STScI | SPACE TELESCOPE SCIENCE INSTITUTE

EXPANDING THE FRONTIERS OF SPACE ASTRONOMY

STScI and HTCondor
Reporting on Today and Exploring Tomorrow

Mary Romelfanger and Mike Swam

21 May 2019

OUR VISION

Expanding the frontiers of space astronomy

STScI | SPACE TELESCOPE SCIENCE INSTITUTE

OUR MISSION

We help humanity explore the universe with advanced space telescopes and archives

OUR STRATEGIC GOALS

Excel in the science operations of NASA's current and future astrophysics flagship missions

Advance state-of-the-art astronomical data, archives, and tools for scientific discovery

Make the world's astronomical information accessible to all

STScI | SPACE TELESCOPE SCIENCE INSTITUTE

Mikulski Archive for Space Telescopes (MAST)

- Active Missions
  - Hubble Space Telescope
  - James Webb Space Telescope
  - Transiting Exoplanet Survey Satellite
  - K2
  - Neil Gehels Swift Observatory
  - X-ray Multi-Mirror

- Ground Based
  - Panoramic Survey Telescope & Rapid Response System (PANSTARRS)
  - Faint Images of the Radio Sky at Twenty-cm (VLA-FIRST)

- Legacy Missions
  - International Ultraviolet Explorer (IUE)
  - Kepler
  - Galaxy Evolution Explorer (GALEX)
  - Far Ultraviolet Spectroscopic Explorer (FUSE)
  - Copernicus
  - Extrasolar Planet Observations and Characterization (EPOCH)
  - Extreme Ultraviolet Explorer (EUVE)
  - Hopkins Ultraviolet Telescope (HUT)
  - Ultraviolet Imaging Telescope (UIT)
  - Wisconsin Ultraviolet Photo Polarimeter Experiment (WUPPE)
  - Halfwave Polarimeter (HPOL)
  - Tübingen Echelle Spectrograph (TUES)
  - Interstellar Medium Absorption Profile Spectrograph(IMAPS)
  - Berkeley Extreme and Far-UV Spectrometer (BEFS)

Previous STScI Presentations at HTCondor Week

HTCondor Week 2016

HTCondor Use In Operational Data Processing
For The Hubble Space Telescope (HST)
And The James Webb Space Telescope (JWST)

Mike Swam

HTCondor Week 2018

Using HTCondor to Calibrate and Archive HST and JWST Data

Matthew Burger

Hubble Space Telescope

Hubble Space Telescope

Hubble Space Telescope (HST)

- Launched 24 April 1990
  - Just celebrated the 29th Anniversary
- 2.4 Meter Primary mirror
- 43.5 feet (13.2 meters) long
  - size of a large school bus

- Present Instruments
  - WFC3 - Wide Field Camera 3
  - COS - Cosmic Origins Spectrograph
  - ACS - Advanced Camera for Surveys (repaired during SM4)
  - STIS - Space Telescope Imaging Spectrograph (repaired during SM4)
  - NICMOS - Near Infrared Camera and Multi-object Spectrometer. (not currently operational)
  - FGS - Fine Guidance Sensors

Hubble Space Telescope

- 5 servicing missions to replace instruments and make repairs.
- Last Serviced 11 May 2009 -- 10 years ago!!
  - No Consumables - built in redundant parts and solar power
- Originally had a 1.2 Gigabit reel-to-reel recorder
- Replaced with a 12 Gigabit(1.5 Gigabyte) solid state drive 24 Dec 1999
- Today HST Science Data Processing (SDP) generates

  about 30 Gigabytes of data per day
  - 11 Terabytes per year
- Today the archive contains
  - 166 Terabytes of HST Science Data
  - More than 1.3 Million unique observations
  - Over 15,000 scientific papers published
- New ways to process data being added
  - Hubble Legacy Archive being moved to SDP
  - more accurate astrometry calculations

