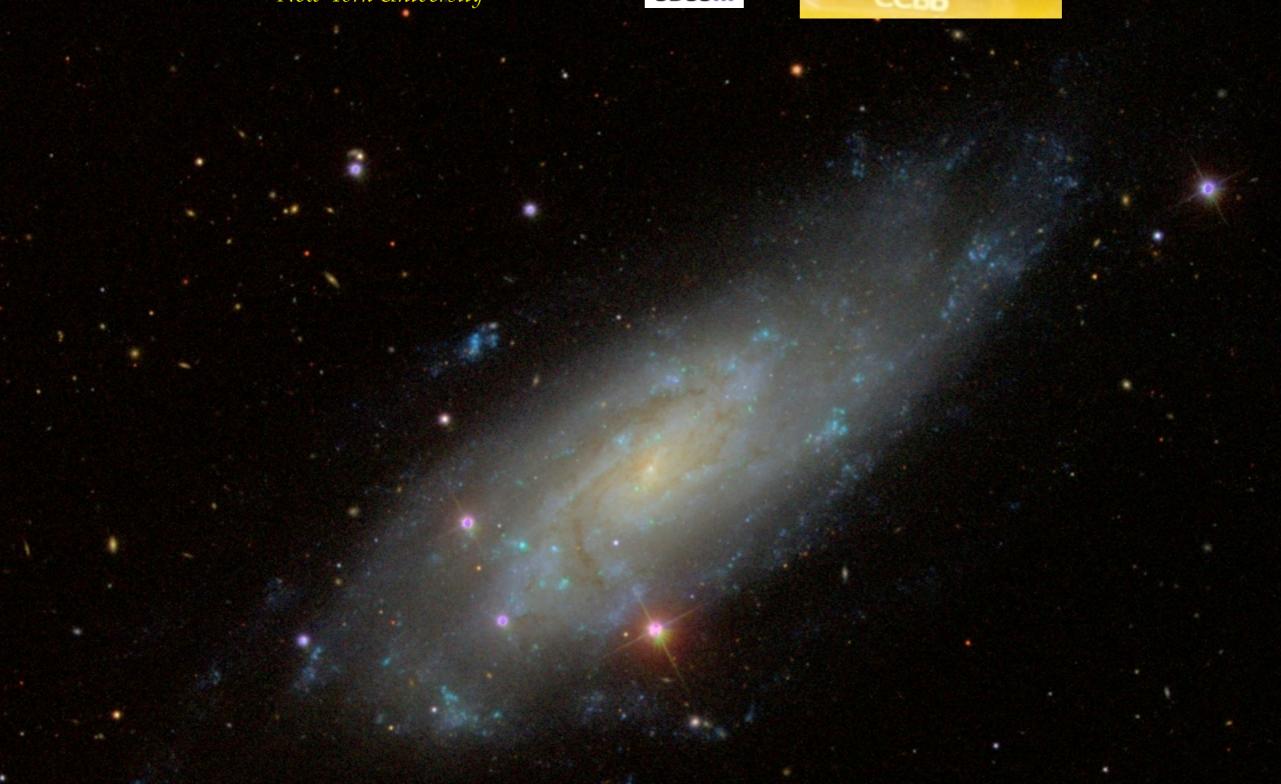
SDSS-III Data Access

Michael Blanton, SDSS-III Data Coordinator Center for Cosmology and Particle Physics New York University









Key people for DR8

1. Science Archive Server team:

Ben Weaver, Demitri Muna, Adrian Price-Whelan (NYU), Tom Throwe (BNL), Ricardo Ogando and BPG team (Observatorio Nacional)

2. Catalog Archive Server team

Ani Thakar, V. Paul, N. Buddana, V. Vamsi, A. Szalay, G. Fekete, N. Li, T. Budavari (JHU)

3. SEGUE-2 team

Fergal Mullally, Steve Bickerton, Craig Loomis (Princeton), Connie Rockosi (UCSC), Phoebe Stierhoff (Case), Brian Yanny (Fermilab)

4. BOSS imaging team

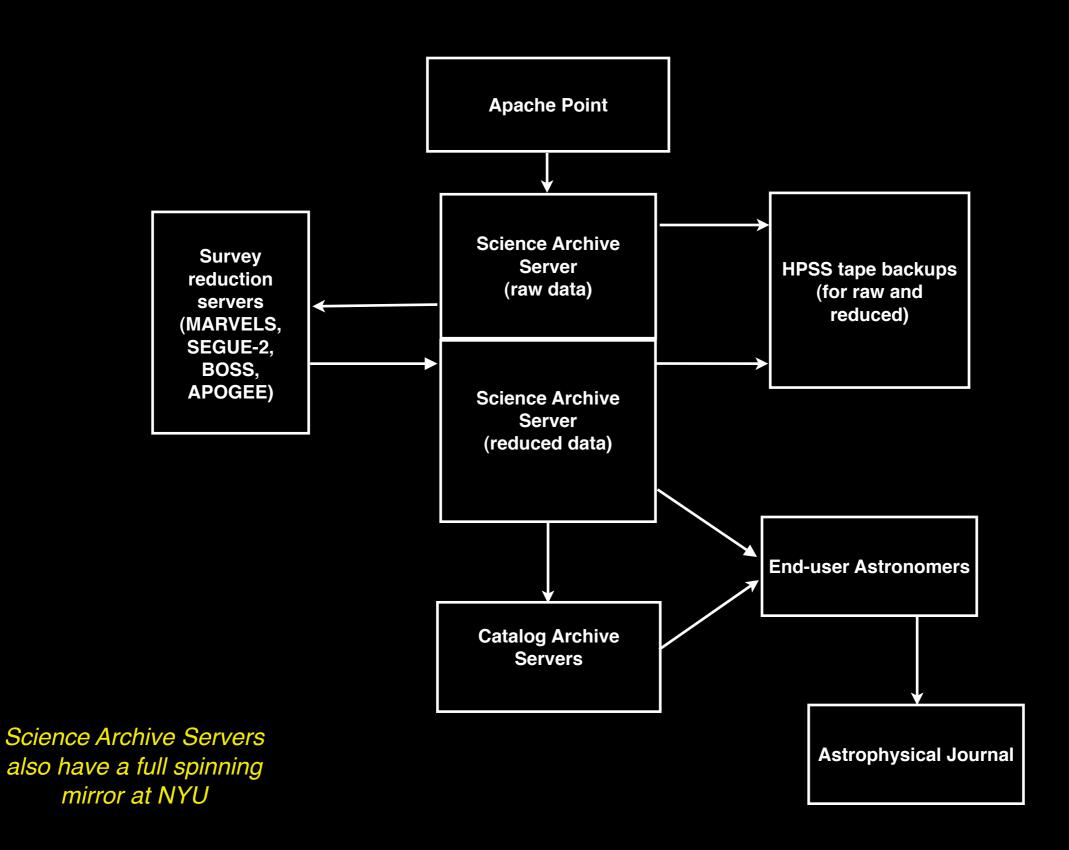
David Schlegel, Gary Kushner (LBL), Nikhil Padmanabhan (Yale)

