et mise en œuvre par de grands acteurs aux capacités reconnues
 - Feuilles de route européenne et nationale
 - Le CDS est sur la Feuille de Route nationale des Infrastructures de Recherche depuis la première version de celle-ci
 - Dans le fait qu'elles/ils pourront trouver les données pertinentes pour leur recherche, y accéder et les réutiliser en connaissance de cause
 - Mise à disposition des données et des métadonnées, et des outils pour les utiliser



COMMENT LES ASTRONOMES PARTAGENT LES DONNÉES

□ Pourquoi partager les données en astronomie?



- Les données sont au cœur de l'approche scientifique de la discipline
 - Astronomie multi-longueur d'onde/multi-messager (ondes gravitationnelles, etc.)
Combiner les observations de différents instruments pour comprendre les phénomènes physiques à l'œuvre dans les objets
 - Variabilité sur des échelles 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 politique des données des infrastructures en astronomie

- Appels à proposition compétitifs pour le temps d'observation
- La plupart des observatoires mettent leurs données à disposition en général après une « période propriétaire » (embargo, en général six mois ou un an)
- Un élément important
 - pour rendre le partage des données acceptables par la communauté
 - pour continuer à avoir les meilleures propositions d'observation, donc les meilleures données du point de vue scientifique dans les archives
- Gros travail
 - sur le pipeline de données dans les observatoires
 - sur la sélection et la curation des données dans les services à valeur ajoutée
- La communauté a confiance dans les données qu'elle trouve dans les archives et dans les services de données

□ Les standards jouent un rôle essentiel

- Point de départ: le format FITS (1977-1981)
 - 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
 - OAI-PMH, [Dublin Core + extensions disciplinaires](#) => interopérabilité
 - Interopérabilité des données et des outils qui permettent de les utiliser

□ Les standards sont ouverts

- Tout le monde peut participer à la définition des standards.
- Les standards sont en accès libre et tout le monde peut les utiliser
- Plus de 100 « autorités » ont déclaré au moins un service dans le registre de ressources de l'IVOA
- Les standards sont customisés par la planétologie, les astroparticules (Clusters ASTERICS et ESCAPE), le Virtual Atomic and Molecular Data Centre

□ 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

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 November 2024 Interoperability Meeting](#)
[15-17 November 2024, La Valletta \(Malta\)](#)

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](#)



□ Les membres de l'IVOA

Member Organizations



- [Argentine Virtual Observatory](#)
- [Armenian Virtual Observatory](#)
- [AstroGrid, United Kingdom](#)
- [Australian All-Sky Virtual Observatory](#)
- [Brazilian Virtual Observatory](#)
- [Chinese Virtual Observatory](#)
- [Canadian Virtual Observatory](#)
- [Chilean Virtual Observatory](#)
- [European Space Agency](#)
- [European Virtual Observatory](#)
- [German Astrophysical Virtual Observatory](#)
- [Japanese Virtual Observatory](#)
- [Kazakhstan Virtual Observatory](#)
- [Netherlands Virtual Observatory](#)
- [Observatoire Virtuel France](#)
- [Russian Virtual Observatory](#)
- [Square Kilometer Array Observatory](#)
- [South African Astromatics Alliance](#)
- [Spanish Virtual Observatory](#)
- [Italian Virtual Observatory](#)
- [Ukrainian Virtual Observatory](#)
- [US Virtual Observatory Alliance](#)
- [Virtual Observatory India](#)

