Eros II, c'est 90 millions d'étoiles analysées à partir de 2 millions d'images FITS correspondant à 7 ans de campagne soit 14 To de fichiers.

C'est 2 000 nuits d'observation.

C'est plus de 60 pointages du télescope par nuit soit une prise de vue toutes les 5 à 10 mn pour près de 1000 images chaque nuit.

C'est 70 000 fichiers binaires de mesures pour 6 To.

C'est 90 millions de courbes de lumière ASCII, correspondant aux étoiles suivies, soit 500 Go sous forme compressée.

Toutes ces observations ont été conduites grâce au télescope de 1m de l'expérience, le Marly, à La Silla, le site de l'ESO au Chili de Juin 1996 à Février 2003.

Les données sont au Centre de calcul de l'IN2P3, dans le système de fichiers distribué Irods.

L'ensemble des données est référencé dans une base de données Oracle.

Expérience de Recherche d'Objets Sombres : Eros

2

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)

Eros, ou Expérience de Recherche d'Objets Sombres.

L'expérience s'inscrit dans la problématique de la "masse manquante des galaxies" : si on compare le nombre d'étoiles visibles des grandes galaxies à la masse qu'elles devraient avoir pour rester homogènes, 80% est invisible.

Une hypothèse était que de "petits" objets célestes – genre Jupiter – pouvaient exister à la périphérie et contribuer à cette masse manquante.

Non lumineux, aucune observation directe n'est possible, mais on peut détecter leur présence par l'effet de (micro) lentilles gravitationnelles. Comme la probabilité d'observer un tel phénomène est très rare, il faut multiplier les observations pour augmenter les chances de détection.

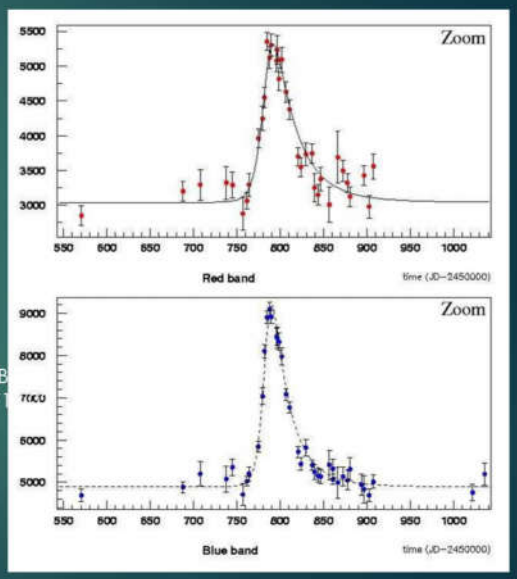
L'expérience Eros a été conçue dans ce but. Malheureusement, l'hypothèse n'a pas été validée. La masse manquante est ailleurs.

