

Sauver les données Eros II



Jna

Jean-Noël Albert – LAL / IJClab

Juin 2020

Expérience de Recherche d'Objets Sombres : Eros

90 millions d'étoiles

2 millions d'images

7 années de campagne

2000 nuits d'observation

70 000 fichiers de mesures

90 M courbes de lumières

14 To d'images et

6 To de mesures

La Voie lactée à La Silla (Chili).
En haut à droite, le Grand Nuage de Magellan,
en bas à droite, le Petit Nuage (ESO/Z.Bardon)



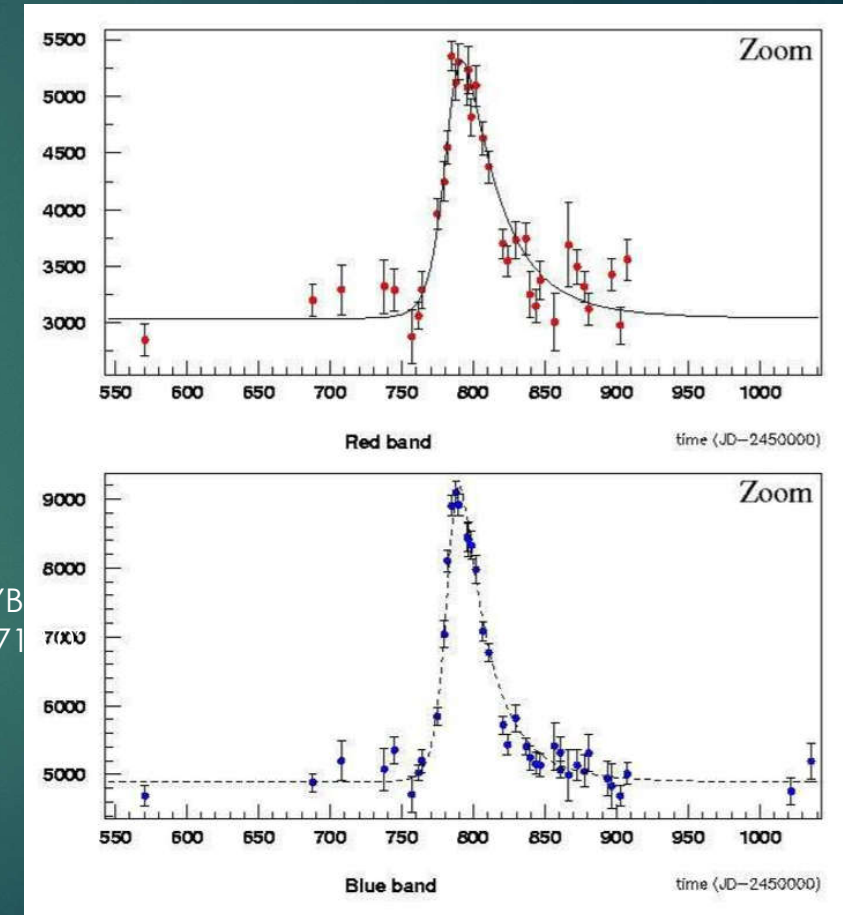
Microlentilles gravitationnelles

- ▶ Recherche de microlentilles gravitationnelles
 - ▶ Objets obscures massifs dans le halo galactique
 - ▶ Variations du flux de la lumière d'étoiles lointaines
 - ▶ Observation des mêmes zones sur une longue durée
- ▶ Exemple d'une courbe de lumière
- ▶ Exemple du fichier des mesures

```

$ head lm0153n22431.time
# star: erosid      MagR  ErrMR  XR      YR      MagB  ErrMB  XB      YB
#   lm0153n22431 19.133  0.124  478.78  684.37  19.238  0.070  431.84  71
#
#   date      MagR  ErMagR  MagB  ErMagB
348.81416  19.224  0.037  19.330  0.025
366.79045  99.999  9.999  19.337  0.035
373.68669  19.115  0.051  19.299  0.032
376.84666  19.195  0.058  19.287  0.036
380.86332  19.210  0.052  99.999  9.999
381.70421  19.186  0.081  99.999  9.999