5. Documentation team

Natalia Connolly (Hamilton), Jordan Raddick (JHU), Michael Strauss (Princeton), Bob Nichol (Portsmouth)



Data distribution for SDSS-III





SAS & CAS

Science Archive Server (currently running)

all data from all surveys
flat files (FITS, FTCL formats)
full dynamic range images
full dynamic range spectra
raw data and complete meta-data
rsync, wget access to full directory trees
[web form access for simple searches]
full data model with detailed descriptions of formats

SDSS-III SAS

| Author | Certing Started | Latest Headlines | Author | Comple | Comp

Catalog Archive Server (hardware installed, testbed version running)

SQL database with catalogs
Form interface for simple queries
CASJobs and MyDB servers for complex queries
SkyServer\browsing of images, cutouts and spectra
Attached to web page with full user documentation





Data processing and data releases in a nutshell

1. Data flow

- a. All data comes off mountain to SAS, backed up to tape and on mirror at NYU
- b. BOSS data processed on SAS by BOSS team at LBL
- c. SEGUE-2 data processed at Princeton, mirrored to SAS for data access
- d. APOGEE data processed at UVa, mirrored to SAS for data access

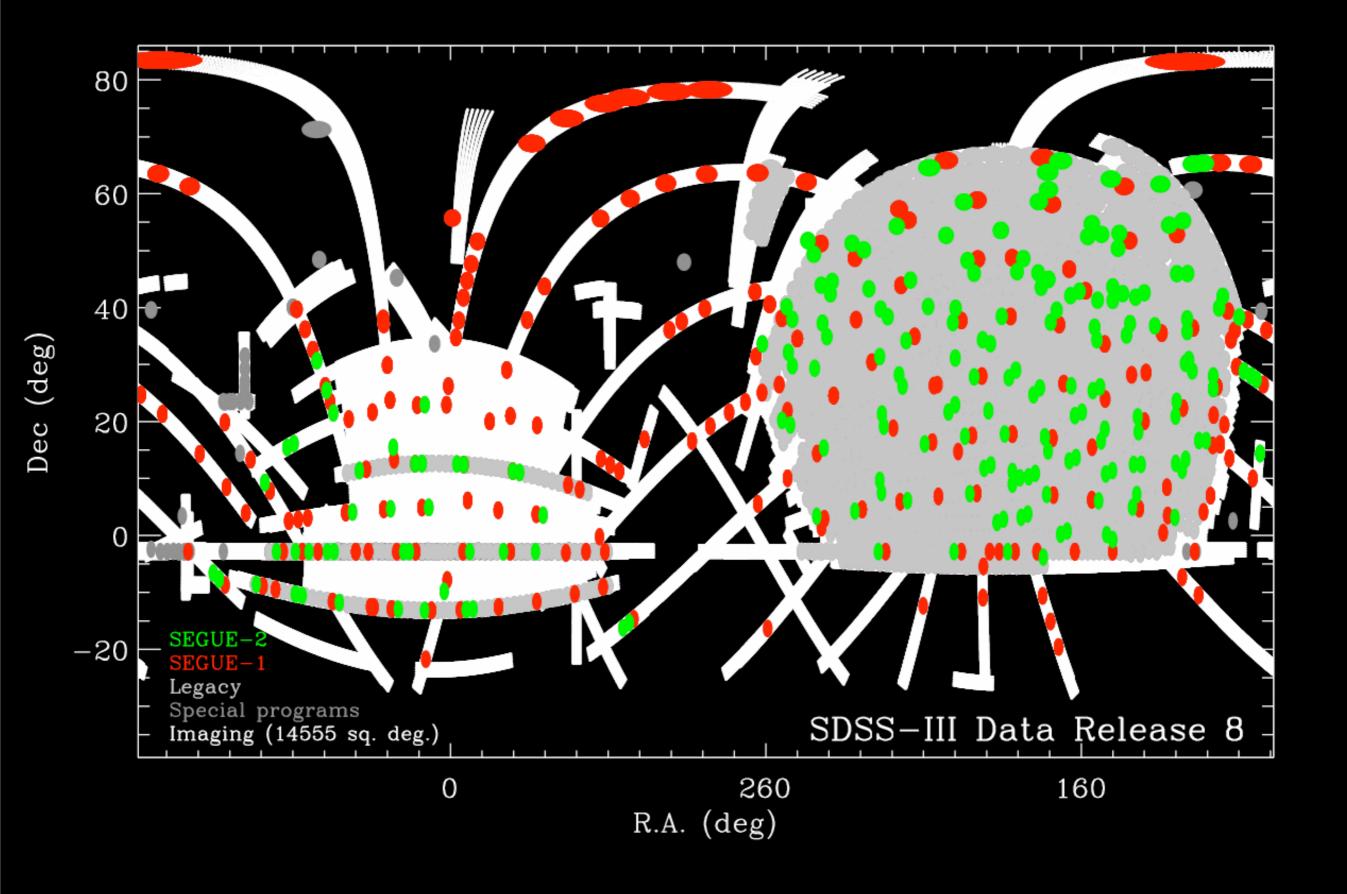
2. Archive structure and contents

- a. Web site and documentation: http://www.sdss3.org
- b. Catalog Archive Server (CAS) for catalogs at JHU: http://skyservice.pha.jhu.edu/casjobs/
- c. Science Archive Server (SAS) for images and spectra at LBL: http://data.sdss3.org

3. Data release schedule

Date	Data Release	APOGEE	BOSS	MARVELS	SEGUE-2
Dec. 2010	DR8		Final imaging		Final spectra
Jul. 2012	DR9		•	Radial velocities (up to Jul. 2011)	
Jul. 2013		_	Spectra (up to Jul. 2012)		
Dec. 2014	DR12	Final spectra	Final spectra	Final radial velocities	

DR8 Scope





How is DR8 different from DR7?