Microlentilles gravitationnelles

3

- 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
```

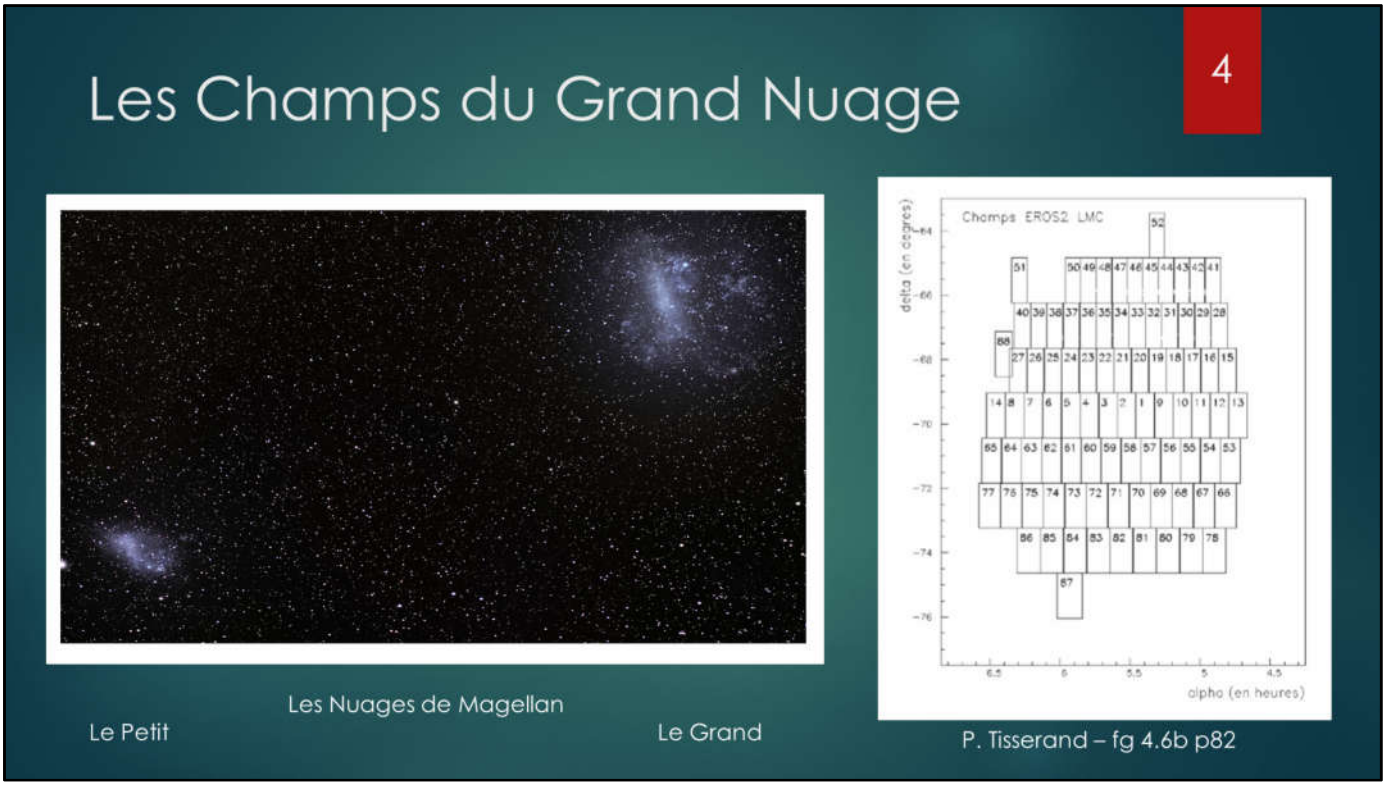


P. Tisserand – fg 9.4b pg 191

Le passage d'un objet sombre massif dans la ligne de visée d'une étoile lointaine conduit à une variation significative de la lumière de l'étoile, comme le montre le diagramme (thèse de Patrick Tisserand, 2004).

Les Grand et Petit Nuages de Magellan, des galaxies naines satellites de la Voie Lactée, constituent un vivier de telles étoiles. L'expérience a donc observée nuit après nuit, durant plus de 2 445 jours, soit 7 années, les étoiles de ces deux Nuages.

Les mesures réalisées sont converties à la fin des productions en fichiers ASCII regroupant l'ensemble des observations pour chacune des étoiles suivies. En bas, un extrait d'un tel fichier.



La base d'information que constitue les observations de l'expérience Eros pourrait être précieuse pour des expériences plus récentes. En effet, en astronomie, on ne peut pas « rejouer » une mesure – contrairement aux expériences accélérateurs. Ce qui n'a pas été observé est perdu. Ce qui a été observé peut servir à mieux comprendre les observations à venir.

Ici, une vue des Grand et Petit Nuage de Magellan. Ces deux cibles étaient couvertes par plusieurs pointages successifs du télescope dans la nuit.

Le diagramme montre les 88 "champs", ou zones de pointage couvrant la région du Grand Nuage (thèse de Patrick Tisserand). Chaque champ représente 2 fois 8 images CCD au format FITS.

Les images Eros sont au format FITS, le standard de l'astronomie.

Une image Eros II représente 2048 x 2048 pixels de 16 bits, 8 Mo par image.

Plusieurs types d'images existent :

- des images brutes, conservées pour l'essentiel à Saclay, même s'il existe quelques exemplaires à Lyon ;
- les images réduites, conservées à Lyon ;
- des images de calibration, de différents types : obscurité et coupole – brutes ou réduites ;
- des images construites à partir d'autres images, afin d'en améliorer la qualité – nommées images composées.

L'image présentée est une vue du centre du Grand nuage de Magellan, une zone dense de 20 000 étoiles. L'image du bas est un gros plan de 20" d'arc montrant les étoiles (voir la thèse de P. Tisserand).

