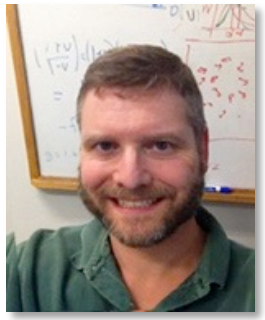
The Roman Space Telescope Science Archive: An Astrophysics Discovery Machine

Marc Postman (STScI), Alex Szalay (JHU)
and the SIT-6 Team

November 16, 2021

Our Science Investigation Team:


Jay Anderson
(STScI)


Tamas Budavari
(JHU)


Andy Connolly
(U. Wash.)


Bridget Falck
(JHU)

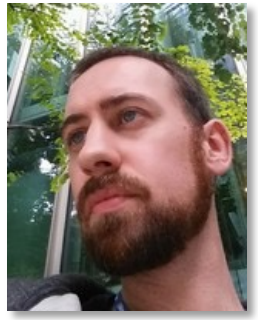

Mike Fall
(STScI)


Sarah Heap
(GSFC)


Gerard Lemson
(JHU)


Thomas McGlynn
(GSFC)


Brice Menard
(JHU)


Arik Mitschang
(JHU)


Josh Peek
(STScI)

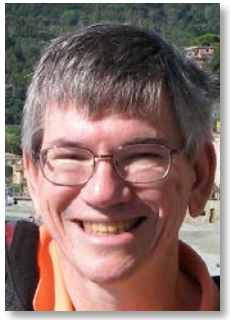

Marc Postman
(STScI)


Swara
Ravindranath
(STScI)


Hayden
Smotherman
(U. Wash)


Greg Snyder
(STScI)


Ani Thakar
(JHU)


Rick White
(STScI)

