

SAG#8: Cosmic Origins Science Enabled by the WFIRST-AFTA Data Archive

Sally Heap & the SAG#8 Team

SAG#8 Charter

How will the WFIRST-AFTA data archive be used for Cosmic Origins science?

- Cross-section of COR science investigations
- High-level science data products
- Catalogs
- Archive interface design
- Calibration requirements
- Data accessibility & distribution
- Computing resources
- Archive operations

What are the data requirements needed to conduct COR science?

How to maximize the return via coordination with other astronomical archives?

Partial Inventory of Objects to be Observed by WFIRST

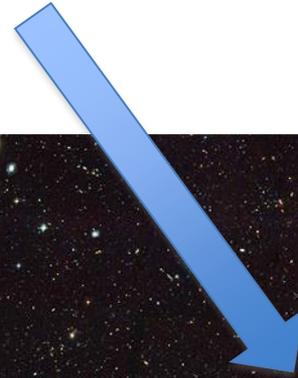
High Latitude Surveys

- 400M galaxies with measured shapes
- 30M galaxies in redshift survey
- 20M H α galaxies at $z=1-2$
- 2M [O III] galaxies at $z=2-3$
- 10^5 galaxies at $z \geq 7.5$ brighter than 26 mag
- 40K massive galaxy clusters
- 2700 SN Ia at $z=0.1-1.7$

Microlensing Survey

- 2×10^8 stars in galactic bulge ($\sim 40,000$ obs. per star)
- 3000 planets; 300 with $M \leq M_{\oplus}$
- 10^5 transiting planets
- 5000 KBO's down to 10 km with orbits