Le but des analyses : localiser chaque étoile sur les images réalisées nuit après nuit et mesurer leurs paramètres, sachant que chaque pointage, aussi précis soit-il, entraine de petites variations tant dans la direction que dans l'inclinaison des images. Par ailleurs, les variations météos compliquaient les observations et impactaient la qualité des observations.

L'expérience Eros II présentait la particularité de photographier la même zone du ciel simultanément avec deux caméras équipées chacune de 8 CCD de 2048 x 2048 pixels 16 bits. La lumière issue du télescope était divisée par un prisme en un flux bleu et un flux rouge. Chaque étoile est donc présente chaque nuit sur deux images

Chaque observation conduit donc à la production de 16 images FITS, soit 128 Mo. Une nuit standard permettait environ 1000 captures, soit 8 Go de données.

Le schéma de gauche montre la disposition des deux caméras. Le diagramme du bas le placement des 8 CCD (thèse de Patrick Tisserand).

Deux vues du télescope, avec un gros plan sur l'appareillage de prise de vue (Images JNA).

Les images issues des CCD sont nommées dans la terminologie Eros « images brutes ».

Ces images ne sont pas directement utilisées dans les analyses. Elles sont préalablement « calibrées » (on dit « réduites »). Cette réduction est réalisée sur le site grâce à des images de « calibration », acquises généralement au début ou en fin de nuit.

Le réseau de La Silla permettait – heureusement – l'envoi de courriers, mais ne supportait pas les transferts massifs. Les images étaient expédiées sous la forme de DLT vers Saclay. Les DLT brutes restaient à Saclay, les DLT réduites continuaient sur Lyon. Durant le temps du transfert, les données étaient conservées localement sur le site.

A Lyon, les images étaient recopiées sur des 3480/3490 (800 Mo) et référencées dans une base de données Oracle avec leurs principales caractéristiques.

Avec l'arrivée du HPSS, les images ont été déplacées vers le robot, puis à l'arrêt de l'expérience, regroupées sous la forme d'archives Tar de grandes tailles.
Les données et la base de données sont donc restées 15 ans en sommeil.

La première phase de la préservation des données Eros a été la recopie des images vers Irods, un système de fichiers distribué, accessible largement et protégé par mot de passe.