1. What is different about the imaging data?

- a. 3000 deg more imaging data in SGC
- b. Imaging data has all been rerun through latest PHOTO v5_6
- c. Reductions, calibrations, "resolve" are under photo-op system (differs from SDSS-2)
- d. Corrected frames now calibrated and sky-subtracted, with correct WCS
- e. Photometric "datasweeps"
- f. 2MASS matches added

2. What is different about the spectroscopy?

- a. A handful (102) old plates not in DR7 now released
- b. SEGUE-2 took 211 plates, or 135,040 more spectra
- c. Spectroscopic data also rerun through SSPP
- d. Galaxy spectroscopy has "Garching" analysis
- e. Target flag overloading resolved (primtarget -> legacy_target, special_target, segue1_target)
- f. Matches between spectra and photometry use flux information

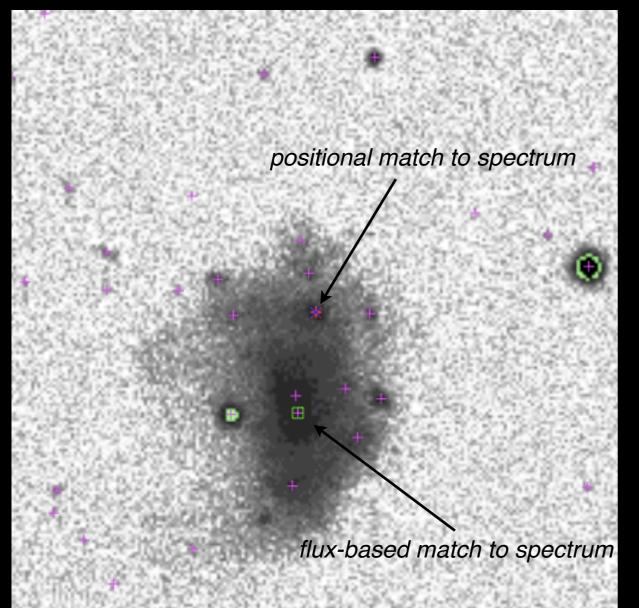
3. What is different about the CAS?

- a. Not much to user, we are trying to keep units, naming conventions the same
- b. Adding geometrical information regarding the photometric window
- c. Pulling some calculations out (e.g. matching) to ensure SAS/CAS are identical



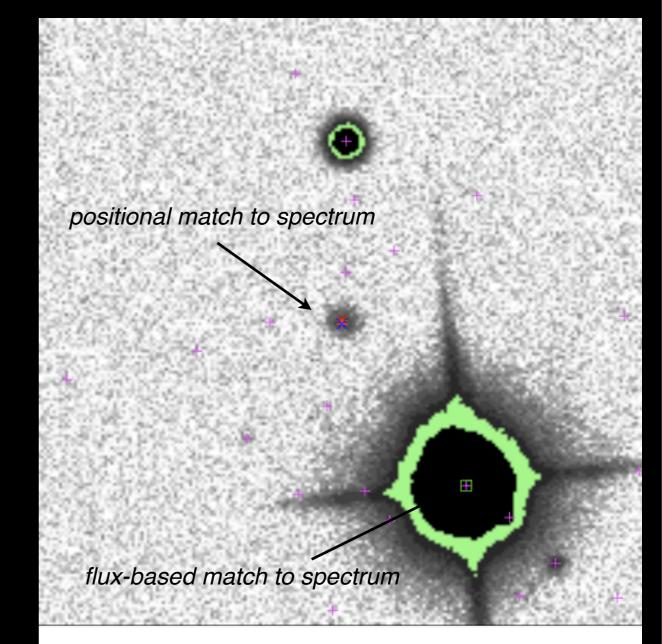
Flux-based matching

Successes (1-2 per plate) recover nearby galaxy matches much better.



- + All unique objects
- * Location to match (J083336.33+293007.2)
- × Positional match 3560/301/5/142/310
- □ Flux-based match 3560/301/5/142/311 flagged because aperflux=4.346 << fiberflux=12.599 used brightest child, with flux=9.193 out of 14.003

Failures (1-2 per plate) reflect errors in photometry in all cases I've seen; e.g. in this case photo underestimates flux of galaxy by more than a factor 10, due to deblending



- + All unique objects
- * Location to match (J084045.91+295200.3)
- × Positional match 3606/301/3/43/60
- □ Flux-based match 3606/301/3/43/56 flagged because aperflux=0.488 << fiberflux=7.039 used brightest child, with flux=6.350 out of 7.297



DR8 & BOSS Data Access Summary

- 1. Science Archive Server:
 - 1. All raw and archived reduced DR8 data available through wget/rsync
 - 2. Interactive web interface under development (see BOSS)
- 2. SEGUE-2 data: reductions finished Fall 2009, targeting finished Spring 2010
 - 1. SAS flat files described at https://sdss3.org/internal/branches/v3/dr8/spectro/spectro catalog access.php
 - 2. CAS interface at http://skyservice.pha.jhu.edu/casjobs/ (use SEGUE2 context)
 - 3. Spectrum tools at https://trac.sdss3.org/wiki/SEGUE2/DataAccess
- 3. BOSS imaging: reductions & calibrations finished July 2010
 - 1. SAS flat files described at https://sdss3.org/internal/branches/v3/dr8/imaging/imaging_access.php
 - 2. Testbed version of data available since Aug. 2009 at: http://skyserver.sdss.org/dr8/en/
 - 3. DR8 version of data still being loaded as we speak (photometry > 90% loaded!)
- 4. BOSS spectroscopy: first year reductions finished July 2010
 - 1. Detailed description of spectrum access and tools at https://trac.sdss3.org/wiki/BOSS/data_access
 - 2. Web form and interface up and running at https://spectra.sdss3.org:8000 (see Price-Whelan talk)
 - 3. Tinker, White etc. working on clustering samples: https://trac.sdss3.org/wiki/BOSS/clustering