```



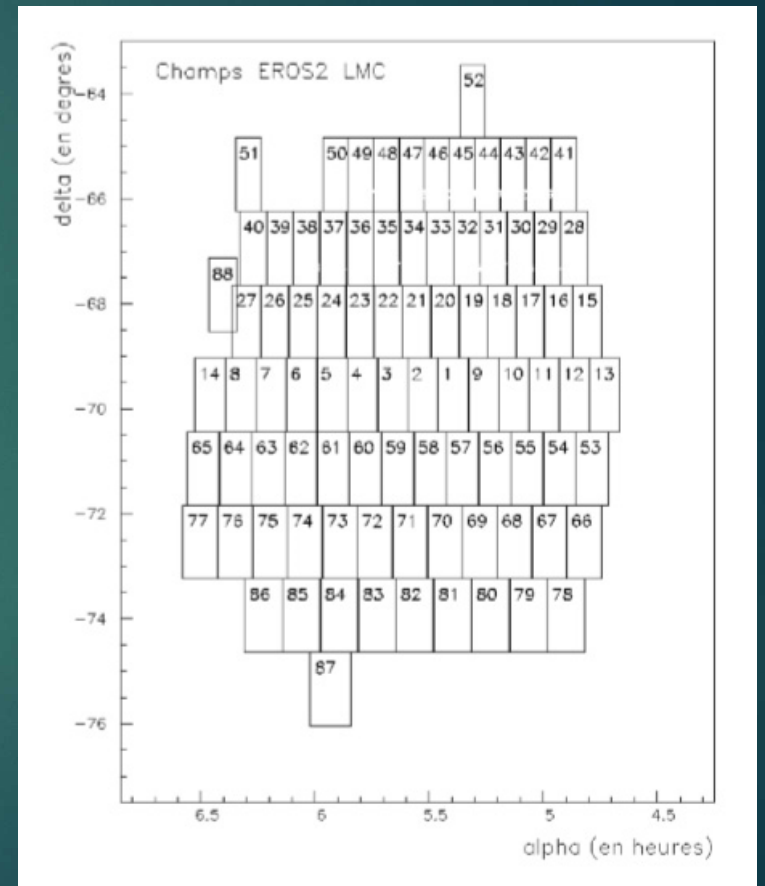
Les Champs du Grand Nuage



Les Nuages de Magellan

Le Petit

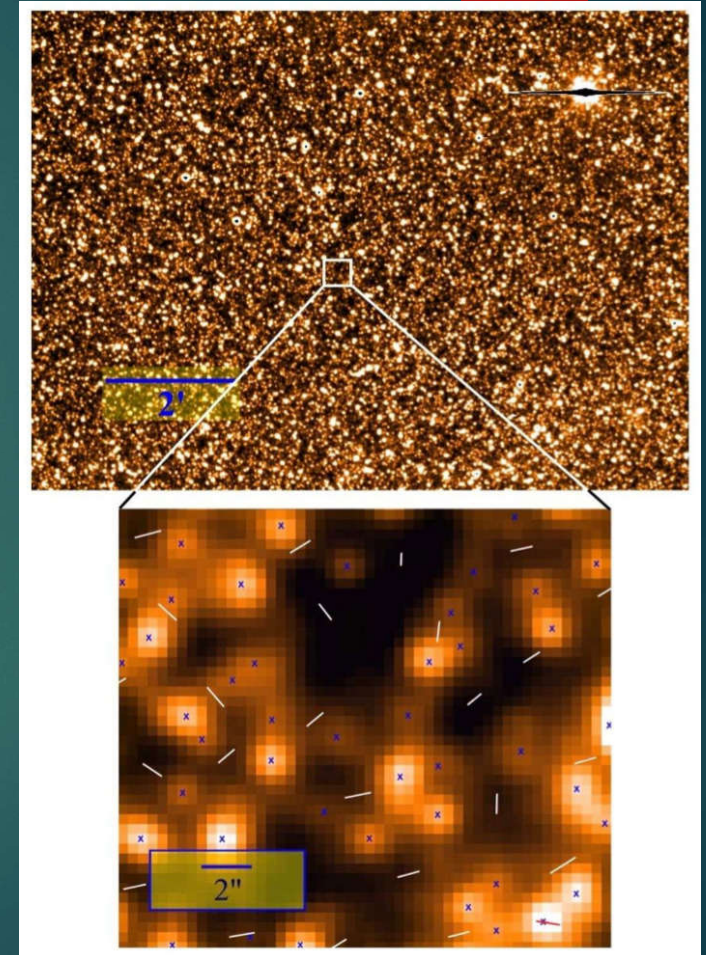
Le Grand



P. Tisserand – fg 4.6b p82

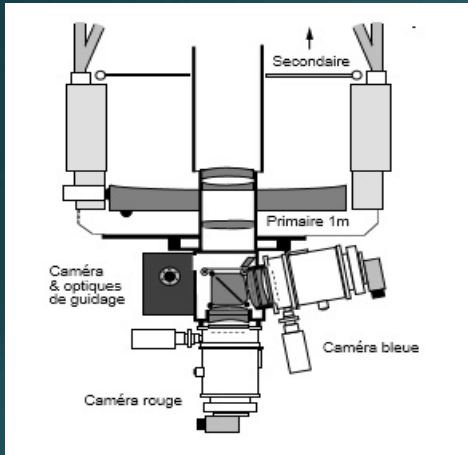
Les images Eros II

- ▶ Vue du centre du Grand Nuage de Magellan, une zone dense de 20 000 étoiles.
- ▶ En bas, un zoom de 20'' d'arc.
- ▶ Le but des analyses : localiser chaque étoile sur les images réalisées nuit après nuit et mesurer leurs paramètres.