Une vue du poste de pilotage du télescope Marly avec les deux champs rouge et bleu (Image JNA).

La migration des données des archives Tar du HPSS vers Irods a montré différentes pertes :

- Les images brutes n'ont pas été sauvées par Saclay.
  - Il est difficile de savoir si les DLT sont encore disponibles, ni comment elles auraient été conservées.
  - Il semble encore plus difficile d'espérer pouvoir les récupérer.
  - Cela représente environ 2 millions d'images – peut-être plus : les images brutes n'étant jamais passées par Lyon, il n'y en a aucune trace dans la base de données…
  - Et au moins 1 600 DLT…

- Des cartouches du HPSS ont cassé, surtout durant une phase délicate qui a duré plusieurs années. Les pertes ne sont toutefois pas énormes, quelques milliers d'images, soit quelques %.

- Un programme est totalement perdu ("Naines rouges"), soit plus de 60 000 images (4 %).
  - Aucune explication à ce stade.
  - Mais grâce à la base de données, on sait que ces images ont existé puisqu'elles ont été référencées à leur entrée à Lyon.

Les images brutes n'étant plus accessibles, aucun espoir de récréer les images perdues.

L'analyse des images était conduite grâce au logiciel Peida, développé pour les besoins de l'expérience. Plusieurs étapes étaient nécessaires.

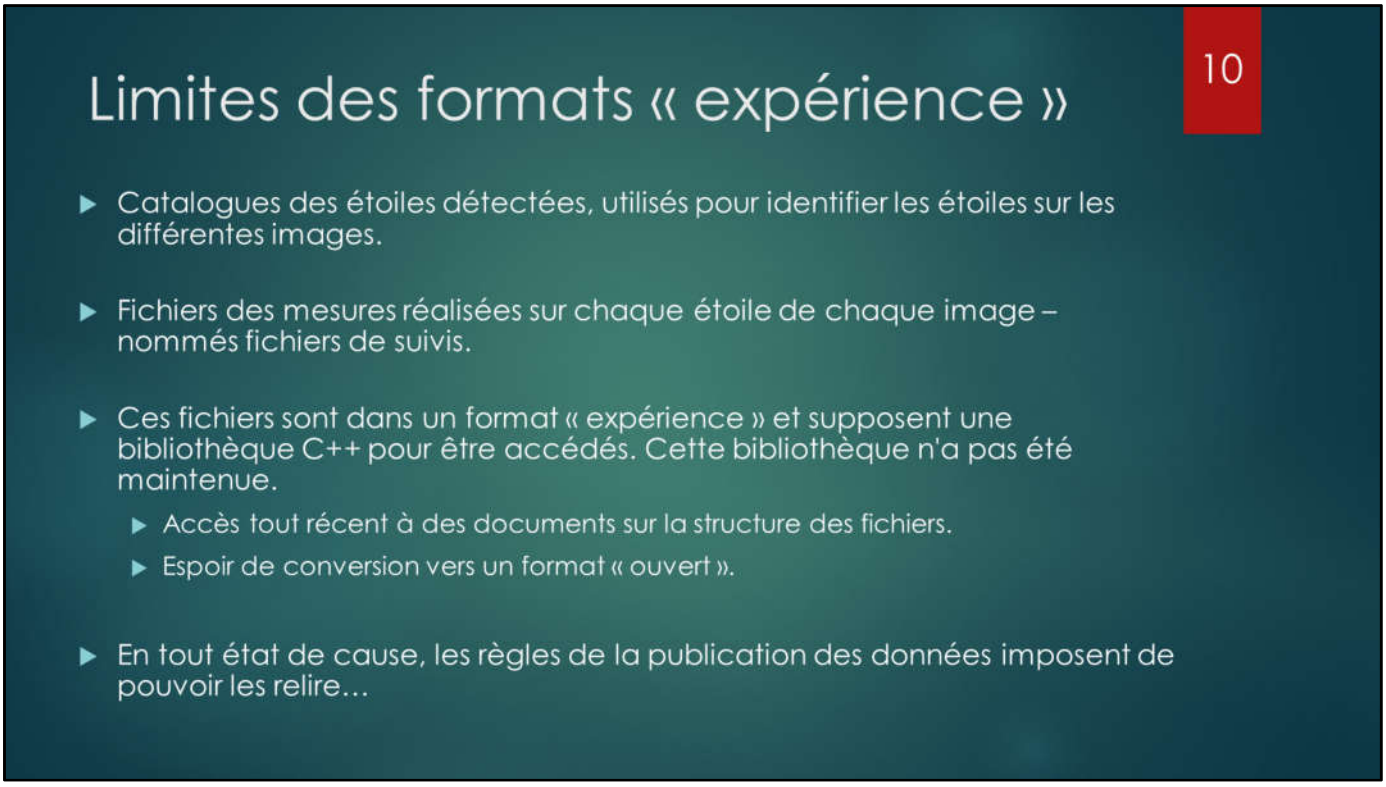
La première étape est l'identification des étoiles observées. Pour améliorer l'efficacité du programme de reconnaissance, une série de quelques bonnes images est sélectionnée pour chaque zone à analyser afin de créer une image dite « composée ». Ces traitements ayant été réalisés au Centre de calcul, ces images sont disponibles dans Irods.

L'identification des étoiles conduit à la création de catalogues, conservés sous la forme de fichiers dit « fichiers de références », ou plus simplement « références ». Ces fichiers utilisent un format binaire propre à l'expérience. Ils ont été créés au Centre de calcul et ont été sauvés dans Irods.

La détection des étoiles est conduite indépendamment pour les images bleues et les images rouges, d'où l'existence de « références bleues » et de « références rouges ». Une ultime phase, dite « association », reprenait le catalogue bleu et le catalogue rouge d'une même zone et associait les étoiles présentes dans les deux catalogues.