Partners



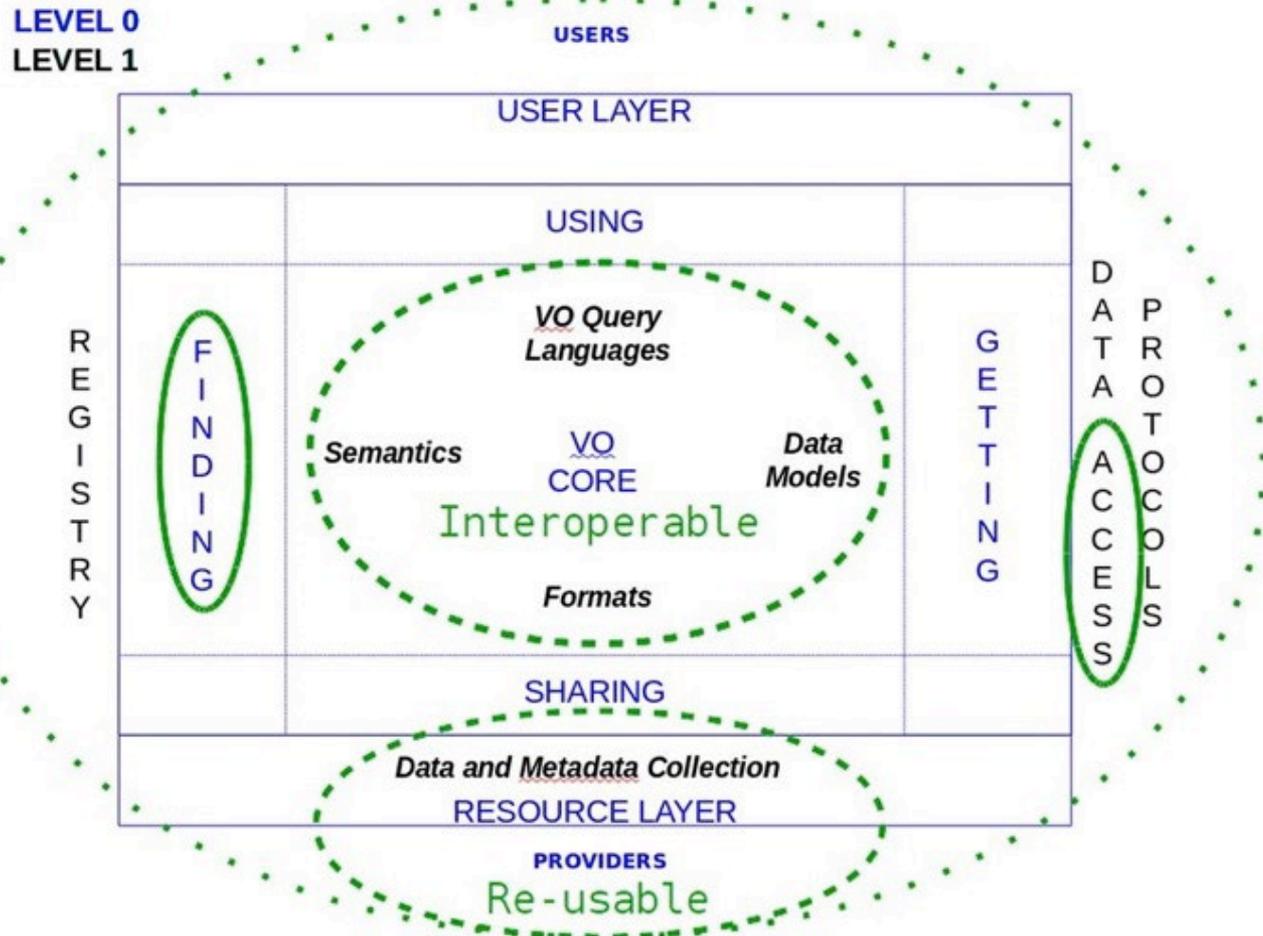
[IAU Office of Astronomy for Development \(OAD\)](#)

□ Les profils des participants

- Parmi les personnes qui participent de façon active aux travaux de l'IVOA, des personnels des archives d'observatoires et des centres de données, des développeurs de services et d'applications
- Profils variés: chercheurs, gestionnaires de projet, ingénieurs informaticiens, curateurs de données
- Pertinence par rapport aux attentes (scientifiques et techniques) des utilisateurs

