DC 2013

Philippe Gris
LPC Clermont-Ferrand

Production

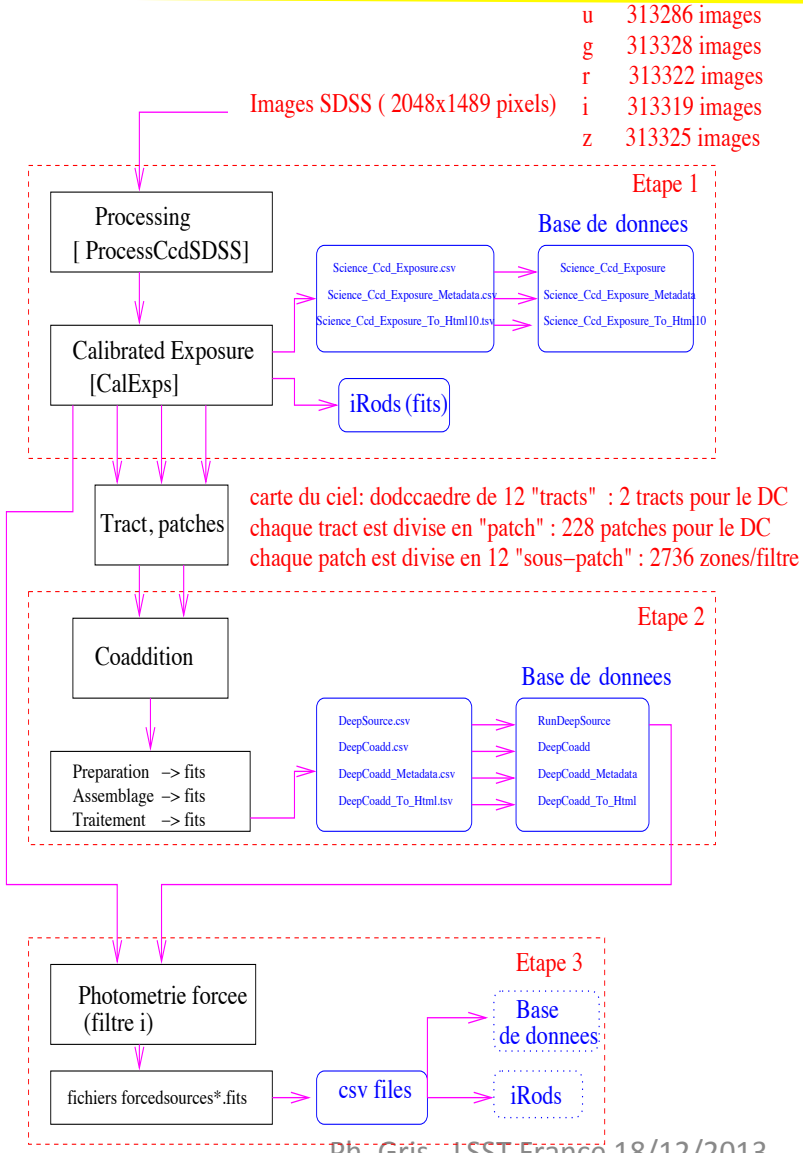
DC 2013: la production

- Objectif: traiter une partie des images SDSS correspondant au Stripe82 dans le software LSST:
  - 5 < RA < 55
  - -1.26 < DEC < 1.26
  - Cinq filtres: u,g,r,i,z
  - Zone de recouvrement avec les USA: 5 < RA < 10

- A notre disposition, nous avions:
  - Les données – copiées dans iRods depuis un serveur SDSS.
  - Le software LSST (release v7_2)

- Plusieurs étapes ont conduit à la mise en œuvre du DC:
  - Série de TP (3) pour exécution/compréhension de chaque étape de la production (entrées/ sorties)
  - Mise en place de scripts pour chaque étape de la production -> test à 1% + comparaison USA
  - Mise en place d'une structure permettant une production massive -> test à 1% +comparaison USA
  - Traitement de toutes les données pour toutes les étapes de production: 31/07 -> 27/09