La seconde grande phase des traitements est la mesure des paramètres des étoiles présentes sur les différentes images réduites. Les étoiles analysées sont celles décrites dans les catalogues. Les résultats sont conservés dans des fichiers dit « de suivis », également dans un format binaire propre à l'expérience. Pour une même zone du ciel, il y a donc des « suivis bleus » et des « suivis rouges ». Les traitements sont segmentés par groupes de nuits. Le but était de réduire la taille des fichiers et de permettre une analyse progressive des mesures, au fur et à mesure des arrivées, tout en minimisant les mouvements de fichiers entre les cartouches et les disques.

Plusieurs campagnes de traitement ont été réalisées. La référence « officielle » est la production P5, couvrant plus de 70 % des mesures réalisées.



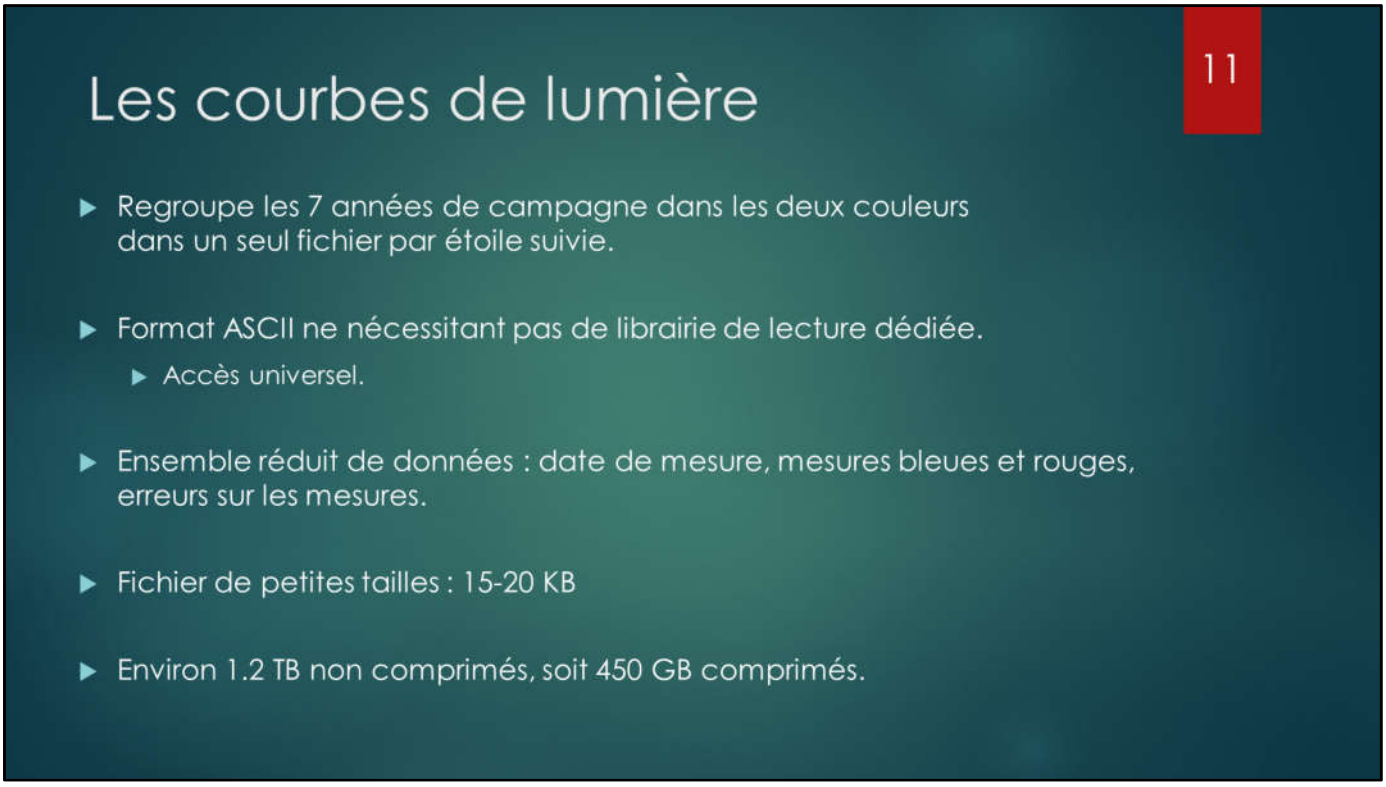
Les résultats des mesures réalisées sur les étoiles sont conservés dans deux types de fichiers : des catalogues des étoiles détectées, nommés *fichiers de références*, et les fichiers des mesures, ou *fichiers de suivi*.

L'accès à ces fichiers nécessite une bibliothèque C++ qui n'a pas été maintenue. Jusqu'à cette semaine, ces fichiers pouvaient être classés dans la catégorie des fichiers « perdus » - bien *qu'inaccessibles* serait plus approprié…

Cela représente tout de même 70 000 suivis et 15 000 références pour un total de 7 To. Les fichiers de la production P5 ont été transférés dans Irods. Les autres productions semblent pouvoir être abandonnées.

J'ai fini par avoir accès à des documents sur la structure interne – au moins des suivis. Il y a donc un espoir de pouvoir les convertir dans un format « ouvert » - restant à définir. Après tout, Java peut tout !

Les règles de la publication des données en astronomie imposent de pouvoir relire les données. On ne peut donc pas publier les fichiers de suivi s'ils ne sont pas lisibles.



Les fichiers de suivi étaient bien adaptés au traitement massif des images. Ils étaient callés sur la structure des observations et permettaient de minimiser les mouvements de fichiers entre les cartouches et les disques – souvenons-nous de l'état de la technologie il y a 20 ans.