VO is FAIR

Merci à Mark Allen

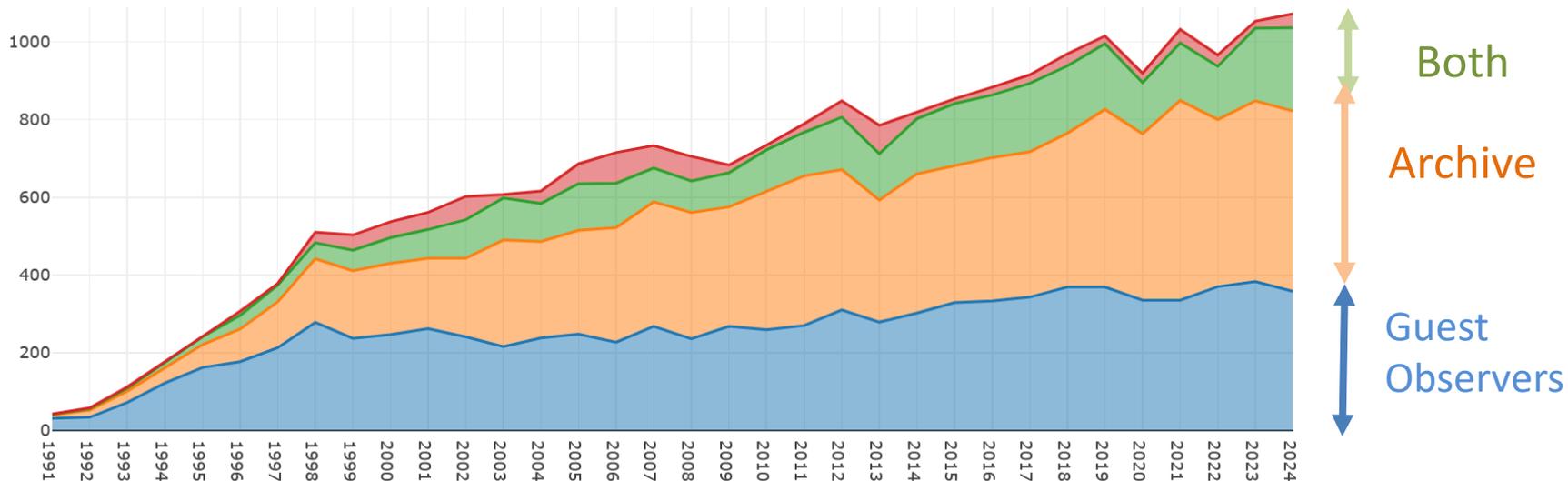
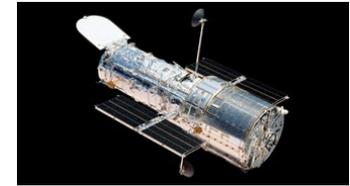


□ Les Groupes d'Intérêt et la procédure de standardisation

- Les Groupes d'Intérêt (*Interest Groups*) rassemblent les besoins et agissent en soutien des Groupes de Travail pour la définition des standards
 - [Data Curation and Preservation](#)
 - [Education](#)
 - [Knowledge Discovery in Data Bases](#)
 - [Operations](#)
 - [Solar System](#)
 - [Theory](#)
 - [Time Domain](#)
 - [Radio Astronomy](#)
 - [High Energy](#)
- La [procédure de standardisation](#) est inspirée de celle du W3C
Working Draft, Proposed Recommendation, Request for Comments, Recommendation