Horsehead Nebula
Hubble Heritage
2013

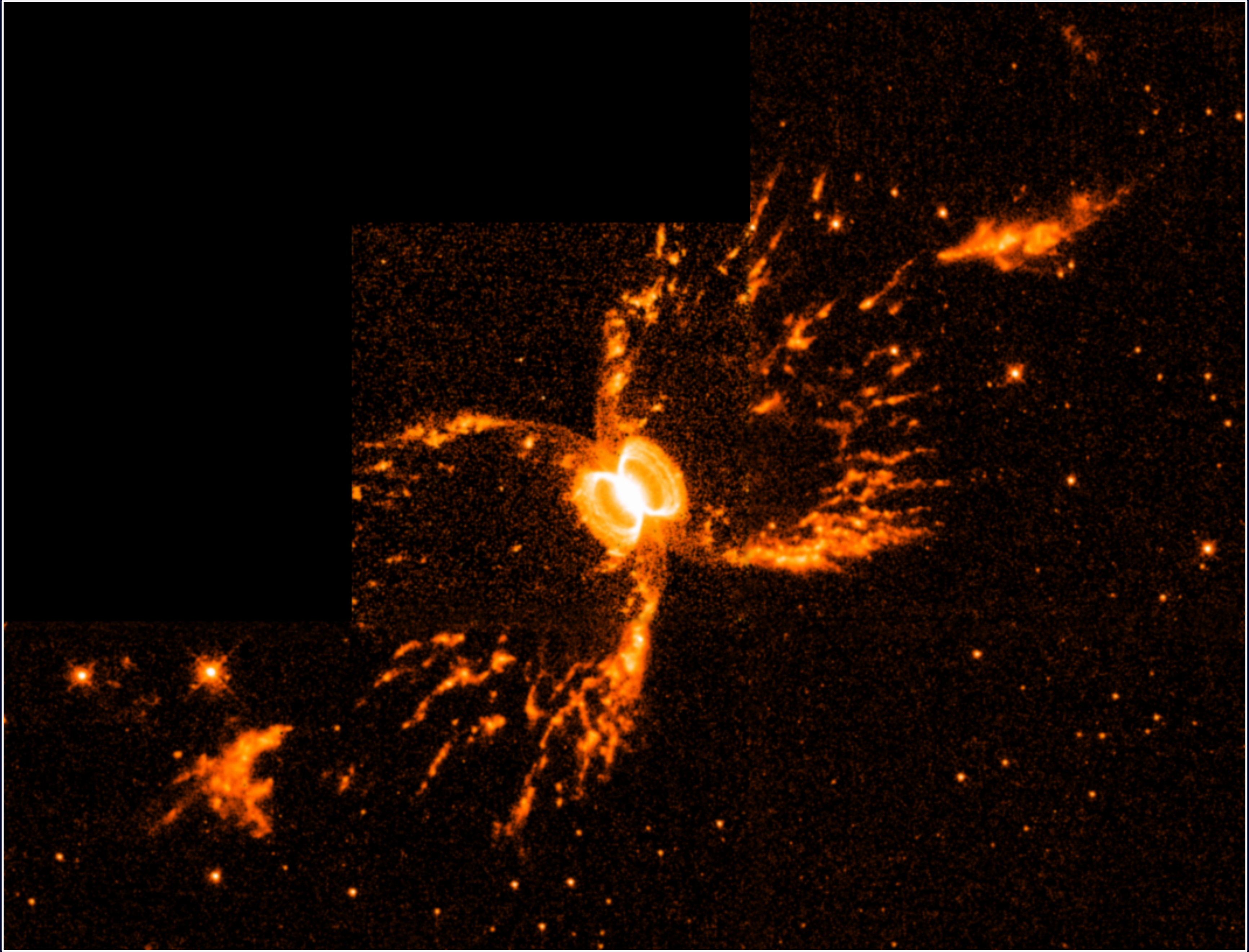
Hubble Space Telescope

- We have many very small very quick jobs
- We process HST data in a 240 core HTCondor pool with a shared nfs mounted partition.
- This is an operational environment with a single user. This single user environment allows us to use job priorities to control the order of the processing and reprocessing.
- The population of the database table used to track dataset progress through workflows is not completely reliable.
  - We see sporadic HTCondor job hook failures, likely due to database connectivity issues, so Operators can have difficulty in discerning real failures from errant reports in the database (most--often the workflows completed OK, but the DB says otherwise)