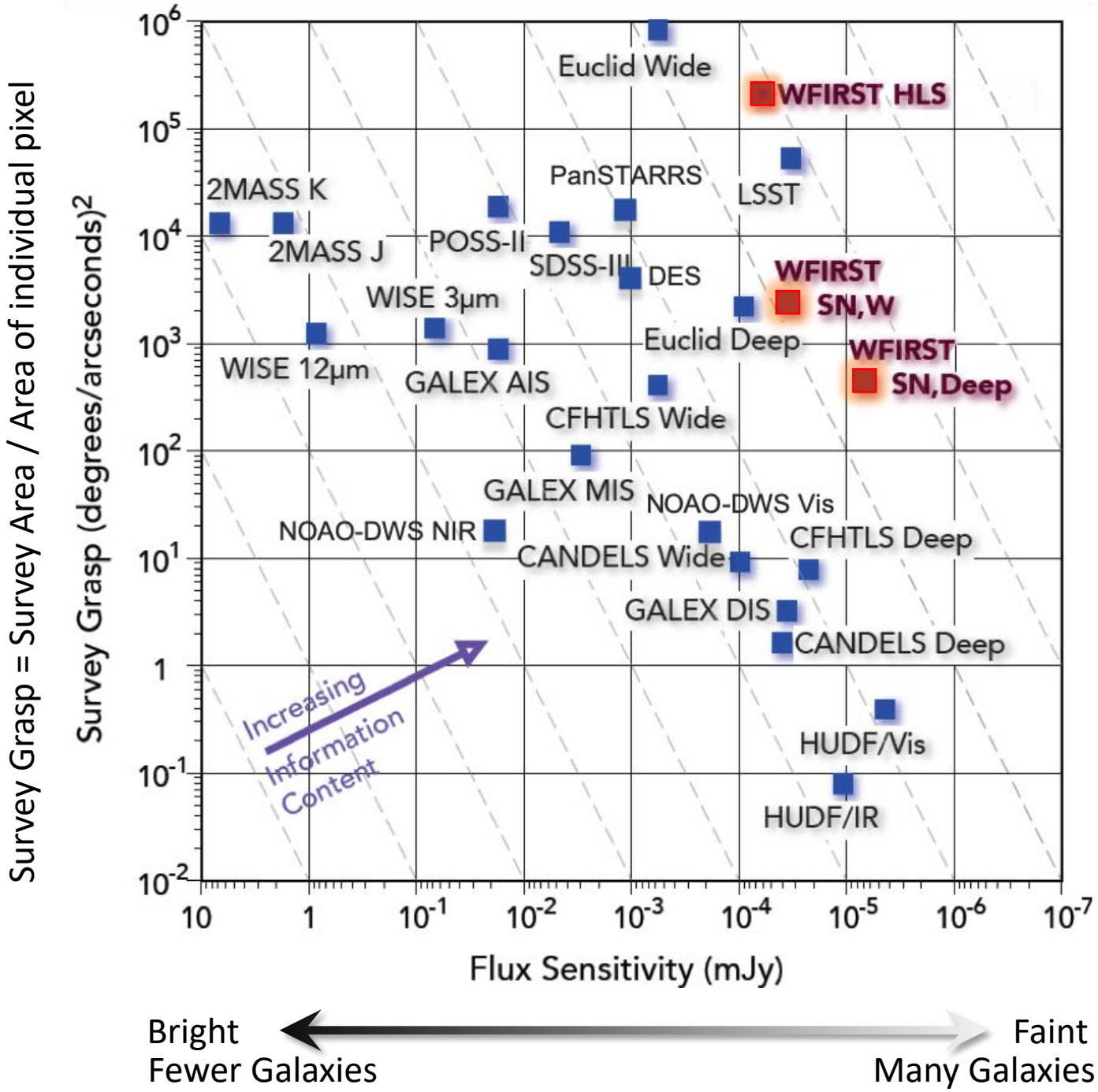

Alex Szalay, P.I.
(JHU)

Our view of the Roman Space Telescope:

- A combination of a large survey telescope with a Great Observatory.
- Key Challenge for our SIT: how do you design the RST database, advanced machine learning algorithms, and the science archive interfaces to enable a broad community to extract a wide array of science from the Roman datasets?

The surveys performed with Roman will be amongst the most information-rich datasets ever created.

The Roman Space Telescope mission will produce >15 Petabytes of data over its life cycle.



Goals for Designing Roman Archive Science Tools

- Study and evaluate the best archival practices from across all of astronomy.
- Design and build simple end-to-end simulations. Incorporate these simulations into a scalable, queryable database.
- Establish a common platform for data query interfaces and data exploration.
- Design and prototype a unified object catalog using a simulated catalog, with cross-matches to external surveys.
- Investigate and develop capabilities needed for a GO+GI hybrid archive.
- Identify forward-looking technologies not in production today anywhere (e.g., scripting, fast parallel analysis tools).
- Develop and build prototypes of highly scalable parallel tools (e.g. cross-correlations inside the archive database).
- Implement novel object classification codes based on machine learning techniques.

Design Philosophy

- The Roman Science Archive must Integrate data from a wide variety of sources, as it is likely that user queries will span not only the RST database but other databases, particularly LSST (but also PanSTARRS, Gaia, WISE, VISTA, DES, HSC, Euclid, eRosita, etc.).
- Apply the lessons learned from archives across astronomy, from GO-based (like MAST) to GI-based (like SDSS), to find best practices and identify potential solutions to the particular issues relevant to RST.
- Make the archive tools and data easily accessible to non-specialist astronomers.

This Session:

Speaker	Title	Duration
Hayden Smotherman	Maximizing Science in Big Data Astronomy	20 min
Gerard Lemson	Intro to SciServer	10 min
Bridget Falck	Overview of Indra, a peta-scale suite of cosmological simulations	10 min
Arik Mitschang	Use of the Spark distributed computing engine to enhance the big data capabilities of the SciServer	10 min
Swara Ravindranath	Roman Space Telescope Grism simulation code on SciServer	10 min
Ani Thakar	Summary / Q&A	10 min