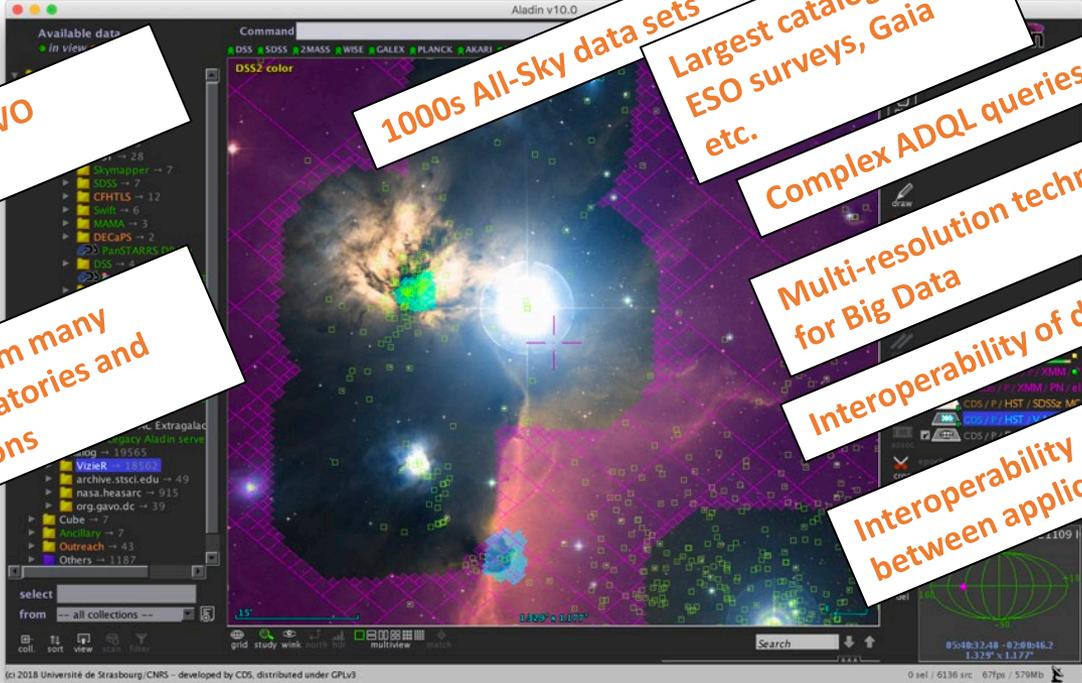
□ L'impact des données ouvertes sur les infrastructures: Hubble Space Telescope, articles publiés

Les données sont trouvable, accessibles et réutilisables
Elles sont trouvées, accédées et réutilisées!



<https://archive.stsci.edu/hst/bibliography/pubstat.html>

One view of the VO from an application/portal :



Available data
in view

Command
DSS SDSS ZMASS WISE GALEX PLANCK AKARI

DSS2 color

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

Built from VO
Registry

Data from many
observatories and
missions

select
from -- all collections --

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

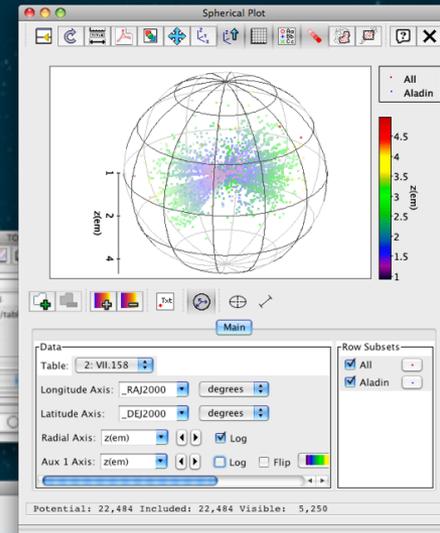
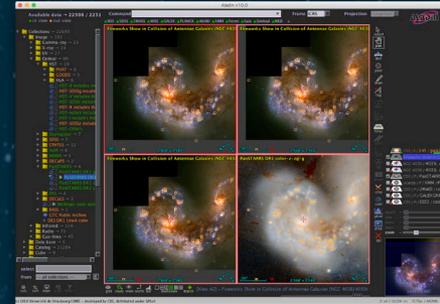
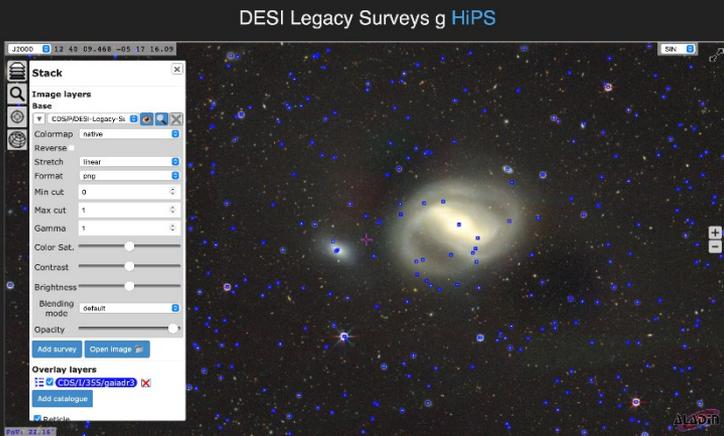
0 sel / 6136 src. 67fps / 570Mb