The "Rose"
Hubble 21st Anniversary 2011

Southern Crab Nebula

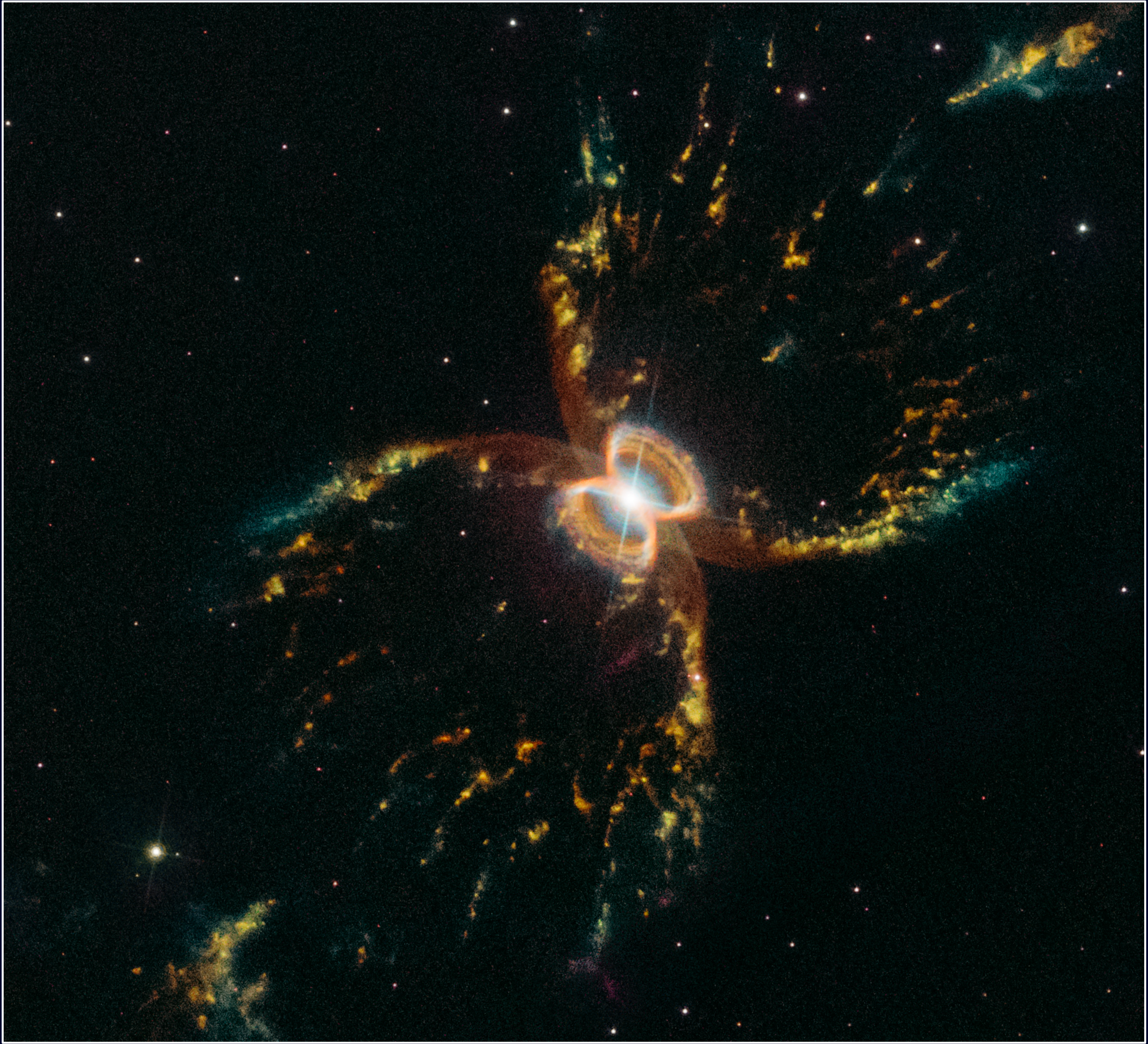
This image was taken in 1999 by Wide Field and Planetary Camera 2(WFPC2). This is a pair of aging stars buried in the glow of the tiny, central nebula. One of them is a red giant, a star that is exhausting its nuclear fuel and is currently releasing its outer layers in a powerful