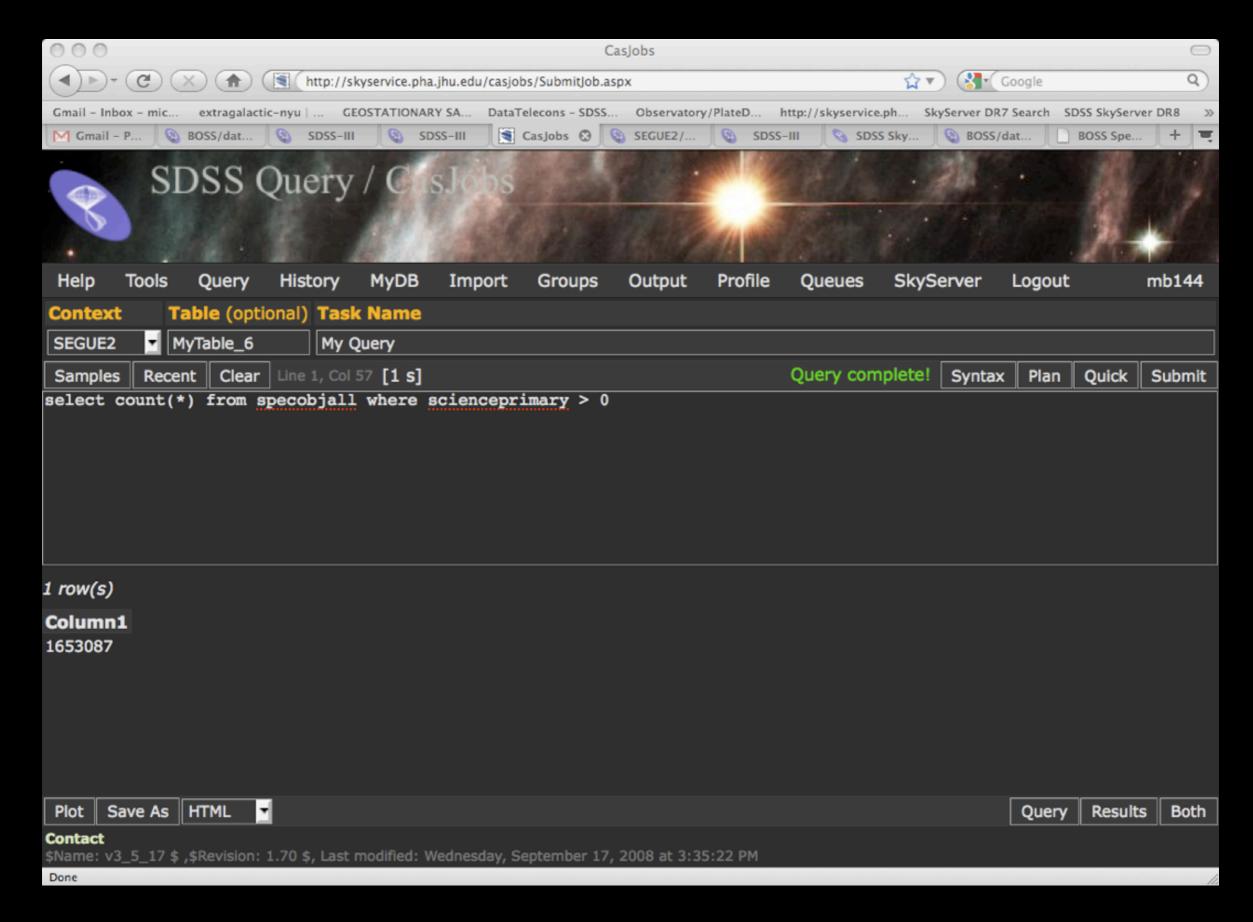
DR8 spectroscopic data access

https://sdss3.org/internal/branches/v3/dr8/spectro/spectro_catalog_access.php

SAS location	CAS table	Description
plates-dr8.fits (2.6 Mb) (in SPECTRO_REDUX; see datamodel)	plateX	List of all DR8 plates
specObj-dr8.fits (2.2 Gb) (in SPECTRO_REDUX;(see datamodel)	specObjAll	Redshifts and classifications of all SDSS spectra
ssppOut-dr8.fits (1.6 Gb) (in SSPP_REDUX; see datamodel)	sppParams	Stellar parameters from SSPP
ssppOut-dr8.lineindex.fits (2.4 Gb) (in SSPP_REDUX; see datamodel)	sppLines	Stellar line indices from SSPP
photoPlate-dr8.fits (5.5 Gb) (in SPECTRO_REDUX; see datamodel)	join on photoObjAll	Photometric flux-based matches
photoPosPlate-dr8.fits (5.5 Gb) (in SPECTRO_REDUX; see datamodel)	join on photoObjAll	Photometric position-based matches
photoMatchPlate-dr8.fits (255 Mb) (in SPECTRO_REDUX; see datamodel)		details of match results
galSpecInfo-dr8.fits (364 Mb) (in SPECTRO_REDUX; see datamodel)	galSpecInfo	Information for each galaxy processed by Garching group
galSpecIndx-dr8.fits (1.9 Gb) (in SPECTRO_REDUX; see datamodel)	galSpecIndx	Index measurements for each galaxy
galSpecLine-dr8.fits (1.7 Gb) (in SPECTRO_REDUX; see datamodel)	galSpecLine	Line measurements for each galaxy
galSpecExtra-dr8.fits (339 Mb) (in SPECTRO_REDUX; see datamodel)	galSpecExtra	Extra physical parameters for each galaxy
spPlate-PLATE-MJD.fits (in SPECTRO_REDUX/[RUN2D]/[PLATE]; see datamodel)		Individual spectra for each observation of each plate
spec-image-PLATE-MJD-FIBER.png (in SPECTRO_REDUX/images/[RUN2D] /[PLATE])	specObjAll	PNG images of each spectrum