Mais pour analyser *une* étoile, il fallait redescendre les deux catalogues d'étoiles, les différents blocs de suivis et ce dans les deux couleurs. Et maitriser le C++. Sur les deux plateformes du CC : IBM et HP. Plus les DEC du LAL. Puis des SUN…

Des contraintes un peu rudes pour les barons de l'expérience qui ont obtenu que les mesures soient converties en fichiers ASCII, à raison d'un fichier par étoile, mais contenant toutes les mesures de la campagne dans les deux couleurs. Ces fichiers ASCII, ou *courbes de lumière*, sont le seul résultat pérenne, avec les images FITS, de l'expérience. Ces courbes de lumière présentent en outre l'intérêt d'être de petites tailles – de 15 à 20 Ko.

Plusieurs campagnes de production ont été réalisées, plus ou moins abouties. Les deux plus importantes, en termes de suivis générés, sont la production P1 et la production P5. La production considérée comme *la production officielle* est la production P5. C'est à partir des fichiers de suivi de cette production que les courbes de lumière ASCII ont été produites.

Les paramètres des étoiles sont référencés dans des fichiers également ASCII constituant des catalogues ASCII d'étoiles, ou *catalogues*.

Chercher les mesures d'une étoile consiste donc à trouver l'étoile par ces coordonnées astrométriques dans les catalogues et récupérer la courbe de lumière.

Les catalogues et les courbes de lumière ont été transférés vers Irods. Les courbes de lumière sont conservées sous la forme d'archives Tar comprimées correspondants aux différents secteurs observés. L'intérêt de Irods est qu'il supporte le format Tar et que les fichiers peuvent être accédés « directement ». Mais pour un accès global à toutes les étoiles d'une zone, il est bien sûr plus efficace de recopier toute l'archive.

Les catalogues sont également conservés dans la base de données, ce qui permet de réaliser des recherches

directement *via* un accès à la base de données – au travers d'un outil (Java) qui assure également le transfert des courbes de lumière, images, ...