stellar wind. Its companion is a hot, white dwarf, a burned-out star.



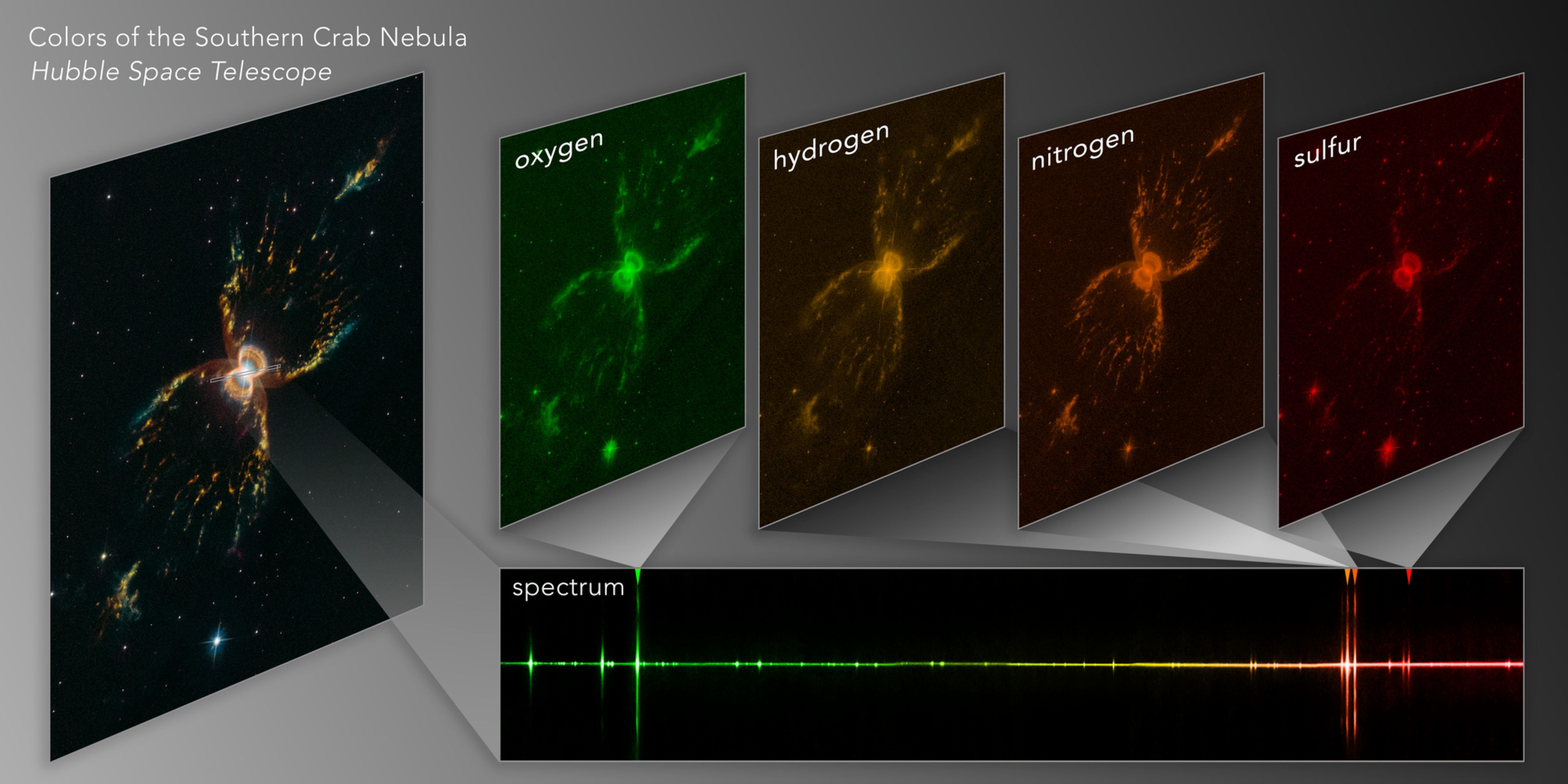
Credits: Romano Corradi, Instituto de Astrofisica de Canarias, Tenerife, Spain; Mario Livio, Space Telescope Science Institute, Baltimore, Md.; Ulisse Munari, Osservatorio Astronomico di Padova-Asiago, Italy; Hugo Schwarz, Nordic Optical Telescope, Canarias, Spain; and NASA