DC 2013: la production



Traitement des images

Coaddition

Photométrie contrainte

DC 2013: la production

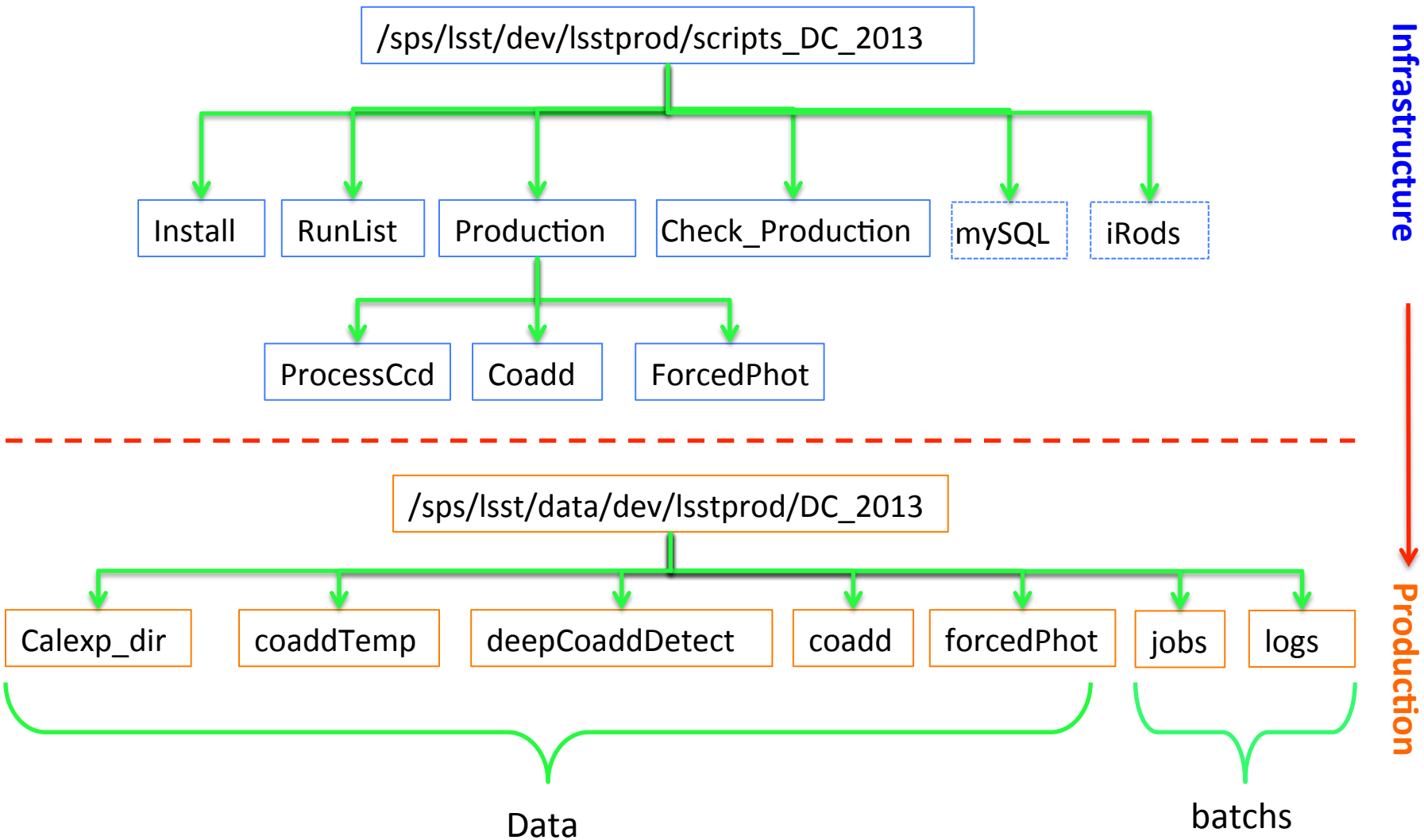
- *Mise en place d'un structure permettant une production massive:*
  - *Création d'un cadre de travail paramétrable, évolutif, utilisé comme base par les utilisateurs participant au DC*
  - *Ensemble de scripts (shell et python) permettant de gérer la production:*
    - *Script pour définition de l'environnement*
    - *Soumission massive de jobs (paramètres définis en partie à partir du test 1%)*
    - *Vérification des résultats pour les différentes étapes*
  - *Historique de toutes les activités répertorié dans un elogbook (http://lsst.in2p3.fr/wiki/index.php/Logbook_DC_2013_new).*