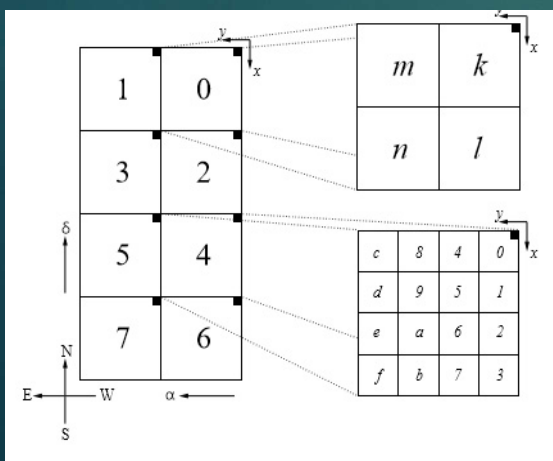


La double caméra

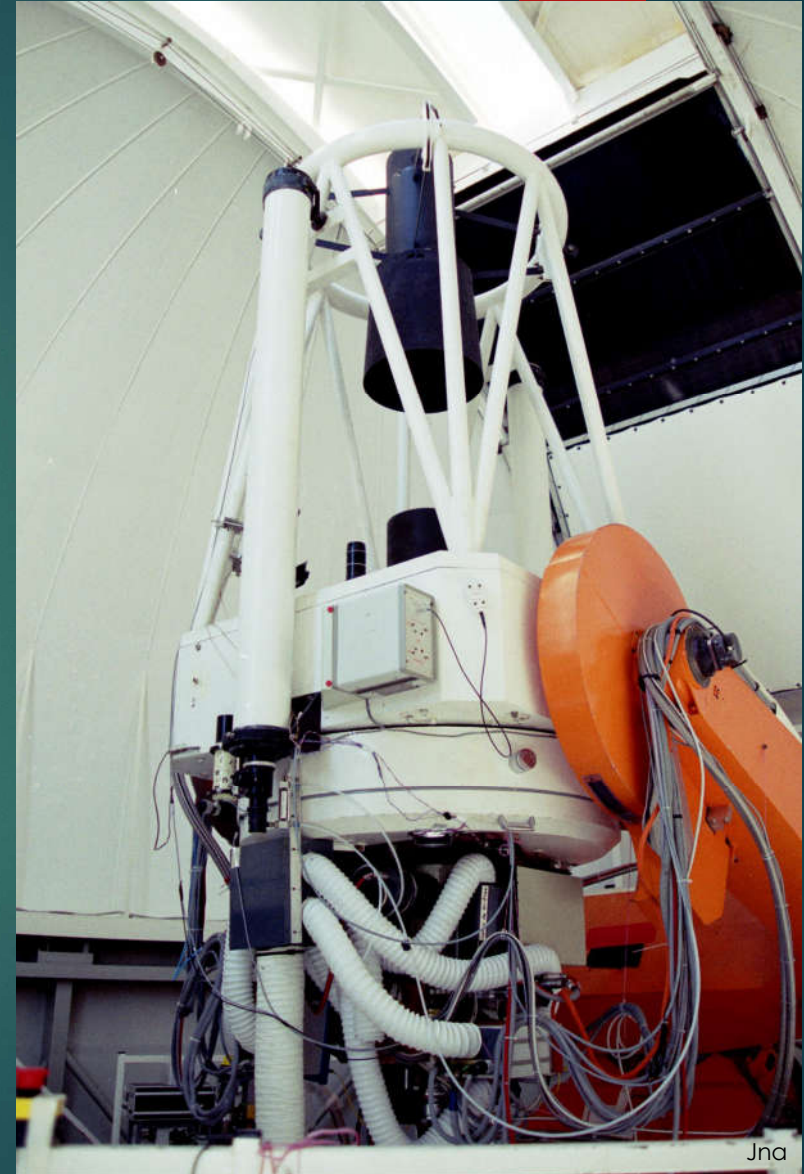
P. Tisserand – fg 4.1 pg 75



P. Tisserand – fg 4.2 pg 76



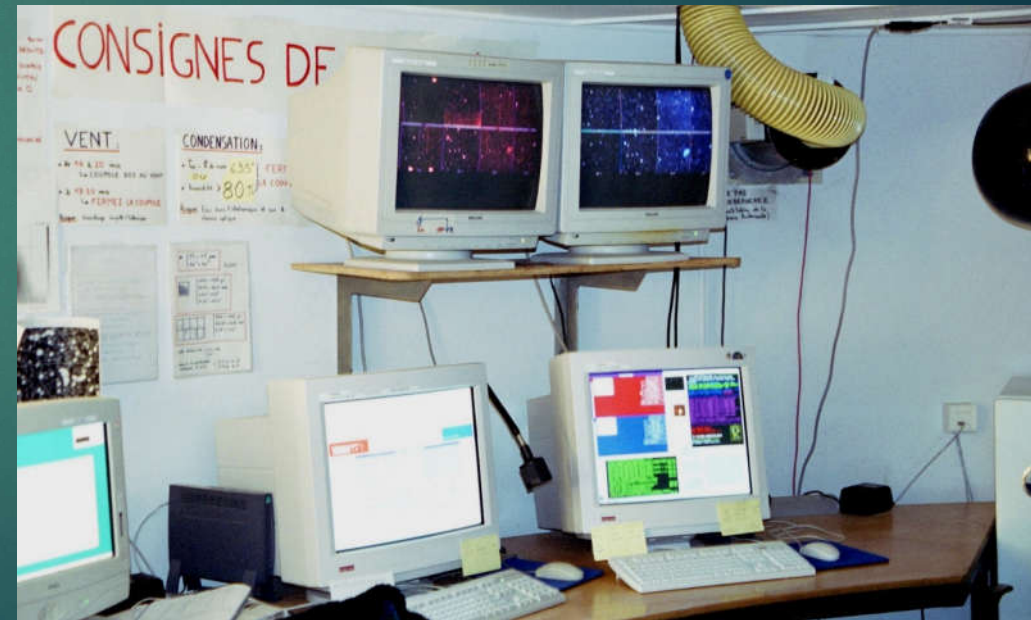
Jna



Jna

Images brutes et réduites

- ▶ Les images issues des CCD sont des images « brutes ».
 - ▶ « Normalisées » grâce à des images « de calibration » pour former des images « réduites » utilisées dans les analyses.
 - ▶ une image « brute » pour chaque image « réduite ».
 - ▶ Réduction est directement sur le site.
- ▶ Transfert des images par DLT.
 - ▶ Les images brutes à Saclay.
 - ▶ Les images réduites à Lyon.
 - ▶ Recopie sur 3480/3490 (200 Mo) et indexation dans la base de données Oracle
- ▶ Migration vers HPSS et désormais Irods.