Southern Crab Nebula

This latest image of the Southern Crab Nebula was created from several exposures that were taken in March 2019 by Wide Field Camera 3. This image is a composite of exposures taken to detect various wavelengths of light that correspond to the glowing gases in the nebula.



Credits
NASA, ESA, and STScI

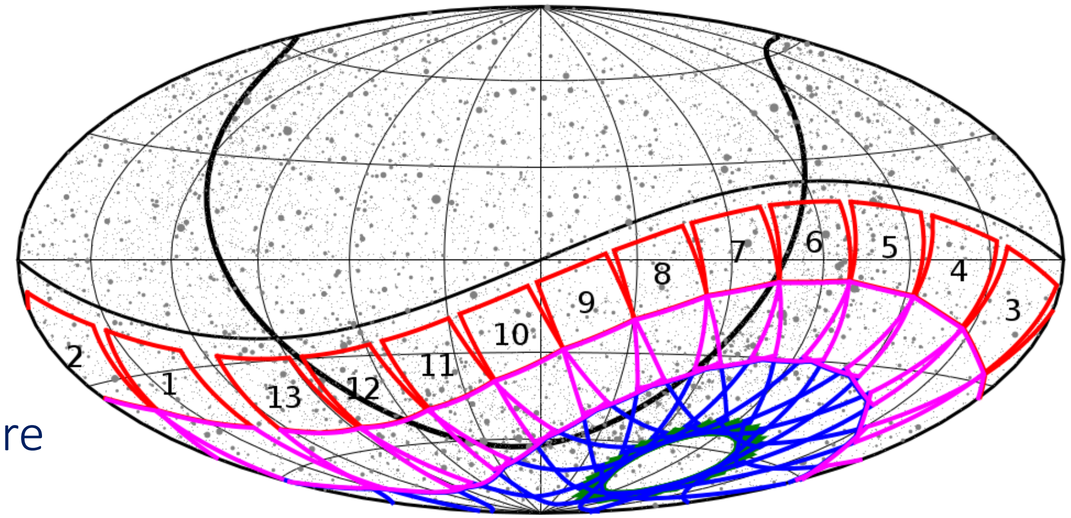


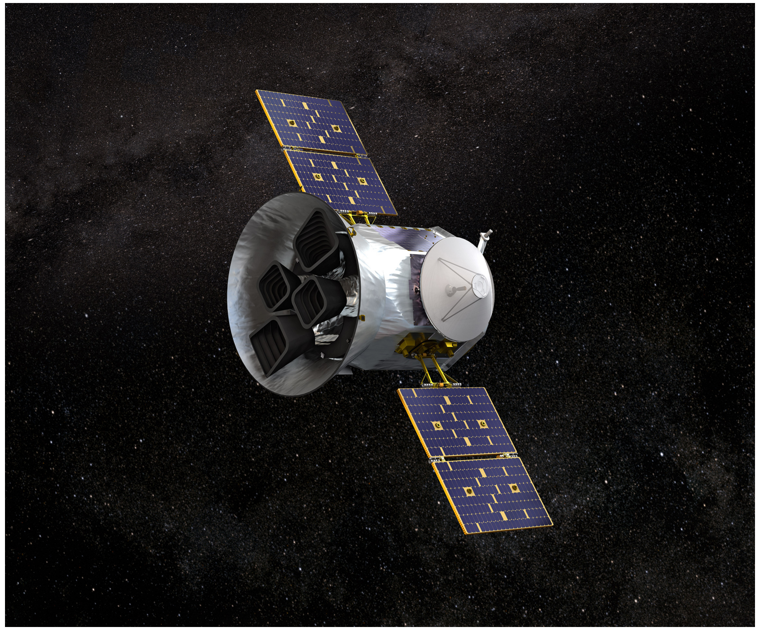
[How HST photos are created]() is explained here.

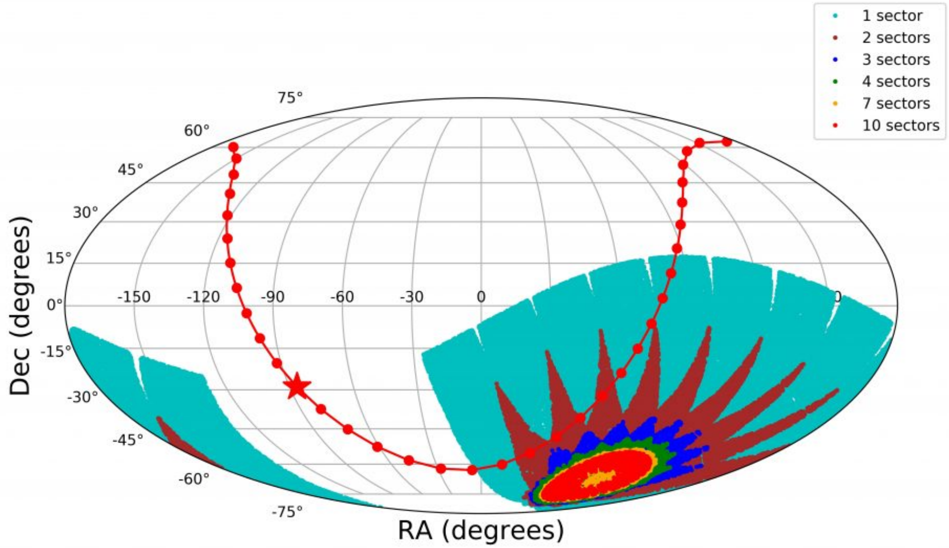
Transiting Exoplanet Survey Satellite

Transiting Exoplanet Survey Satellite (TESS)

- Launched 18 April 2018
- Primary mission is a two year survey
- 4 cameras observe each sector for 27 days
- Maps show 1st year planned sectors and the sectors that have been completed