Diapositive de Mark Allen, CDS



Merci à Mark Allen

VO interoperability –



Access to PB-scale all-sky data



Table List

Current Table Properties

Label: VII.158

Location: VII/158/tab

Name: 22484

Rows: 22484

Columns: 14

Sort Order: All

Row Subset: All

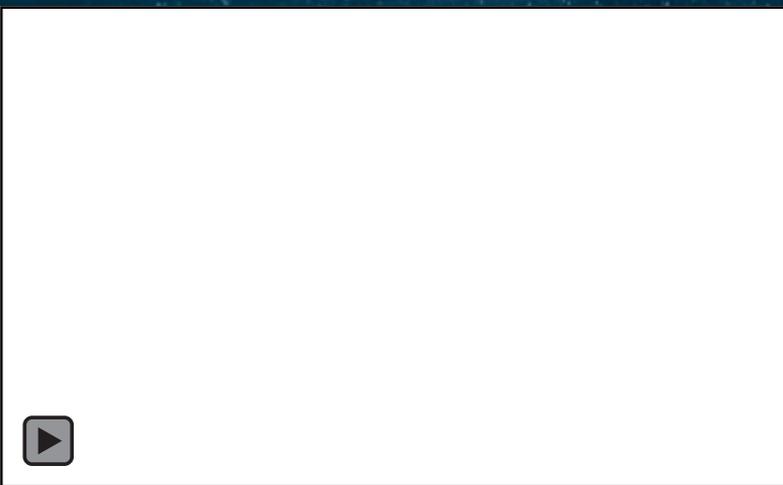
SAMP

Messages:

TOPCAT(1). Table Browser

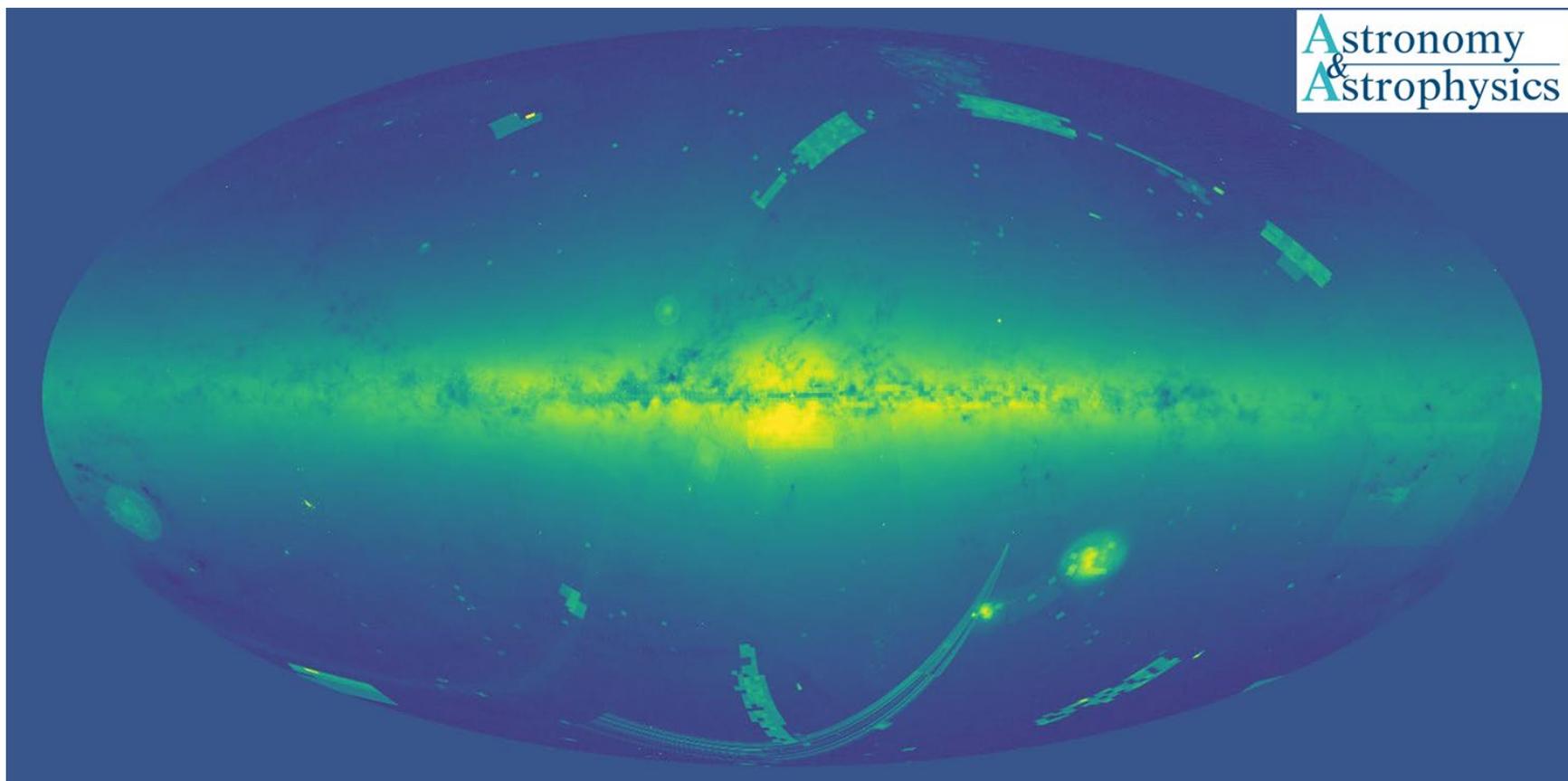
Table Browser for I. III.157

Seq	QSO	Name	z	vmag	Type	Camp	ObsSp
31	1133+704	Mk 180	0.946	14.49	BLZ	1	CaSp ObsSp
32	1146+037	PKS	0.341	14.9	QSO	1	CaSp ObsSp
33	1146+549	PG	0.969	15.82	QSO	1	CaSp ObsSp
34	1156+295	4C 29.45	0.729	14.41	BLZ	1	CaSp ObsSp
35	1202+281	PG	0.145	15.51	QSO	1	CaSp ObsSp
36	1211+141	PG	0.085	14.63	QSO	2	CaSp ObsSp
37	1219+755	Mk 205	0.07	14.5	YL	2	CaSp ObsSp
38	1225+317	IC	2.219	15.87	QSO	1	CaSp ObsSp
39	1229+074	IC 273	0.158	12.66	QSO	2	CaSp ObsSp
40	1229+204	TON 3542	0.044	15.3	YL	2	CaSp ObsSp
41	1241+376	PG	1.273	15.38	QSO	1	CaSp ObsSp
42	1253+095	IC 279	0.538	17.75	BLZ	2	CaSp ObsSp
43	1302+102	PKS	0.286	14.92	QSO	2	CaSp ObsSp



Virtual Research Environment of:
Services, apps, **notebooks**, platforms, APIs, ...