Situation des images

- ▶ Perte des images brutes de Saclay.
 - ▶ Non transférées sur des supports plus modernes.
 - ▶ 2 millions de fichiers perdus - 1 600 DLT.
- ▶ Cassettes de cartouches HPSS.
 - ▶ Quelques milliers de fichiers perdus.
- ▶ Un programme non transféré.
 - ▶ 60 000 images perdues.
- ▶ Au total, hors images brutes, 4 à 5 % de pertes.

Traitement des données

- ▶ Une étape sur le site : la réduction des images brutes
- ▶ Plusieurs étapes au Centre de calcul, en batch.
 - ▶ Création de très bonnes images pour une meilleure détection des étoiles
 - ▶ Détection des étoiles et création de catalogues
 - ▶ Mesure des paramètres des étoiles des différentes images
 - ▶ Sauvegarde dans des fichiers binaires dans un format « expérience ».
 - ▶ Pour réduire la taille des fichiers et les volumes traités, les analyses sont conduites par quart de CCD, par flux de couleur et par segment de nuit.

Limites des formats « expérience »

- ▶ Catalogues des étoiles détectées, utilisés pour identifier les étoiles sur les différentes images.
- ▶ Fichiers des mesures réalisées sur chaque étoile de chaque image – nommés fichiers de suivis.
- ▶ Ces fichiers sont dans un format « expérience » et supposent une bibliothèque C++ pour être accédés. Cette bibliothèque n'a pas été maintenue.
 - ▶ Accès tout récent à des documents sur la structure des fichiers.
 - ▶ Espoir de conversion vers un format « ouvert ».
- ▶ En tout état de cause, les règles de la publication des données imposent de pouvoir les relire...

Les courbes de lumière

- ▶ Regroupe les 7 années de campagne dans les deux couleurs dans un seul fichier par étoile suivie.
- ▶ Format ASCII ne nécessitant pas de librairie de lecture dédiée.
 - ▶ Accès universel.
- ▶ Ensemble réduit de données : date de mesure, mesures bleues et rouges, erreurs sur les mesures.
- ▶ Fichier de petites tailles : 15-20 KB
- ▶ Environ 1.2 TB non comprimés, soit 450 GB comprimés.

Base de données

- ▶ Base de données Oracle utilisée comme catalogue des fichiers.
 - ▶ Mise en œuvre dès le début de l'expérience.
- ▶ Tous les fichiers « officiels » de l'expérience y sont référencés – ce qui a beaucoup facilité les migrations et les validations post-transferts – à l'exception notable des courbes de lumière.
- ▶ Dépôt pour les catalogues ASCII des courbes de lumière sous la forme de BLOB comprimés.
- ▶ Support pour les outils des recherches des données – écrits en Java et utilisant l'ORB Hibernate et l'API Java à Irods : Jargon.

Préservation et Seconde source

- ▶ Discussion avec le Centre de calcul pour la préservation des données de l'expérience.
 - ▶ Un processus semble-t-il assez complexe nécessitant une aide solide.
- ▶ Besoin de trouver une seconde source de stockage indépendant du CC, pour éviter les incidents rencontrés avant la migration HPSS → Irods.
 - ▶ Des discussions avec le SI LAL/IJCLab – au point mort...

Références et URL

- ▶ Le blog Eros / Anastasis : <https://groups.lal.in2p3.fr/erosanastasis/>
- ▶ Le site ErosDb II : <http://eros.lal.in2p3.fr/ErosDB/>
- ▶ Le site historique Eros : <http://eros.in2p3.fr/>