DR8 spectroscopic data access





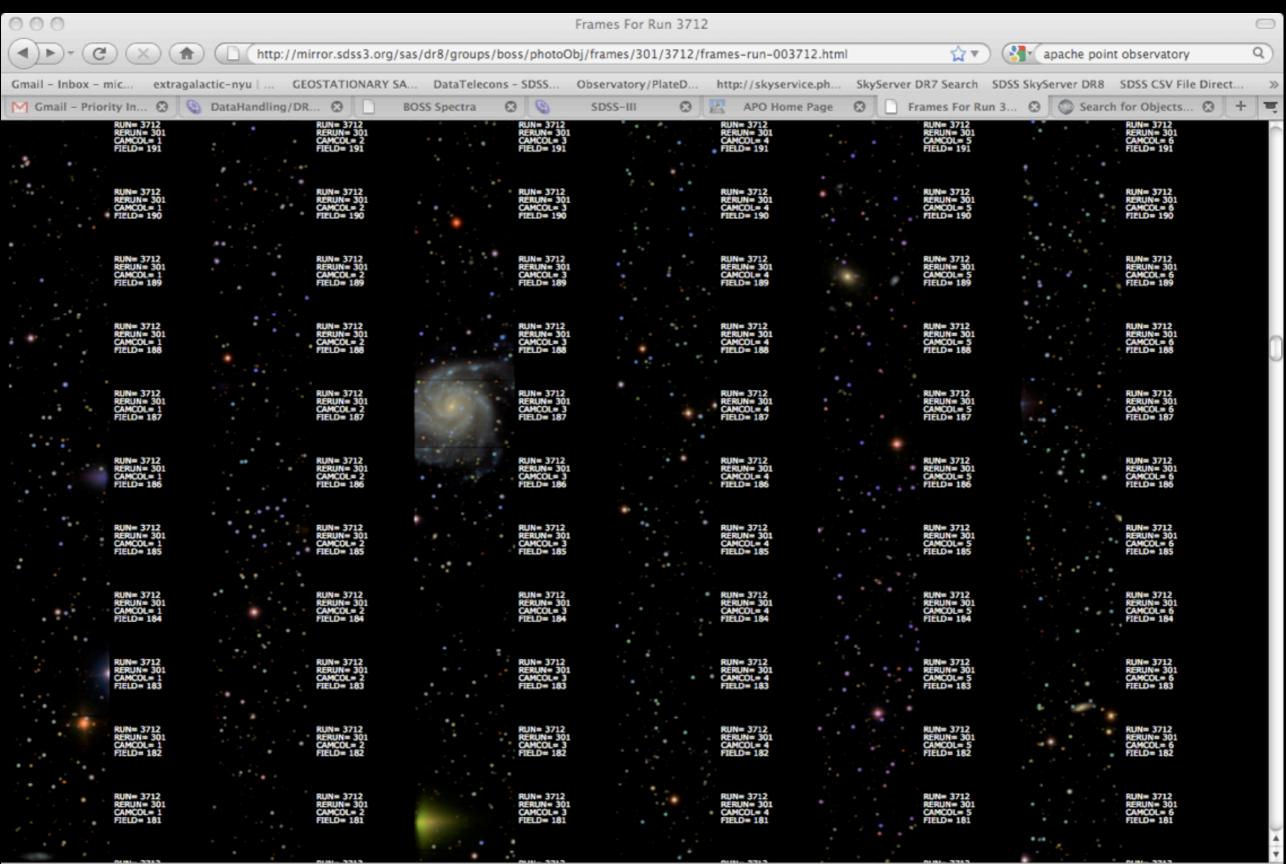
DR8 imaging data access

https://sdss3.org/internal/branches/v3/dr8/imaging/imaging_access.php

SAS location	CAS table	Description
photoObj-RUN-CAMCOL-FIELD.fits (in BOSS_PHOTOOBI; see datamodel)	photoObjAll	Photometric parameters for objects in each imaging field. (around 938,000 files total, each around 3.5Mb)
photoField-RUN-CAMCOL.fits (in <u>BOSS_PHOTOOBJ</u> ; see <u>datamodel</u>)	Field	Metadata for each imaging field in a camcol (4,590 files total, each around 1.5Mb)
photoRun-RUN.fits (in BOSS_PHOTOOBJ; see datamodel)	Run	Metadata for each imaging run (765 files total, one for each run, each around 15kb)
calibObj-RUN-CAMCOL-TYPE.fits.gz (in <u>PHOTO_SWEEP</u> ; see <u>datamodel</u>)		"Sweep" data for each imaging camcol (4590 files total, one for each run, each around 70Mb)
photoEXT-RUN-CAMCOL.fits (in BOSS_PHOTOOBJ/external; see datamodels)	TWOMASS, TWOMASSXSC, USNO, First, RC3, ROSAT, properMotions	External catalog matches and ancillary information for each camcol. Only lists SDSS objects that match. Field-by-field version in subdirectories.
frame-BAND-RUN-CAMCOL-FIELD.fits.bz2 (in BOSS_PHOTOOBJ/frames; see datamodels)		Calibrated, sky-subtracted corrected frames (SDSS-III format). JPG versions can be browsed (as in this example)
Resolve information (in <u>PHOTO_RESOLVE</u> ; see <u>datamodel</u>)		Information about the SDSS window function, including a list of all fields in window_flist.fits (336 Mb)