- *Détails techniques:*
  - *Cadre de base: /sps/lsst/dev/lsstprod/scripts_orig*
  - *./Prepare_To_Run.sh fichconfig.txt*

```
MAINDIR /sps/lsst/dev/lsstprod
DCNAME DC_2013
DATAMAIN /sps/lsst/data/dev/lsstprod
DBHOST ccoratest03.in2p3.fr
DBUSER lsst_prod
```

Suffisant pour créer toute l'arborescence nécessaire à la production

DC 2013: la production



DC 2013: la production

- Nombre de jobs nécessaires (par filtre)

Etape	Njobs	Info
ProcessCcd	1254	production
	63	vérification
	23	copie DB
	6	copie iRods
Coaddition	1368	production
	28	vérification
	1	copie DB
forcedPhot	2089	production
	224	création csv
	1	copie csv DB
	5057	

~ 40000 jobs soumis
~ 1 million d'heures HS06

Très bonne réactivité du centre de calcul pour adapter une puissance de calcul adaptée à nos besoins.

-> cf présentation de Rachid

Gestion des jobs de façon « standard » (qsub) + scripts bash et python, sauf pour la production des calexps en filtre z (Dirac: cf présentation de Johann)

DC 2013: la production

- Vérifications systématiques des résultats des batchs:
  - Pour chaque étape: nb traités = nb evts entrée ?
  - Vérification à partir des fichiers logs (pas optimal)
  - A permis de détecter divers problèmes:
    - Crash de batchs -> resoumission possible
    - Non traitement de certains événements

- Exemple : vérification de l'étape ProcessCcd, filtre :

Events processed: 313319 out of 313319:

ok: 309324 - 98.7 %

astrometry problem: 2131 - .680 %

matches problem: 0 - 0 %

missing files: 674 - .215 %

Wrong number or type of arguments for overloaded function 'new_ExposureF': 1188 - .379 %

Wrong number or type of arguments for overloaded function 'new_MaskedImageF'.: 0 - .000 %

Another problem: 1 - .000 %

- Cette vérification est quantitative et ne préjuge en rien de la qualité des données produites.

DC 2013: la production

- *Quantité de données produites*

Fichiers SDSS à traiter: ~ 7.1 TB

Etape	Taille	format	Stockage
ProcessCcd	43.5 TB	fits	iRods
	9.1 GB	csv	DB
Coadd	57.2 GB	csv	DB
	1.2 TB	fits+boost	iRods
	119 GB	csv	iRods
forcedPhot	2.4 TB	fits	iRods
	6.3 TB	csv	DB

DC 2013: la production

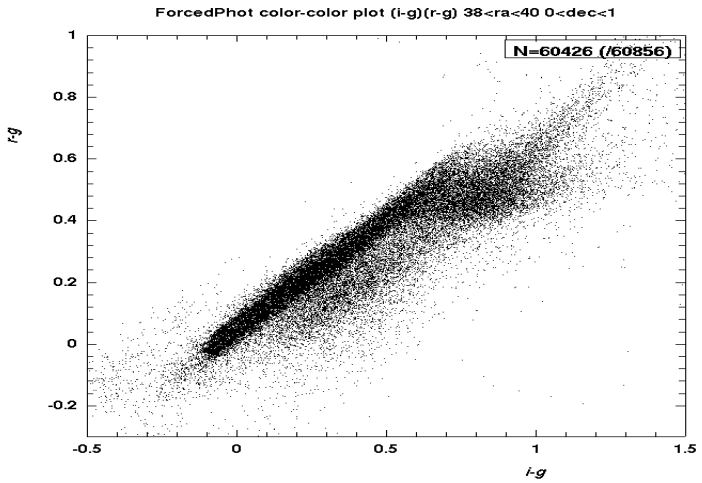
- Problèmes rencontrés
  - Au début du DC: utilisation non optimale des ressources mises à disposition par le CC.
  - Stockage dans iRods: mauvaise procédure de copie – corrigée par la suite
  - Gestion difficile du nombre de fichiers produits
  - Lecture de la base de données à l'étape de la photométrie contrainte:
    - Nombre de connexions simultanées possibles insuffisant
    - Problème de désactivation des indices de la table DeepSource (vue mySQL de RunDeepSource)
    - Altération de la base de données (après étape précédente)-> nouvelle paramétrisation nécessaire
    - Un travail complexe au niveau de la base de données est toujours en cours (création d'un unique vue à partir de 5 bases - Cf présentation d'Osman)