- Données des tables attachées à des publications sur une carte du ciel (service VizieR du CDS)



□ Les logiciels

ASCL.net

Astrophysics Source Code Library

Making codes discoverable since 1999

[Home](#) [About](#) [Resources](#) [Browse](#) [Submissions](#) [News](#) [Dashboard](#)

Welcome to the ASCL

The Astrophysics Source Code Library (ASCL) is a free online registry and repository for source codes of interest to astronomers and astrophysicists, including solar system astronomers, and lists codes that have been used in research that has appeared in, or been submitted to, peer-reviewed publications. The ASCL is indexed by the [SAO/NASA Astrophysics Data System](#) (ADS) and Web of Science and [is citable](#) by using the unique ascl ID assigned to each code. The ascl ID can be used to link to the code entry by prefacing the number with ascl.net (*i.e.*, [ascl.net/1201.001](#)).

Most Recently Added Codes

2025 Mar 31

[submitted] [GalClass: Visual Galaxy Classification tool](#)

[Pastras, Stavros](#)

GalClass facilitates visual morphological classifications of large samples of galaxies taking advantage of multi-wavelength imaging and ancillary information. It offers a versatile Graphic User Interface (GUI), which adapts to the provided classification scheme. It displays a series of pre-prepared PDF files for classification, grouping by galaxy and filter, while also listing relevant metadata and displaying a color image of each source. It enables easy navigation through the sample and continuously outputs classification results in a JSON file. Finally, it offers an analysis submodule which combines and processes output files of multiple classifications.