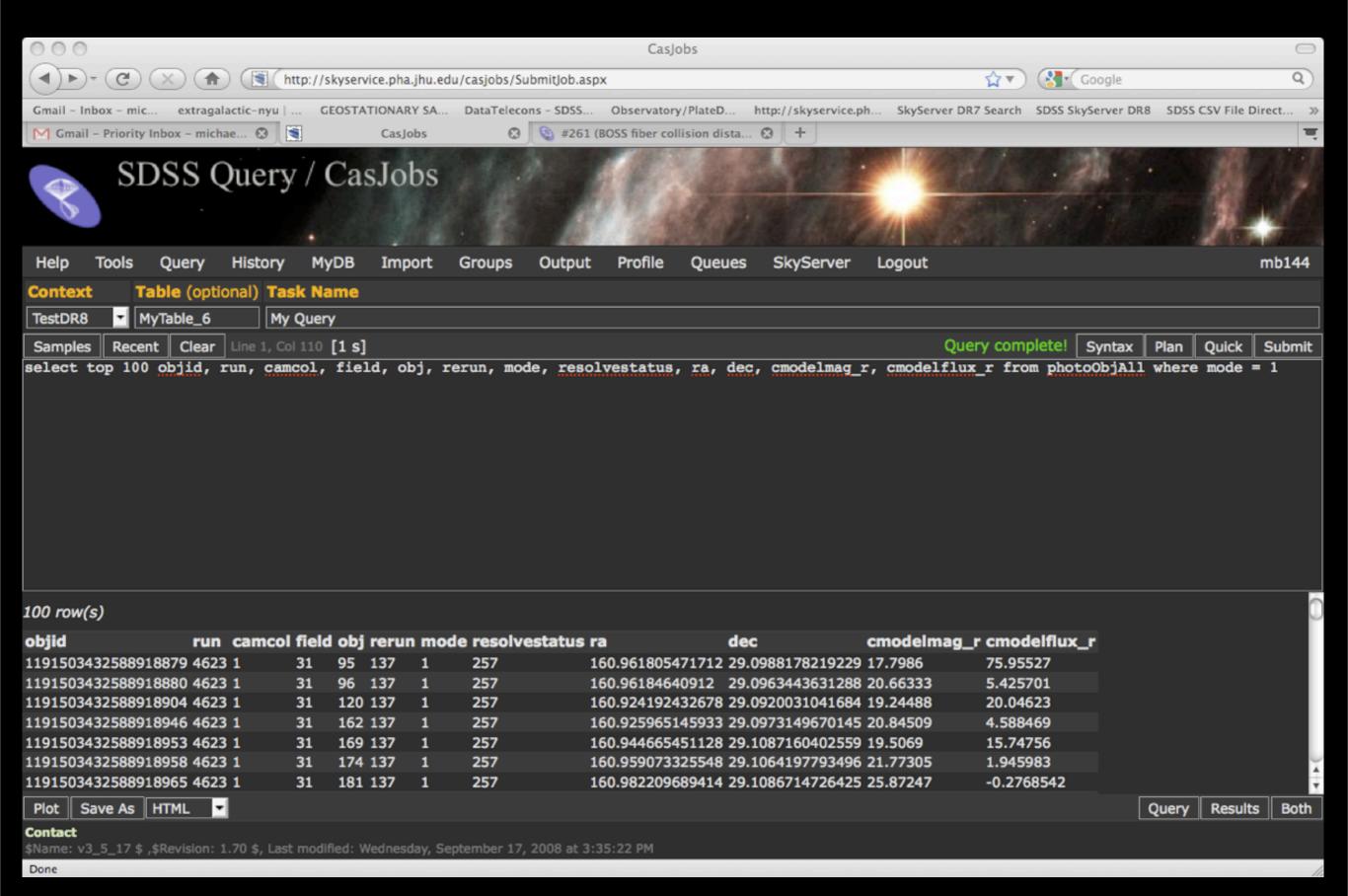


Corrected frames, new format: calibrated, sky-subtracted, correct WCS



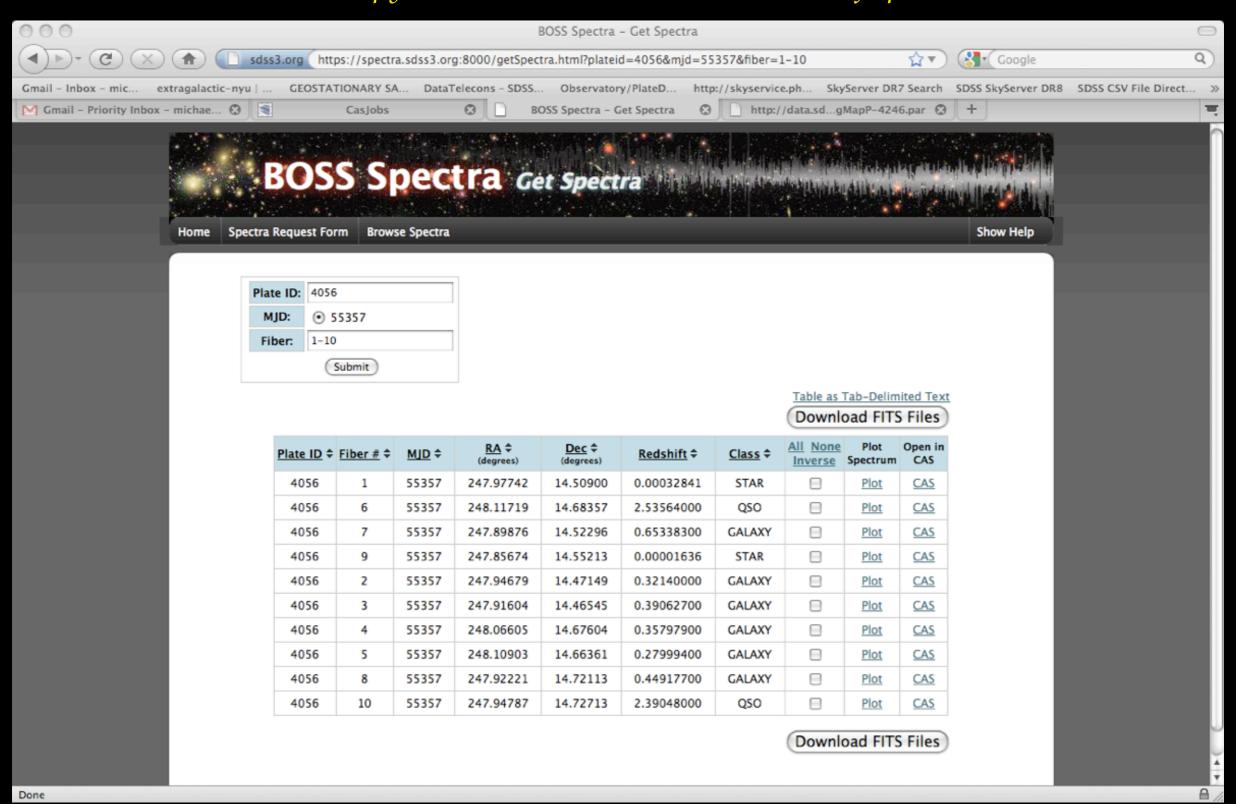


DR8 imaging data access



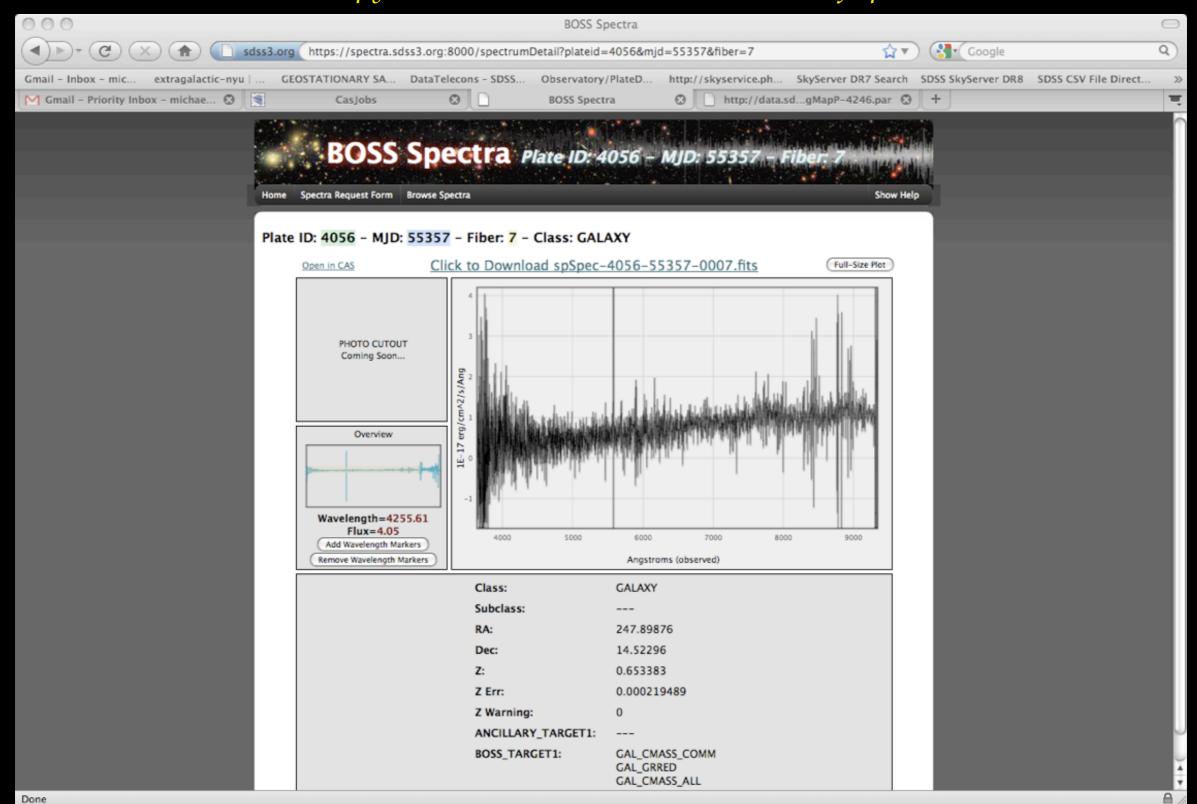


See https://trac.sdss3.org/wiki/BOSS/data_access for general data access; this is most useful for getting at the catalog or getting at a large number of