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

Sally Heap & the SAG#8 Team

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

SAG #8 will:

- Solicit from the community the types of investigations that will be conducted

- Analyze how the archive is to be used and scope the data requirements* necessary to conduct COR science investigations

- Identify the kinds of data products that are valued and needed

- Consider what other assets or efforts may be needed to maximize the science return from the WFIRST archive (e.g., coordination of WFIRST-AFTA data with LSST, Euclid, or JWST)

* high-level science products, catalogs, archive interface design, calibration requirements, data accessibility and distribution, computing resources, and archive operations

The answer is here ...somewhere

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

How to find what *you* want?

Partial Inventory of Objects to be Observed by WFIRST

WFIRST-AFTA SDT Interim Report, April 2014

High Latitude Surveys

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

Microlensing Survey

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

Our Approach to WFIRST Archives

WFIRST SDT
Reports
+
1-Page Science
Ideas

WFIRST-AFTA SDT Final Report, May 24, 2013
Including Appendix A: 1-page science ideas
(referenced as A-<n>)

WFIRST-AFTA SDT Interim Report, Apr 30, 2014

+

Typical Queries
of the SDSS
Archive

Designing & Mining Multi-TB Ast. Archives: SDSS
Szalay et al. (2000) Proc. ACM SIGMOD 2000, p. 451

=

Typical Queries of
the WFIRST
Archive

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

04/30/2014   WFIRST-AFTA SDT Interim Report   15   SDT Interim Report

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 and 23 and 25 mag per square arcseconds, and -10<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 deVaucouleours profile ($r^{¼}$ falloff of intensity on disk) and the photometric colors consistent with an elliptical galaxy. The deVaucouleours 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<redshift<2.7.

Q10: Find galaxies with spectra that have an equivalent width in Ha >40Å (Ha 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<declination<70, and 200<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<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<right ascension<170, -25<declination<35 count of galaxies within 30"of it that have a photoz within 0.05 of that galaxy.

Application to WFIRST

Revised (and added to) the 20 questions for WFIRST in order to:

- Help astronomers judge how the WFIRST archive would work for them
- Help guide designers of the WFIRST archives in devising the overall architecture of the archive: the data repository, indexing strategies, and query tools.
- Help WFIRST project reviewers conduct realistic stress tests of competing archive designs

Main changes

- Made queries more concrete by referring to specific astronomical objects (may be placeholder for other objects)
- Organized the questions by survey (microlensing, high-latitude imaging & spectroscopy, supernovae)
- Included coordination with other surveys (LSST, WISE, etc.)

WFIRST Sample Queries*

- ## Microlensing (Z087, W149)

Find all microlensing events of stars in the galactic bulge in which the apparent position of the lens shifted by xx micro-arcsec during the event (This is a search for neutron stars and stellar-mass black holes in the galaxy). (Sahu, A-18)

- ## High latitude imaging (R?, Y, J, H, F184)

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

Find all galaxies showing red tidal tails (Conselice, A-32)

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

* All queries inspired by references; not direct quotes

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~2 galaxies observed by both WFIRST and Euclid (1.1-2.0 μm) having Hα, [O II], and [O III] emission lines (Scarlata, A-47). Check for [O III] λ4363.

- ## Supernovae

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

Next Step of SAG#8 Study

- First step: How to find what you want in the WFIRST archives – almost finished (this presentation)

- Next step: How to understand what you got (via visualizations, WFIRST database analytics, etc.)

    Example: Suppose you queried the WFIRST HLS + Euclid catalogs and found all z=1.5-2.0 star-forming galaxies with measured luminosities and nebular metallicities. We would expect to find that more luminous (massive) galaxies have higher metallicities. How tight is that correlation? Which galaxies are outliers? In what other respects do these outliers deviate from the main sample? What properties does the main sample have in common? Show 3D cluster diagrams.

The WFIRST database has all the data to make n-dimensional correlations. It makes sense to do the analysis in place.

We need your help! What high-level output do you need?
What analysis tools do you need?

Make it better

Get the full list of queries for WFIRST archives & send suggestions and criticisms to: Sally.Heap@NASA.gov

(SAG8 website to come)

Better yet, join the COPAG and participate in SAG#8
- ✓ Solicit from the community the types of investigations that will be conducted
- ❑ Analyze how the archive is to be used and scope the data requirements
- ❑ Identify the kinds of data products that are valued and needed
- ❑ Consider what other data sets may be needed to maximize the science return

Suggestion to WFIRST Study:
Get 1-page analysis ideas

Outline

- Scientific/calibration question to be answered

- Sources of data

- Expected Number of objects

- Type of data wanted, e.g. calibrated image, extracted spectra, catalog items such a z, line flux, magnitudes

- Plans for analysis
  - Where: at home institution or WFIRST computational database
  - How: method of analysis
  - Software tools needed, e.g. n-dimensional correlations
  - Algorithms you want to try out
  - Simulations you want to compare to observed data