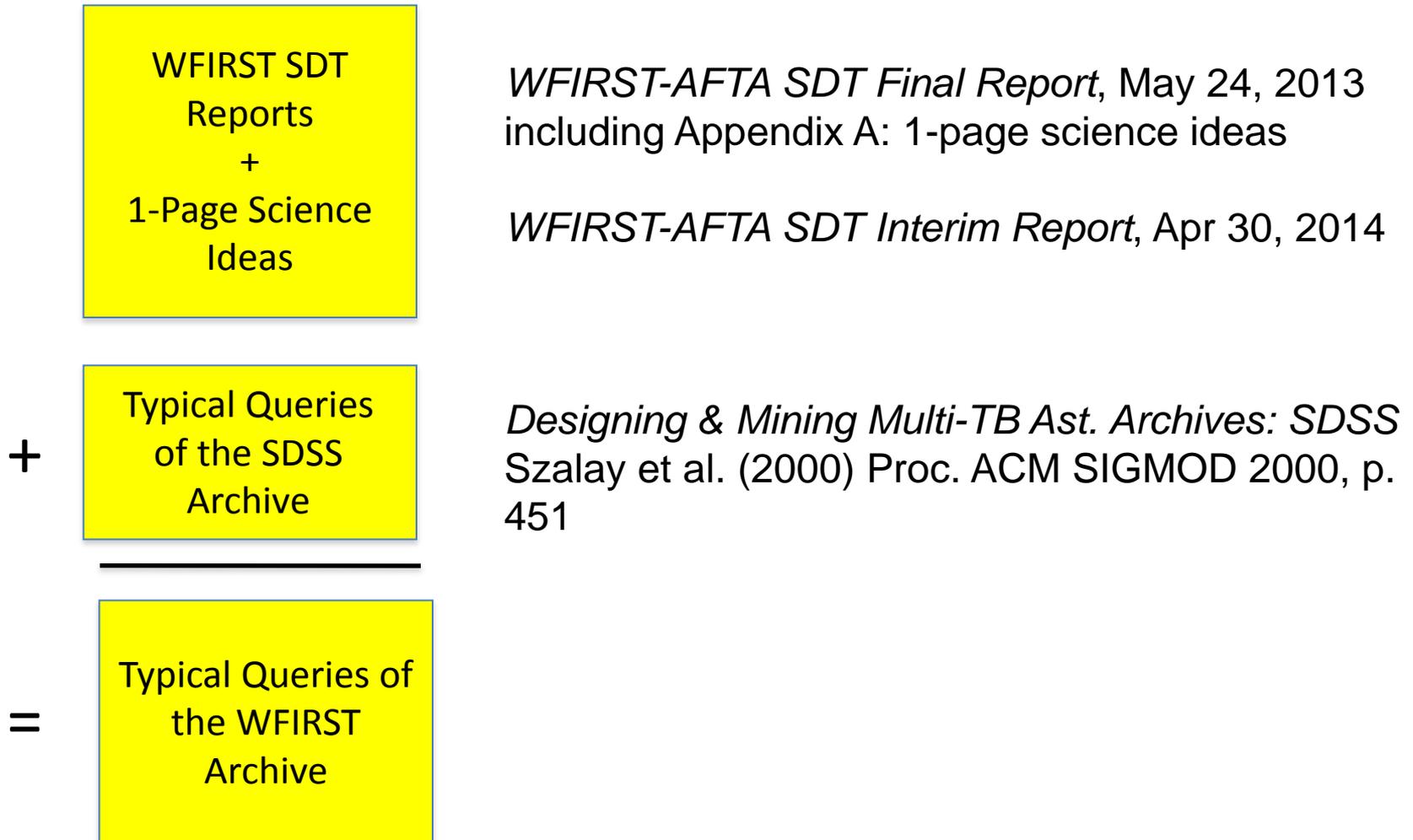
The answer is here ...somewhere



WFIRST-AFTA Deep Field reaches >1,000,000 galaxies in each image

How to find what *you* want?

Our Approach to WFIRST Archives



Cosmic Origins scientists have told us how they want to use the WFIRST archives



Community Members that Submitted 1-page Descriptions of Potential GO Science Programs in the 2013 SDT Report



WFIRST Sample Queries*

- Microlensing Field (Z087, W149) – COPAG queries
- *ML1: Find all microlensing events of stars in the galactic bulge in which the apparent position of the lens shifted by a measurable amount during the microlensing event. (Sahu, A-18). This is a search for neutron stars and stellar-mass black holes in the Galaxy.*
- *ML2: Provide a complete database of the fluxes, positions, proper motions and parallaxes of all bulge and disk stars ($\sim 10^8$ stars) in microlensing survey of the galactic bulge. (Gaudi, A-19)*
- *ML3: Provide a list of all bulge stars showing evidence of having a transiting planet(s). (IR-53,57)*
- *ML4: Find all KBO's (Gould 2014, IR-57)*
- *ML5: Find all objects whose absolute magnitudes and colors are consistent with blue stragglers / red giants / white dwarfs / <keyword>. (31Oct14 WFIRST SDT telecon)*
- All queries inspired by references; not direct quotes
- References: author of 1-page science idea; Appendix A, page #

WFIRST Sample Queries*

- High latitude imaging ($\sim R, Y, J, H, F184$)

Find all stars brighter than $J \sim 25$ whose WFIRST+WISE colors are consistent with an L or T brown dwarf (Tanner, A-12)

Find all galaxies showing double nuclei (Conselice, A-32)

Find all galaxies whose LSST + WFIRST SED's indicate a $z_{\text{phot}} > 7$

WFIRST Sample Queries

- High-latitude spectra (1.35-1.95 μm)

Find all elliptical $z > 1$ galaxies whose spectra show an anomalous emission line (Szalay, Q11)

Find all $z > 1$ galaxies observed by both WFIRST and Euclid (0.9-2.0 μm) having $\text{H}\alpha$ and [O III] emission lines (Scarlata, A-47).