- Year 2 surveys the northern ecliptic hemisphere







Transiting Exoplanet Survey Satellite (TESS)

MIT led NASA Mission

- Data processed and calibrated at NASA's Ames Research Center and MIT
- Data is then delivered to STScI for Archive and Distribution in batches by sector

Per sector: 1.78 TB of Full Frame Images that cover 30 minute exposures of each 96 x 13 degree sector in the sky. They also produce an additional 1TB of individual target data where known good targets are monitored. This results in over 81,000 delivered files for each sector

Total archived (9 Sectors): 26TB in about 750,000 files

2 year (26 sectors) total expected: over 60TB in over 2 Million files.

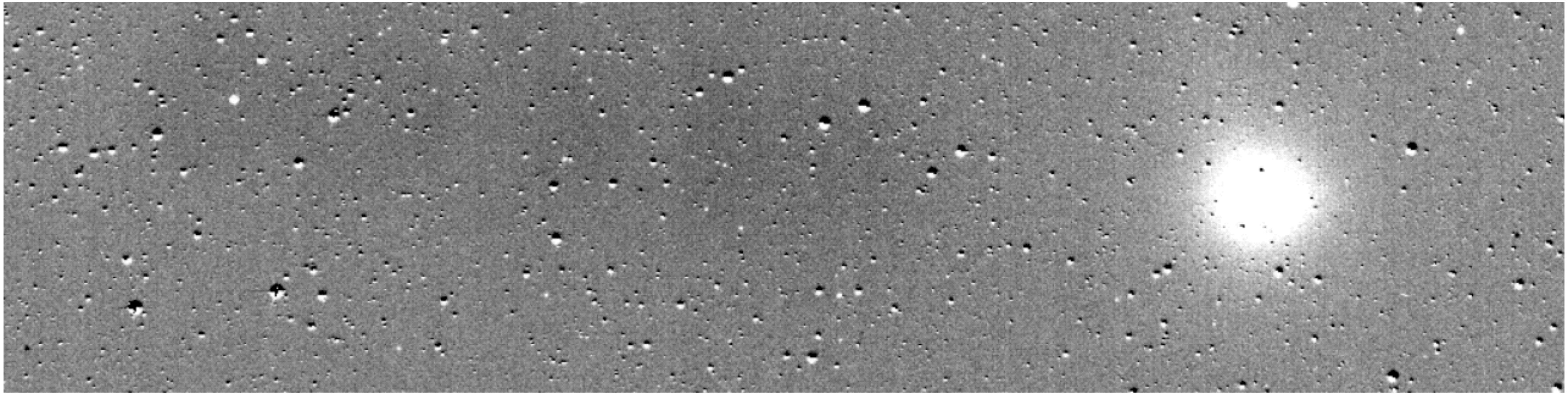
For the expected years 3 and 4, these 26 sectors will repeat but at a higher resolution