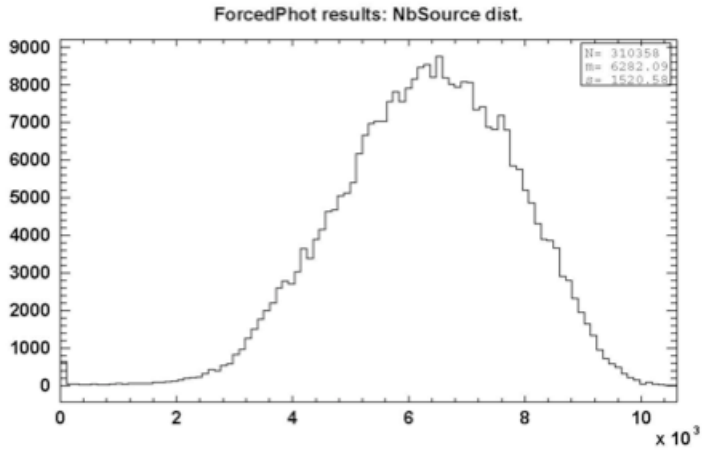
-> Problèmes résolus rapidement et efficacement grâce aux promptes interventions du CC (Rachid, Jean-Yves, Loïc, Osman)

Analyse

DC 2013: Analyse

- Quantité de données accessibles (DB et iRods) pour chaque étape de la production (ProcessCcd, Coaddition, photométrie forcée).
- Qu'en est-il de la qualité des données produites ?
- Plusieurs voies possibles:
  - Comparaison avec les résultats de nos collègues américains (zone de recouvrement)
  - Comparaison avec les résultats SDSS (traitement des mêmes images)
- Développement d'outils de visualisation/comparaison.





Réza

DC 2013: Analyse

Comparaison aux données SDSS (Bogdan)

Extract data for a patch from the DC2013@CCIN2P3

- find which patches cover a given region (overlap CC/ NCSA):

  SELECT tract, patch FROM DeepCoadd WHERE ra > 6.0 and ra < 7.0 and decl > -0.2 and decl < 0.0;

- choose the patch: 0-380,5 (from 20) and get the bounding coordinates:

  SELECT corner1Ra, corner1Decl, corner3Ra, corner2Decl FROM DeepCoadd WHERE tract = 0 AND patch = '380,5';

- query the DB (lsst_prod_DC_2013_2):

  SELECT ra, decl, ravar, declvar, psfFlux, psfFluxSigma, modelFlux, modelFluxSigma, filterId FROM DeepSource WHERE ra > 6.27664 AND ra < 6.50324 AND decl > -0.21301 AND decl < 0.00335 LIMIT 100000

---> 35117 sources (5484-u, 10688-g, 6970-r, 8855-i, 3120-z)

Extract data for a patch from the SQL Stripe82 tool

- http://cas.sdss.org/stripe82/en/tools/search/sql.asp

- SELECT TOP 100000 psfMag_u, psfMag_g, psfMag_r, psfMag_i, psfMag_z, u, g, r, i, z, ra, dec, run FROM PhotoPrimary WHERE ra > 6.27664 and ra < 6.50324 and dec > -0.21301 and dec < 0.00335

- See the schema browser at: http://cas.sdss.org/stripe82/en/help/browser/browser.asp