La base de données Oracle a été mise en œuvre dès le début d'Eros II. Elle constitue un catalogue de l'ensemble des fichiers de l'expérience – du moins de ceux qui sont passés par Lyon – et des principaux paramètres des données.

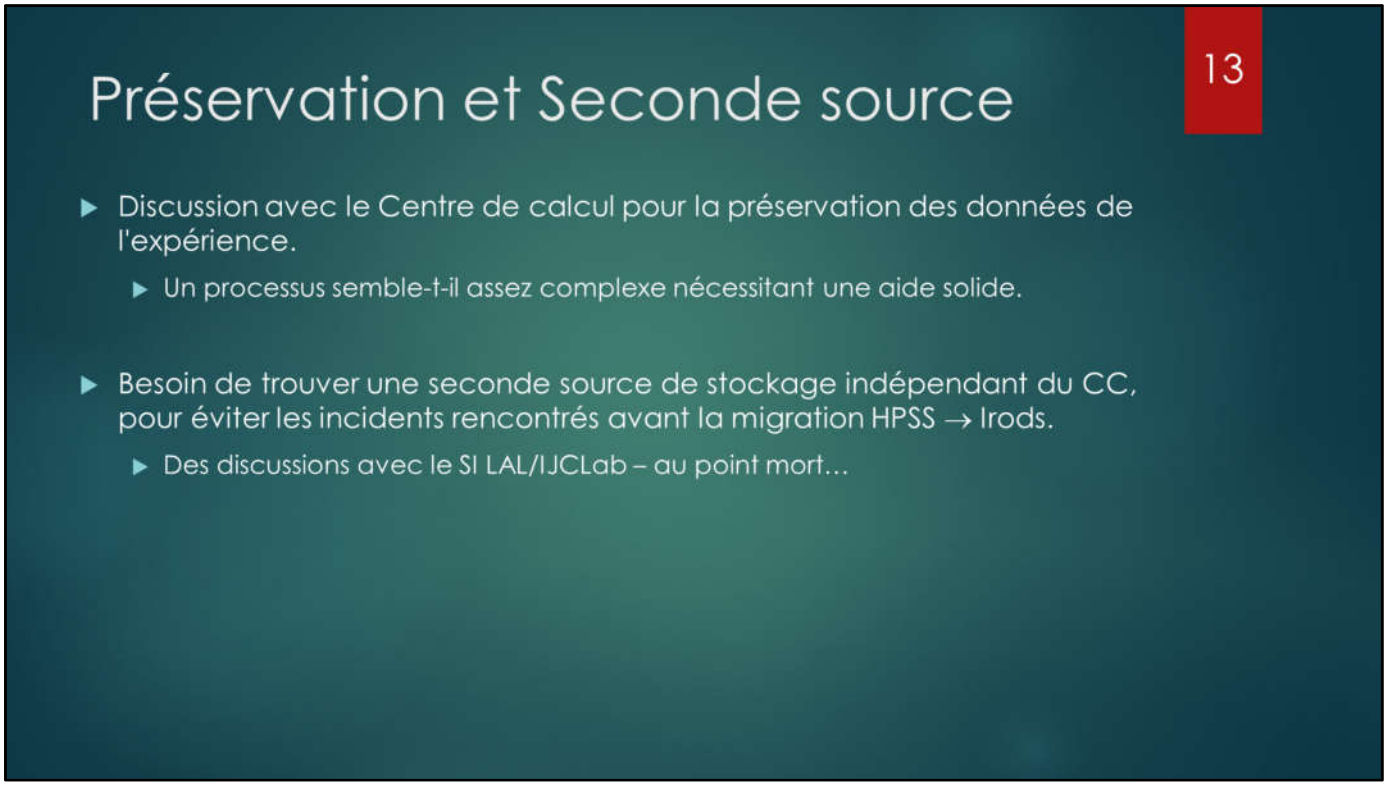
Les informations sur les images sont issues de leur entête FITS. L'intérêt du format FITS – outre son aspect de standard de fait de l'astronomie – est sa capacité à supporter des métadonnées décrivant entièrement le format du fichier.

Les informations sur les catalogues binaires, ou fichiers de référence, et sur les fichiers de suivis proviennent des logs des jobs qui les ont produit. Hélas, ces logs n'ont pas été conservés.

Malheureusement, les courbes de lumière ASCII et les catalogues d'étoiles ASCII ont été créés en dehors du schéma de production « officiel » et ne sont donc pas référencés. Il est donc difficile de contrôler les résultats des migrations de HPSS vers Irods – contrairement aux autres données, indexées dans la base de données.

La base de données sert également de support pour différents outils de recherche : sélection d'images (et de suivis), identification des courbes de lumière, …

Tous les outils de gestion et de recherche des données sont écrits en Java et utilise le « pont objet/relationnel » Hibernate ainsi qu'une intégration de Irods aux systèmes d'accès aux fichiers de Java, NIO, *via* la librairie Java d'accès à Irods, Jargon.



Les pertes de données mises en évidence lors de la migration de HPSS vers Irods a montré la nécessité de disposer d'un second espace de stockage pour la préservation des fichiers (~25 To) dans un espace de stockage protégé, indépendant du Centre de calcul.

Des discussions avec le SI du LAL et Michel Jouvin ont débuté, mais ne progressent guère.

Des discussions ont aussi lieu avec le Centre de calcul qui est disposé à piloter nos tentatives de publier les données de l'expérience. Mais ici encore, les progrès sont maigres. Mais ces discussions ont au moins permis de stabiliser les données existantes.

Références et URL

14

- 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/

Un message aux ingénieurs : allez sur le terrain – participez à la vie des expériences. Ce sont des occasions uniques de connaitre et de comprendre les contraintes auxquelles les physiciens sont confrontés jour après jour.