- Supernovae

Find all galaxies in clusters at $z \sim 0.5-1.5$ in which > 0.5 -mag flux variations were detected

The 20 Queries around which the SDSS SkyServer was built

- Q1: Find all galaxies without unsaturated pixels within 1' of a given point of $ra=75.327$, $dec=21.023$
- Q2: Find all galaxies with blue surface brightness between 23 and 25 mag per square arcseconds, and $-10 < \text{super galactic latitude (sgb)} < 10$, and declination less than zero.
- Q3: Find all galaxies brighter than magnitude 22, where the local extinction is > 0.75 .
- Q4: Find galaxies with an isophotal surface brightness (SB) larger than 24 in the red band, with an ellipticity > 0.5 , and with the major axis of the ellipse having a declination of between 30" and 60" arc seconds.
- Q5: Find all galaxies with a deVaucouleurs profile ($r^{1/4}$ falloff of intensity on disk) and the photometric colors consistent with an elliptical galaxy. The deVaucouleurs profile
- Q6: Find galaxies that are blended with a star, output the deblended galaxy magnitudes.
- Q7: Provide a list of star-like objects that are 1% rare.
- Q8: Find all objects with unclassified spectra.
- Q9: Find quasars with a line width > 2000 km/s and $2.5 < \text{redshift} < 2.7$.
- Q10: Find galaxies with spectra that have an equivalent width in H α $> 40\text{\AA}$ (H α is the main hydrogen spectral line.)
- Q11: Find all elliptical galaxies with spectra that have an anomalous emission line.
- Q12: Create a grided count of galaxies with $u-g > 1$ and $r < 21.5$ over $60 < \text{declination} < 70$, and $200 < \text{right ascension} < 210$, on a grid of 2', and create a map of masks over the same grid.
- Q13: Create a count of galaxies for each of the HTM triangles which satisfy a certain color cut, like $0.7u - 0.5g - 0.2i < 1.25$ && $r < 21.75$, output it in a form adequate for visualization.
- Q14: Find stars with multiple measurements and have magnitude variations > 0.1 . Scan for stars that have a secondary object (observed at a different time) and compare their magnitudes.
- Q15: Provide a list of moving objects consistent with an asteroid.
- Q16: Find all objects similar to the colors of a quasar at $5.5 < \text{redshift} < 6.5$.
- Q17: Find binary stars where at least one of them has the colors of a white dwarf.
- Q18: Find all objects within 30 arcseconds of one another that have very similar colors: that is where the color ratios $u-g$, $g-r$, $r-i$ are less than 0.05m.
- Q19: Find quasars with a broad absorption line in their spectra and at least one galaxy within 10 arcseconds. Return both the quasars and the galaxies.
- Q20: For each galaxy in the BCG data set (brightest color galaxy), in $160 < \text{right ascension} < 170$, $-25 < \text{declination} < 35$ count of galaxies within 30" of it that have a photoz within 0.05 of that galaxy.

Szalay's 20 questions were implemented as SQL queries of the SDSS archive

<http://cas.sdss.org/dr4/en/help/docs/realquery.asp>

“Click on the name of the query from the list below to go directly to that sample query. The queries are roughly in order of increasing complexity. You can cut and paste queries from here into your favorite search tool”.

Basic SQL: Basic SELECT-FROM-WHERE Basic position search Using PhotoTag Search for a Range of Values Rectangular position search More than one table: JOIN...ON Photometry & spectroscopy Counting by type or category Using flags	General Astronomy: Only stars or galaxies Clean photometry Using Field MJD Objects by spectral lines Spectra by classification Moving asteroids Plates with repeat spectra Galaxies blended with stars Counts by type and program Checking SDSS footprint	Galaxies: Clean photometry - Galaxies Galaxies with blue centers Diameter limited sample LRG sample selection Galaxy counts on HTM grid Classifications from Galaxy Zoo BOSS target selection BOSS Stellar Masses BOSS Stellar Vel. Disps.
SQL Jujitsu: Data subsample Objects in close pairs Selected neighbors in run Object counts and logic Repeated high-z objects Splitting 64-bit values Using LEFT OUTER JOIN Using Nested Queries	Stars: Clean photometry - Stars CVs using colors Binary stars colors Using sppLines table Using sppParams table Proper motions	Variability Queries: Stars multiply measured Multiple Detections and Time Series
Quasars: QSOs by spectroscopy QSOs by colors FIRST matches for quasars	Miscellaneous: Photometric Redshifts Spectra in Other Programs - I Spectra in Other Programs - II Using WISE Cross-Match	APOGEE: All APOGEE Plate Visits ASPCAP Parameters and Errors APOGEE Stars No BAD Flags ASPCAP Params for Cluster Mbrs APOGEE Proper Motions APOGEE Stars Near Cluster Ctr RVs for Individual APOGEE Visits APOGEE and SEGUE Spectra SDSS photometry for APOGEE Stars