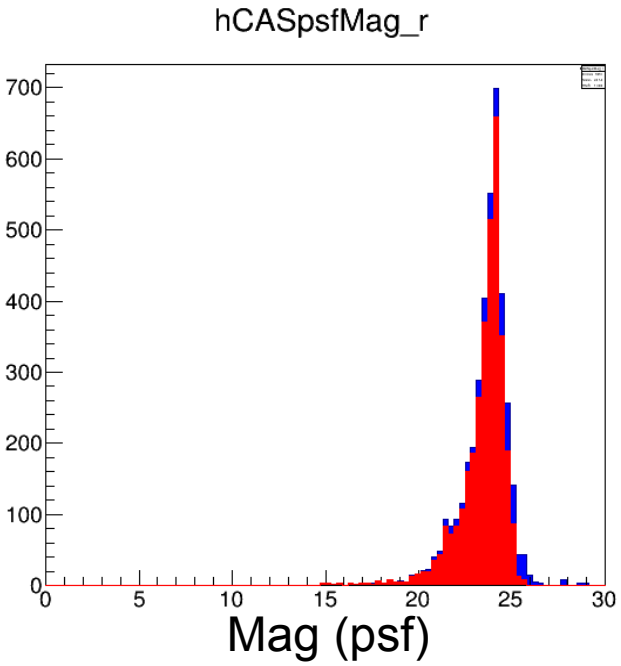
  ---> 85893 objects

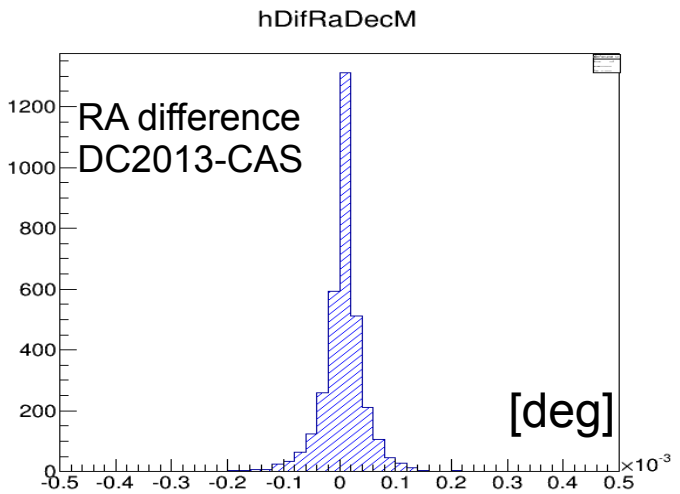
- Select in the analysis only the coadd runs 106 (South strip) and 206 (North strip):

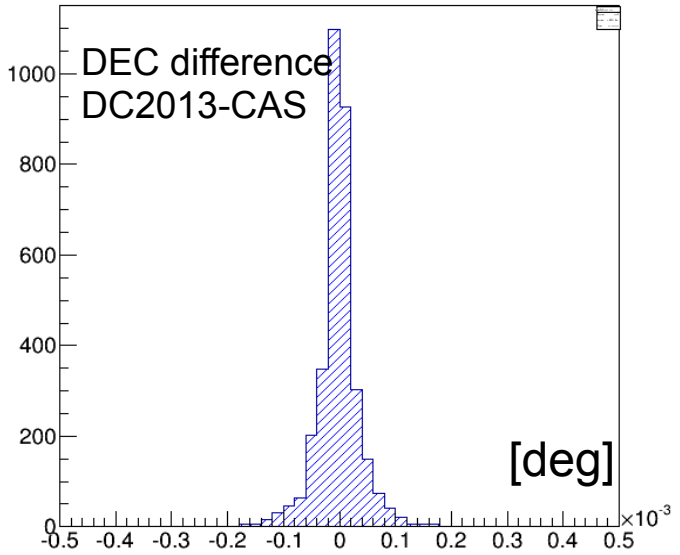
  ---> 3867 objects

Comparaison DC2013@CCIN2P3 avec CAS Stripe82 filtre “r”

Matching: RA/DEC/MAG
RA et DEC :+/- 0.0005 deg = 0.12 arcsec
MAG +/- 1
3867 objets CAS Stripe82
3271 source unique DC2013
(from 5639 variance sel.)