- This will more than double the data volume to about 4TB of data per sector

Fuel is available for many many more years of science.

Transiting Exoplanet Survey Satellite (TESS)

- Once the data delivery is validated, HTCondor is used to parallelize metadata xml generation and Common Archive Observation Model (CAOM) postings for each of the 81,000+ delivered files
- Average time to archive a sector: 24 hours – includes all data validation, archiving, and CAOM delivery steps.
- This runs on a very small system that meets the needs of this project without interfering with the other projects.



Series of TESS exposures catches a comet

TESS reports two more planets May 2019 article

James Webb Space Telescope

James Webb Space Telescope (JWST)

Launch date 30 March 2021

Ariane 5 rocket from Kourou, French Guiana

Shipping from California on a barge that will travel through the panama canal

5 year primary mission

- possible 5 year extension
- There are consumables on board.
- Servicing is not possible

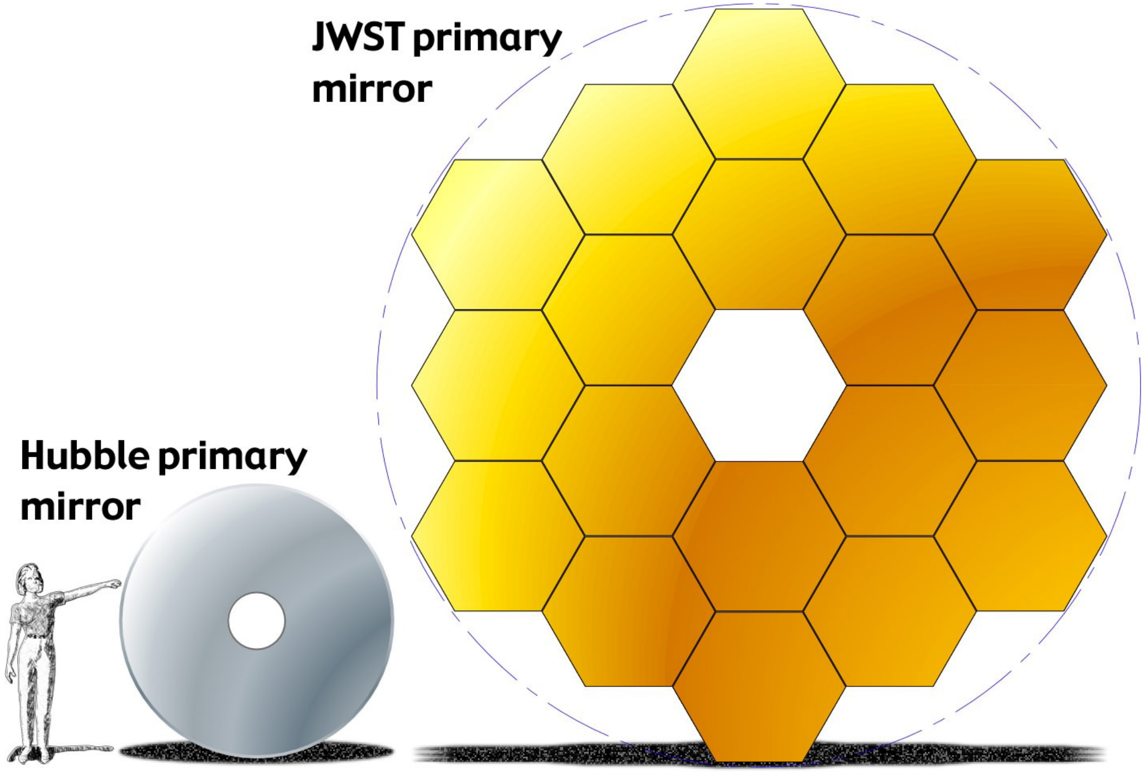
Four Instruments

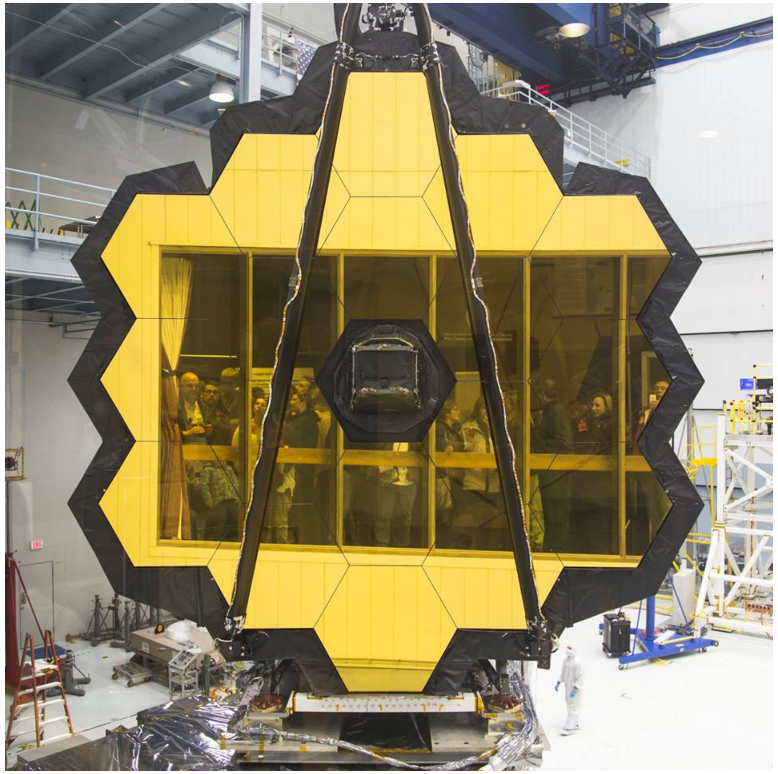
- Near-Infrared Camera, NIRCam
- Near-Infrared Spectrograph, NIRSpec
- Mid-Infrared Instrument, MIRI
- Fine Guidance Sensor/ Near InfraRed Imager and Slitless Spectrograph , FGS/NIRISS