Sample Query

Query written
In SQL

Query contributed
by user

Galaxies with blue centers



```
-- Galaxies with bluer centers, by Michael Strauss. For all galaxies with r_Petro < 18,  
-- not saturated, not bright, and not edge, give me those with centers appreciably bluer  
-- than their outer parts, i.e., define the center color as: u_psf - g_psf and define  
-- the outer color as: u_model - g_model; give me all objs which have  
-- (u_model - g_model) - (u_psf - g_psf) < -0.4  
--  
-- Another flags-based query.  
-- NOTE: This query takes a long time to run without the "TOP 1000".
```

```
SELECT  
colc_u, colc_g, objID  
FROM Galaxy  
WHERE  
( Flags & (dbo.fPhotoFlags('SATURATED') +  
dbo.fPhotoFlags('BRIGHT') +  
dbo.fPhotoFlags('EDGE')) ) = 0  
and petroRad_r < 18  
and ((colc_u - colc_g) - (psfMag_u - psfMag_g)) < -0.4
```

Q19: Find quasars with a broad absorption line in their spectra and at least one galaxy within 10 arcseconds. Return both the quasars and the galaxies.

QSO broadlines near galaxy [Top](#)

```
-- Find quasars with a broad absorption line and a nearby galaxy within 10arcsec.  
-- Return both the quasars and the galaxies.
```

```
SELECT Q.BestObjID as Quasar_candidate_ID , G.ObjID as Galaxy_ID, N.distance  
FROM SpecObj as Q, -- Q is the specObj of the quasar candidate  
     Neighbors as N, -- N is the Neighbors list of Q  
     Galaxy as G, -- G is the nearby galaxy  
     SpecClass as SC, -- Spec Class  
     SpecLine as L, -- L is the broad line we are looking for  
     SpecLineNames as LN -- Line Name  
WHERE Q.SpecClass =SC.value -- Q is a QSO  
     and SC.name in ('QSO', 'HIZ_QSO') -- Spectrum says "QSO"  
     and Q.SpecObjID = L.SpecObjID -- L is a spectral line of Q.  
     and L.LineID = LN.value -- line found and  
     and LN.Name != 'UNKNOWN' -- not not identified  
     and L.ew < -10 -- but its a prominent absorption line  
     and Q.BestObjID = N.ObjID -- N is a neighbor record  
     and G.ObjID = N.NeighborObjID -- G is a neighbor of Q  
     and N.distance < 10.0/60 -- and it is within 10 arcseconds of the Q.
```

Conclusions & Recommendations

To accommodate Cosmic Origins users, the WFIRST data processing system must include Levels 1, 2, and 3

- Level 1 ~ data capture, error checking, formatting, etc.
- Level 2 ~ data reduction and flux & wavelength calibration, position, etc.
- Level 3 ~ classification of the data (Gal, Star, QSO, KBO, variables, etc) measurements (mag, colors, morphology, redshifts, sizes and shapes, environment)

NASA Astrophysics Division has traditionally supported Level 1 and 2 data processing, but not level 3

The usefulness of the WFIRST Archives to Cosmic Origins scientists depends on having access to Level-3 data products

Make it better

Get the full list of 40+ queries for WFIRST archives

Send suggestions and criticisms to:

Sally.Heap@NASA.gov

Better yet, join the COPAG and participate in the
Science Analysis Group #8

We need your help