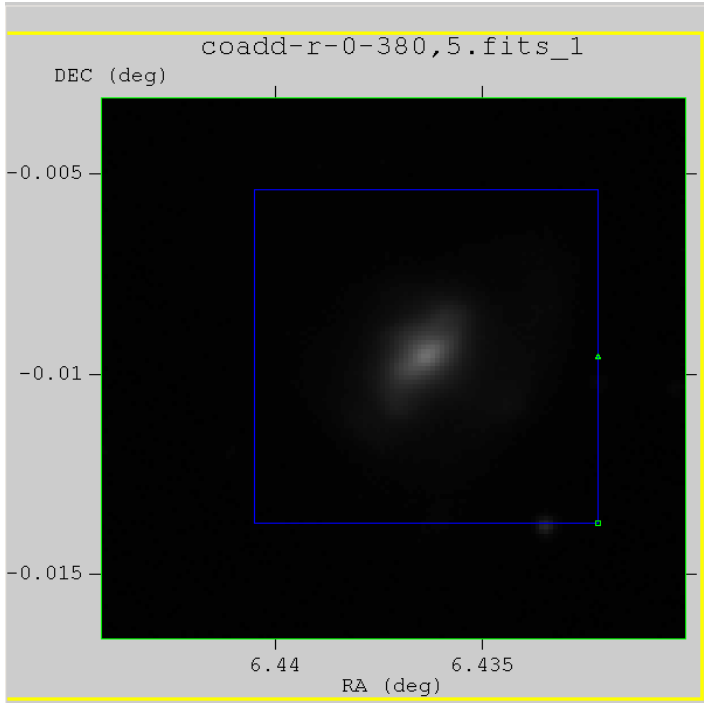


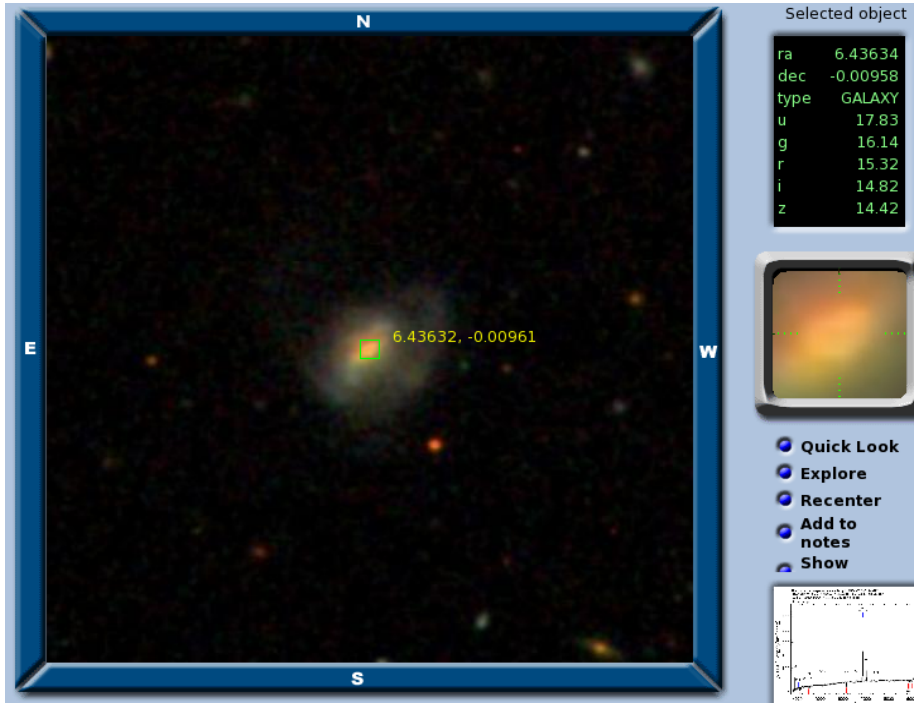
DC 2013: Analyse

Unmatched source (psfMag < 20, filter “r”)