the spectra. Tools there in IDL, python, sm, etc. For small numbers of spectra, consider:



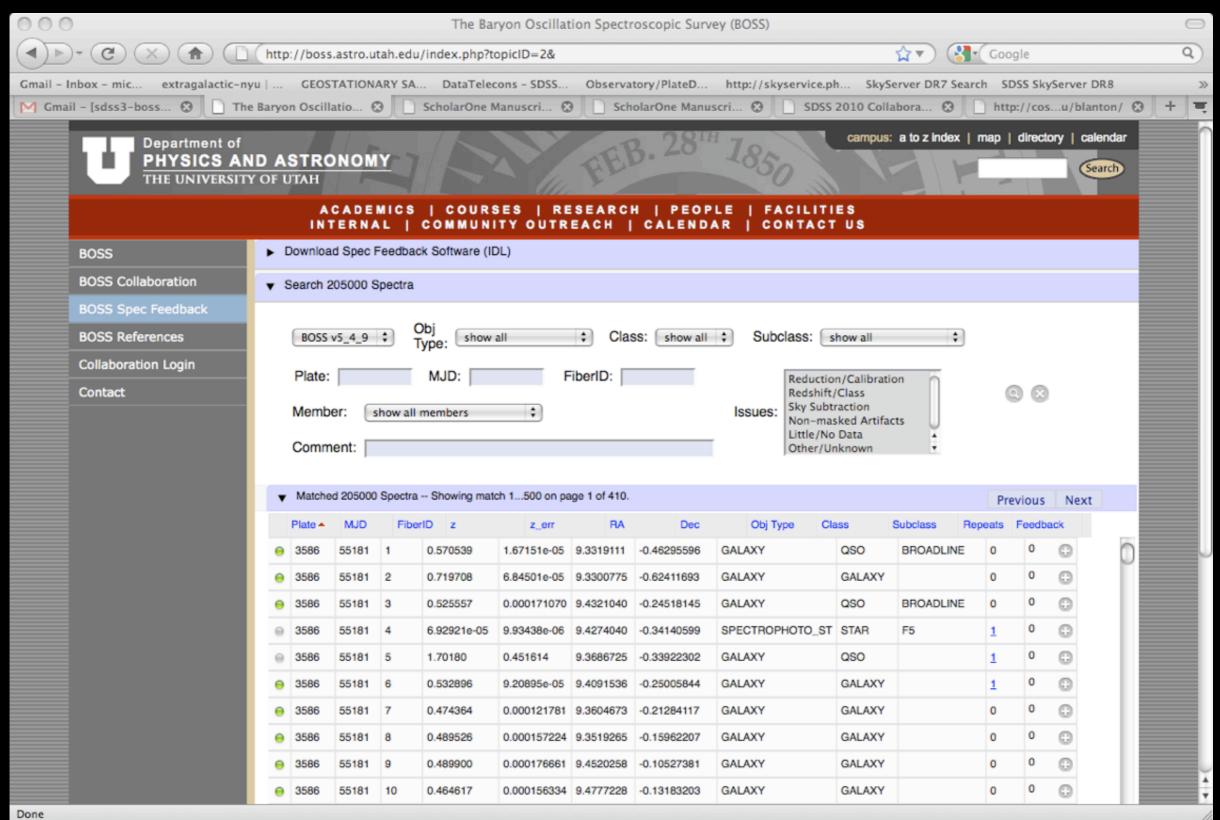


See https://trac.sdss3.org/wiki/BOSS/data_access for general data access; this is most useful for getting at the catalog or getting at a large number of the spectra. Tools there in IDL, python, sm, etc. For small numbers of spectra, consider:





But also, UU BOSS interface: http://boss.astro.utah.edu/
Works with uuplotspec.pro IDL tool





http://www.astro.washington.edu/users/anderson/oddspectra/bin/c.cgi

Odd spectrum page Plate 4041 Fiber 327 MJD=55361 000 http://www.astro.washington.edu/users/and GALAXY BROADLINE z = 0.81090GEOSTATIONARY SA... DataTelecons $X_{r}^{2} = 1.13$ M Gmail - [sdss3-... Misc. Odd 2.5m ... 🖸 The Baryon Osci... 🔾 Target = GAL_CMASS GAL_CMASS Welcome to the SDSS Serendipity odd spectra page. Below you will find a list of number by default, but a number of other options are available: you can sort, sea Quasars: High redshift, Odd BAL, BLLAC, Odd, Odd profile, Bright, Red, Type Galaxies: High redshift, Odd, EA, Supernova White Dwarfs: DZ, DQ, Magnetic, Odd Other Stars: Cataclysmic Variable, Composite, Odd, late type (M or L), Sub I Other: Unknown, Blue featureless, Superposition, Glitch, Not yet updated fiberid plate mjd min RA max RA min dec max dec Search comments for: Classification Find · Hamble from the first of the Flux [10⁻¹⁷ New (last 14 days) i Clear (original page) List of plate updates of 9148 objects selected Prev 25 Next 25 jump to interest -1.23188 51457 135 358.848 188 51457 296 358.043 -0.6591 188 51457 343 358.025 0.4372 190 51457 156 2.8915 -0.00026 1 202 51941 29.9145 -0.9045000 6000 7000 8000 9000 10000 203 32.6798 -0.30551456 141 Observed Wavelength [Ang] 0 203 51456 394 31.8718 0.4118 203 51456 533 32.9284 1.07957 18.69 0 0 0 **BLU-FEAT** 18.96 19.42 19.81 20.15



Also in progress, a more ambitious "Science Portal" being developed by the BPG group





Documentation

- 1. Probably our largest remaining issue.
- 2. Small team has made big progress: Natalia Connolly, Michael Strauss, Jordan Raddick, Bob Nichol, Ben Weaver; much documentation copiable from DR7 web site with some small changes. Results viewable in near real-time at: https://sdss3.org/internal/branches/v3/. This is in the v3 branch of the www product, which can be edited by anyone!
- 3. We need a major push on BOSS imaging documentation and SEGUE-2 target and stellar parameters documentation.
- 4. October 12-15 at NYU we will have a Documentation Festival for any/all interested parties. https://trac.sdss3.org/wiki/DataHandling/DR8MeetingOct10
- 5. Finally, as a general note, recall that survey file format documentation is always maintained at the datamodel: http://data.sdss3.org/datamodel/ (maintained in the sas data product, which also can be edited by anyone).



- 1. The structure of the data in SAS is described by the data model at https://data.sdss3.org/datamodel, maintained in \$SAS_DIR/datamodel
- 2. We track not just the file name, but also its position in the directory structure.
- 3. Of note: we always refer to root directories using environmental variable names for portability (\$PHOTO_REDUX, \$SPECTRO_REDUX, \$BOSS_PHOTOOBJ, \$CAS_LOAD etc.)



Index of Files in the SDSS-III Data Model

Please note: this file is generated automatically. Do not attempt to edit it or check it in to svn.

- asParam (\$PHOTO_REDUX/RERUN/RUN/astrom)
- asPlan (\$PHOTO_REDUX/RERUN/RUN/astrom)
- asQA (\$PHOTO_REDUX/RERUN/RUN/astrom)
- asTrans (\$PHOTO_REDUX/RERUN/RUN/astrom)
- calibMatch (\$PHOTO_REDUX/RERUN/RUN/nfcalib)
- calibObj (\$PHOTO_SWEEP/RERUN)
- calibPhotom (\$PHOTO_REDUX/RERUN/RUN/nfcalib)
- calibPhotomGlobal (\$PHOTO_CALIB/RERUN/RUN/nfcalib)
- csv_ready (\$CAS_LOAD/phCSV/SKYVERSION/RUN)
- datasweep-index (\$PHOTO_SWEEP)
- fpAtlas (\$PHOTO_REDUX/RERUN/RUN/objcs/CAMCOL)
- fpBIN (\$PHOTO_REDUX/RERUN/RUN/objcs/CAMCOL)
- fpC (\$PHOTO_REDUX/RERUN/RUN/objcs/CAMCOL)
- fpFieldStat (\$PHOTO_REDUX/RERUN/RUN/objcs/CAMCOL)
- fpM (\$PHOTO_REDUX/RERUN/RUN/objcs/CAMCOL)
- fpObjc (\$PHOTO_REDUX/RERUN/RUN/objcs/CAMCOL)
- <u>fpParam</u> (\$PHOTO_REDUX/RERUN/RUN/objcs/CAMCOL)
- fpPlan (\$PHOTO_REDUX/RERUN/RUN/objcs/CAMCOL)
- fpPlan (\$PHOTO_REDUX/RERUN/RUN/photo)
- idB (\$PHOTO_REDUX/RERUN/RUN/photo/calib)
- idFrameLog (\$PHOTO_REDUX/RERUN/RUN/logs)
- idGang (\$PHOTO_DATA/RUN/gangs)
- idR (\$PHOTO_DATA/RUN/fields/CAMCOL)
- idReport (\$PHOTO_REDUX/RERUN/RUN/logs)
- idWeather (\$PHOTO_REDUX/RERUN/RUN/logs)
- koCat (\$PHOTO_REDUX/RERUN/RUN/ssc)
- opBC (\$PHOTO_REDUX/RERUN/RUN/logs)
- opCamera (\$PHOTO_REDUX/RERUN/RUN/logs)
- opConfig (\$PHOTO_REDUX/RERUN/RUN/logs)
- opECalib (\$PHOTO_REDUX/RERUN/RUN/logs)
- <u>photoField</u> (\$BOSS_PHOTOOBJ/RERUN/RUN)
- photoMatchPlate (\$BOSS_PHOTOOBJ/plates/RERUN/PLATE4)

Done