2025 Mar 30



CE QU'IL FAUT METTRE EN ŒUVRE – L'EXEMPLE DU CDS

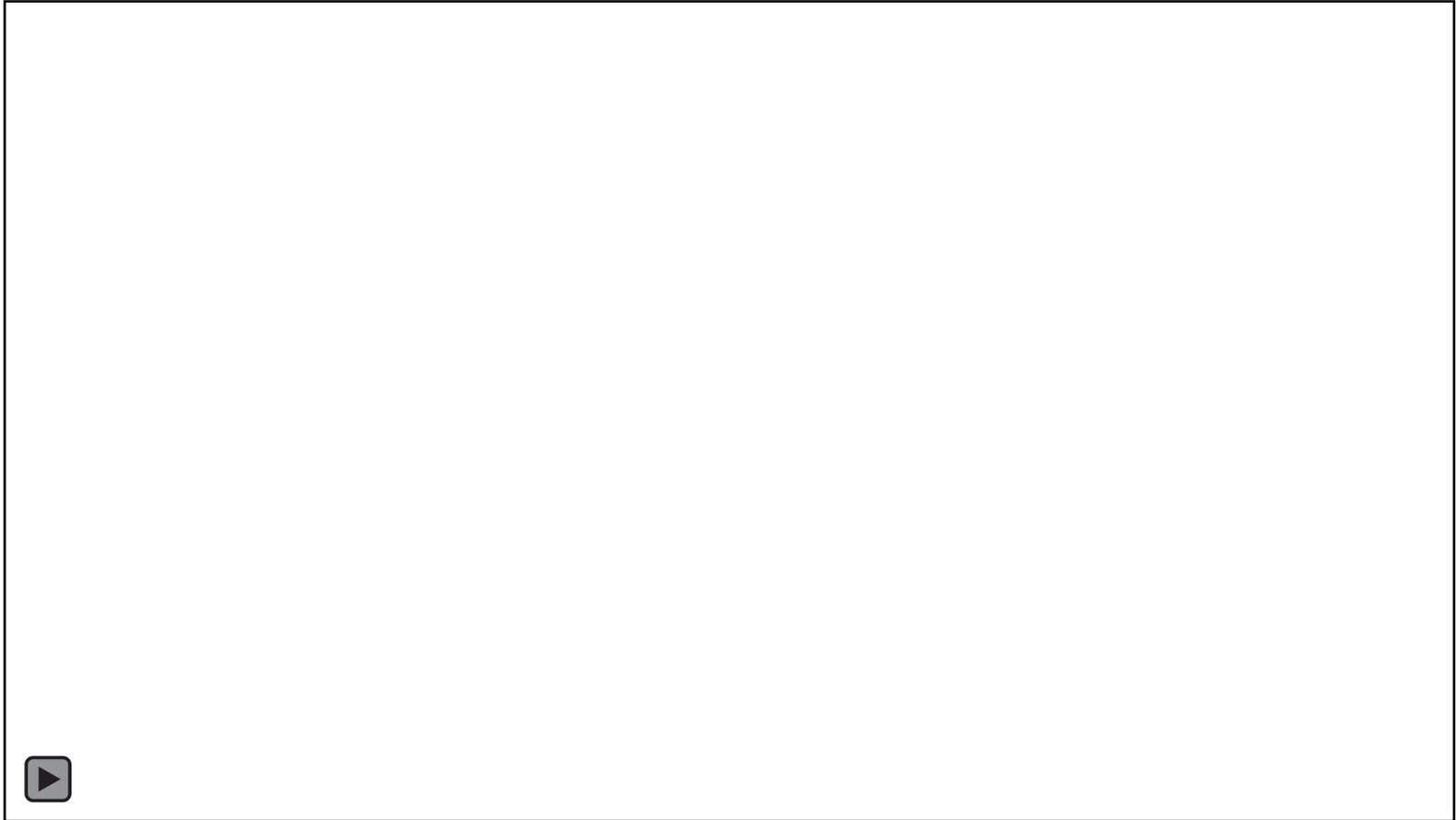
□ Le cœur de la mission du CDS

- 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

□ Beaucoup de travail sous-jacent

- Pilotage par les besoins scientifiques – il faut les identifier et les anticiper!
- Très gros travail de curation des données
 - Vérification et comparaison avec les autres données disponibles
 - Description homogène de données hétérogènes
 - Métadonnées, liens, préservation
- Très gros travail de développement logiciel
 - Veille technologique
 - Tests des nouvelles technologies (stages)
- Il faut rassembler, développer et maintenir les compétences sur la longue durée!

□ Un jour normal au CDS





CONCLUSIONS

□ Message à retenir

Les données astronomiques sont FAIR (pour les besoins de la communauté) et largement réutilisées par celle-ci dans son travail de recherche

Construire la confiance dans le système de production et de partage des données demande à tous les acteurs une recherche constante de la qualité et de s'organiser au niveau international, et beaucoup de travail de beaucoup de gens de profils différents

Panel interdisciplinaire sur le partage des cadres disciplinaires de partage des données 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](#)

□ Les leçons apprises sur la longue durée (1)

- Pourquoi partager les données scientifiques
 - Des incitations fortes au niveau politique
 - La raison dans les faits: une révolution dans la manière de faire de la science! Faciliter le travail des scientifiques et leur ouvrir de nouvelles possibilités
 - Un élément essentiel de la reproductibilité si c'est fait correctement
- La confiance est un élément essentiel
 - Qualité des données, fournisseurs de données « de confiance »
 - Pertinence par rapport aux besoins des utilisateurs

□ Les leçons apprises sur la longue durée (2)

- Evolution constante
 - Du contexte « politique »
 - Du contexte scientifique
 - Du contexte technique – imaginez l'irruption du Web! Et maintenant l'IA
- Un travail sur la longue durée
 - Construction des compétences, des ensembles de données et des services
 - Prenant constamment en compte les évolutions mais en évaluant leur pertinence et si elles ont une pérennité suffisante
 - Assurer le passage de témoin entre les générations
- Beaucoup de travail sous-jacent
 - Données, métadonnées, services, applications, standards
 - Invisible des utilisateurs si c'est réussi!