RA = 6.436371
DEC = -0.009552
MAG = 17.52

What is in DR9 SkyServer ?
MAG = 15.32





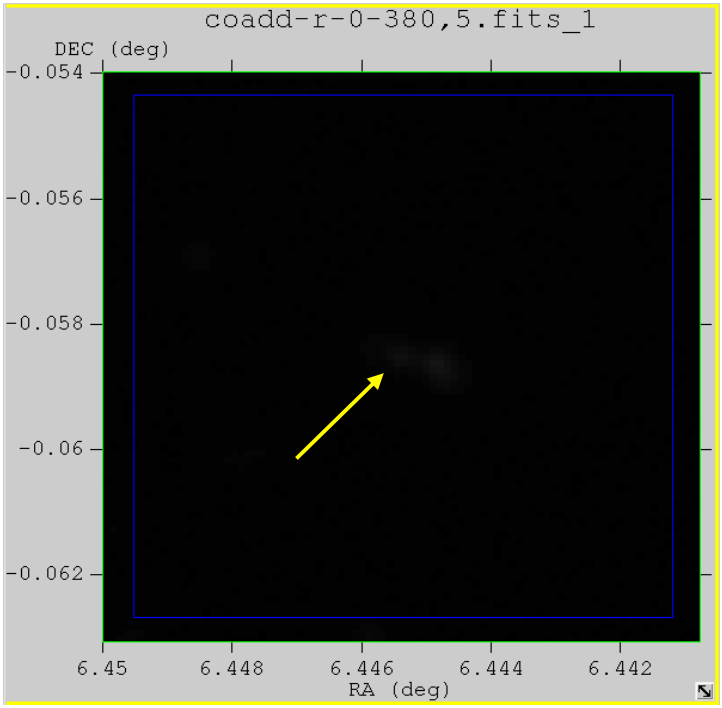
DC 2013: Analyse

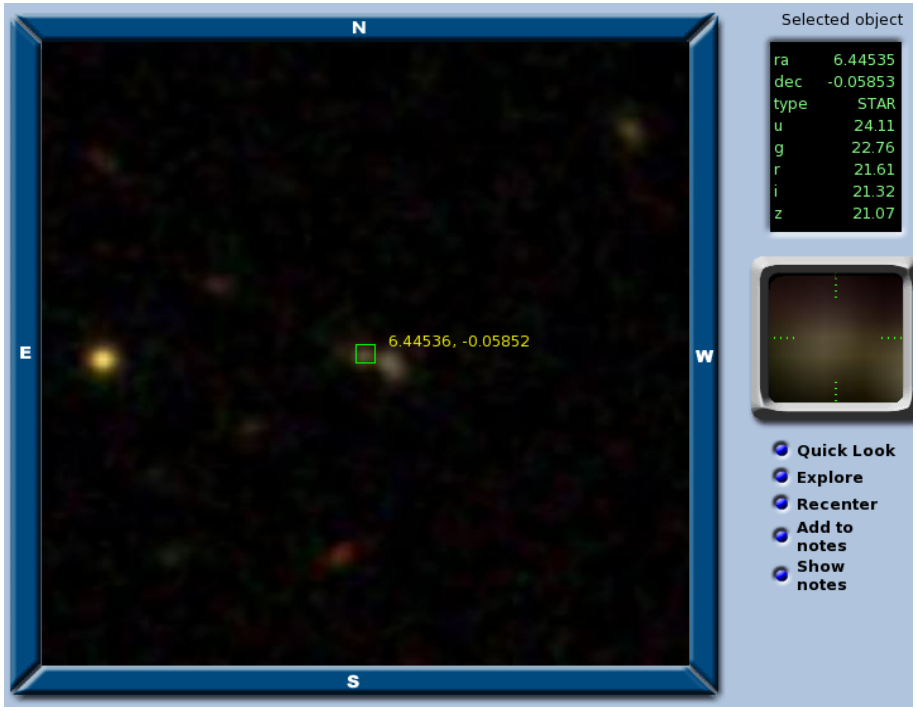
Match only in RA and DEC

DC2013:
RA = 6.445361
DEC = -0.058519
MAG = 21.57

CAS Stripe82:
RA = 6.444815
DEC = -0.058653
MAG = 20.92

SkyServer DR9:
RA = 6.44535
DEC = -0.05853
MAG = 21.61





DC 2013: conclusion

- Production:
  - Passage à l'échelle difficile en l'état actuel de la structure du soft:
    - Gestion du nombre de jobs difficile -> utilisation d'outils appropriés (p. ex. Dirac -> talk de Johann)
    - Vérification des différentes étapes compliquée : scan des fichiers log: long – (batch)
  - Une optimisation de l'enchaînement des étapes devrait être possible:
    - par exemple:
      - en structurant les jobs de façon à traiter les 3 étapes successivement et non séparément (en produisant les calexps par (tract, patch) par exemple)
  - D'une façon plus générale, ce DC2013 fut très instructif et nous a montré qu'une préparation plus approfondie et des échanges plus nombreux avec nos collègues des USA étaient nécessaires (l'importance de la base de données mySQL a été complètement sous-estimée).
- Analyse:
  - Activité qui débute et se développera probablement
  - Nécessaire pour une validation complète du DC et pour la compréhension des données -> physique !
  - La base de données SDSS doit être accessible dans sa totalité pour des études statistiques cohérentes (copie en cours à Clt-Fd)