Full Size Model of JWST at South by Southwest 2013

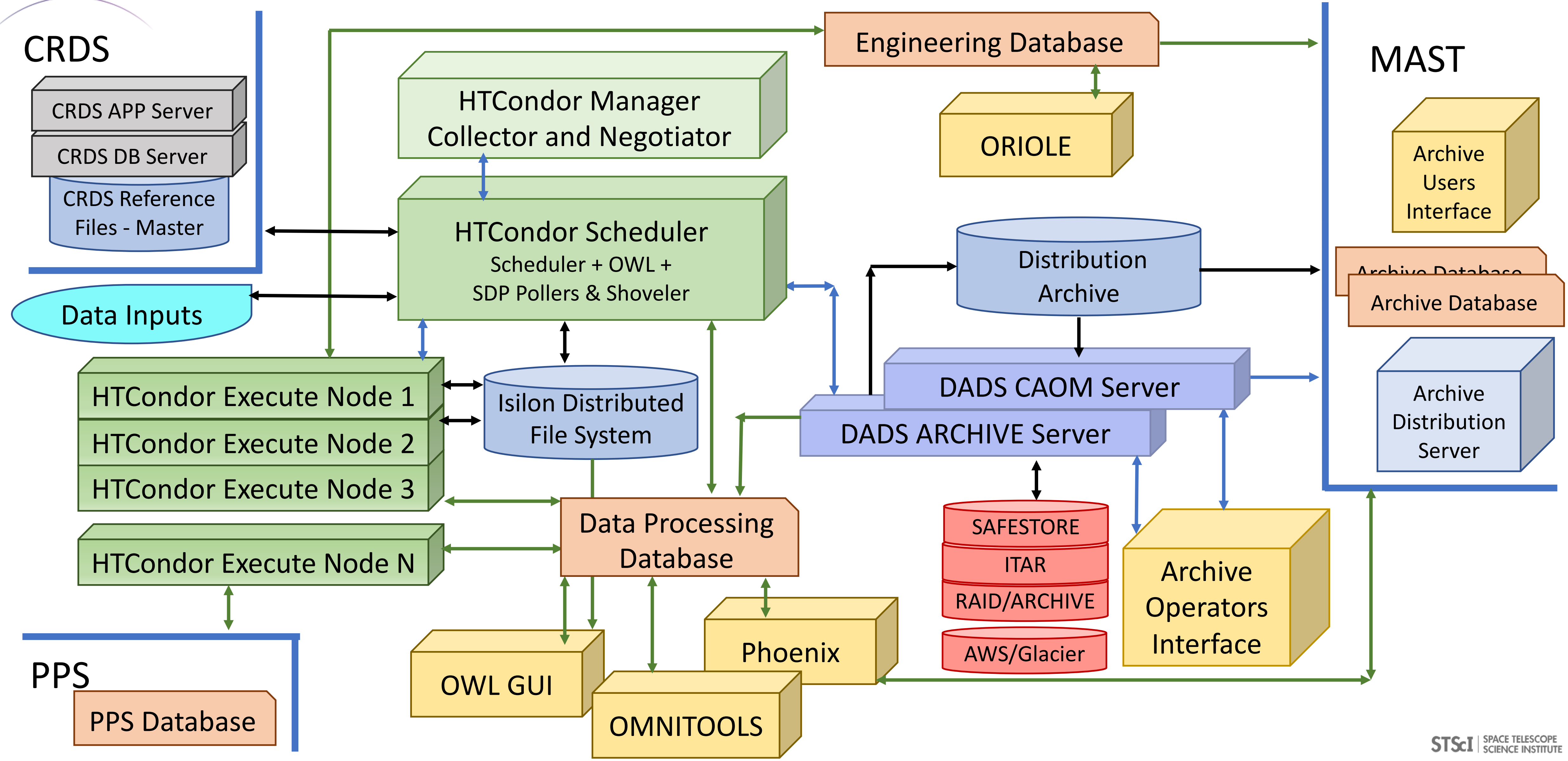
James Webb Space Telescope

- Sun Shield is the size of a tennis court.
- Primary mirror 6.5 meters (21 ft 4 in)
  - Each of the 18 hexagons is 1.32 meters (4.3 ft) and can be individually focused.
  - Each mirror is aligned to 1/10,000th the thickness of a human hair.
- 58.8 Gbytes onboard data storage
- At least 28.6 GB of recorded science data downlinked during each contact.
  - Four hour contact made twice per day

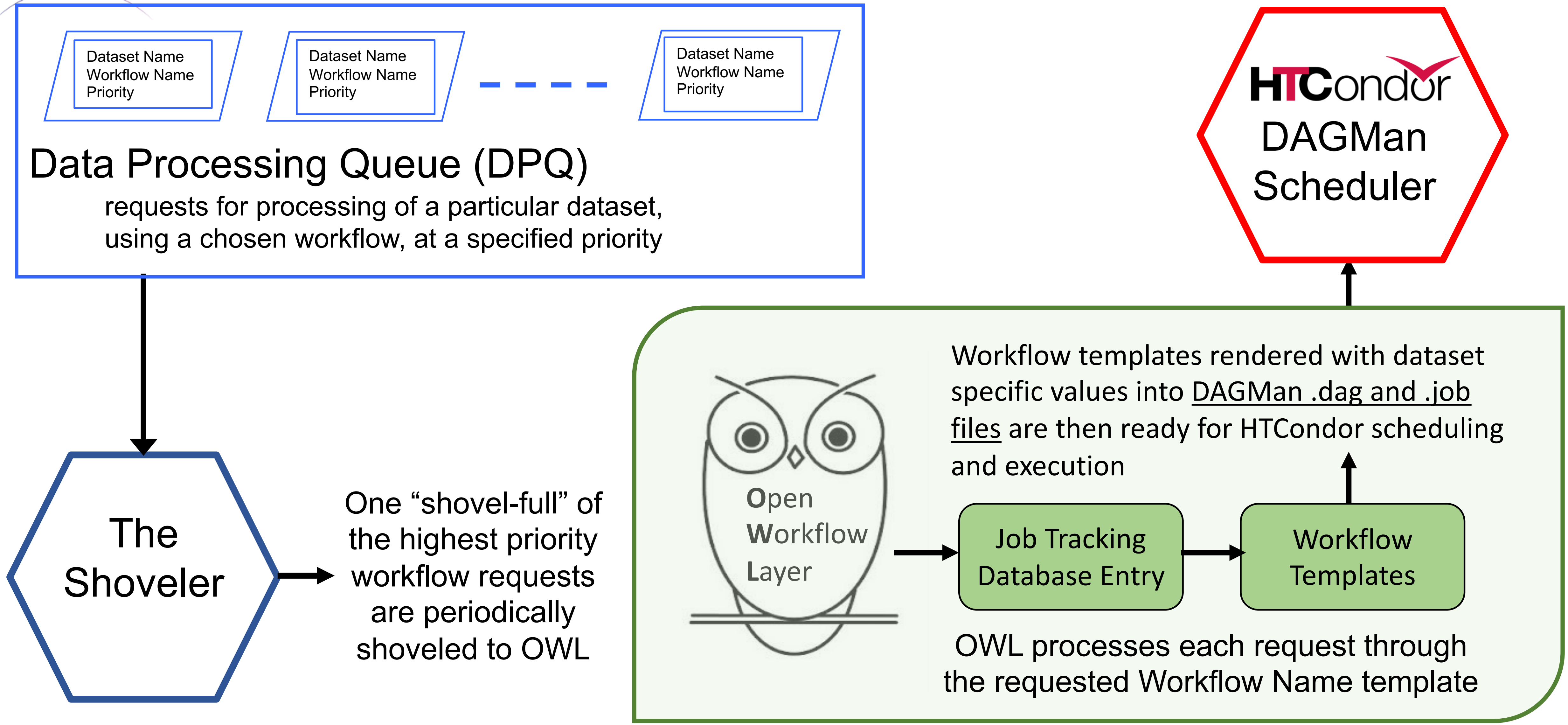




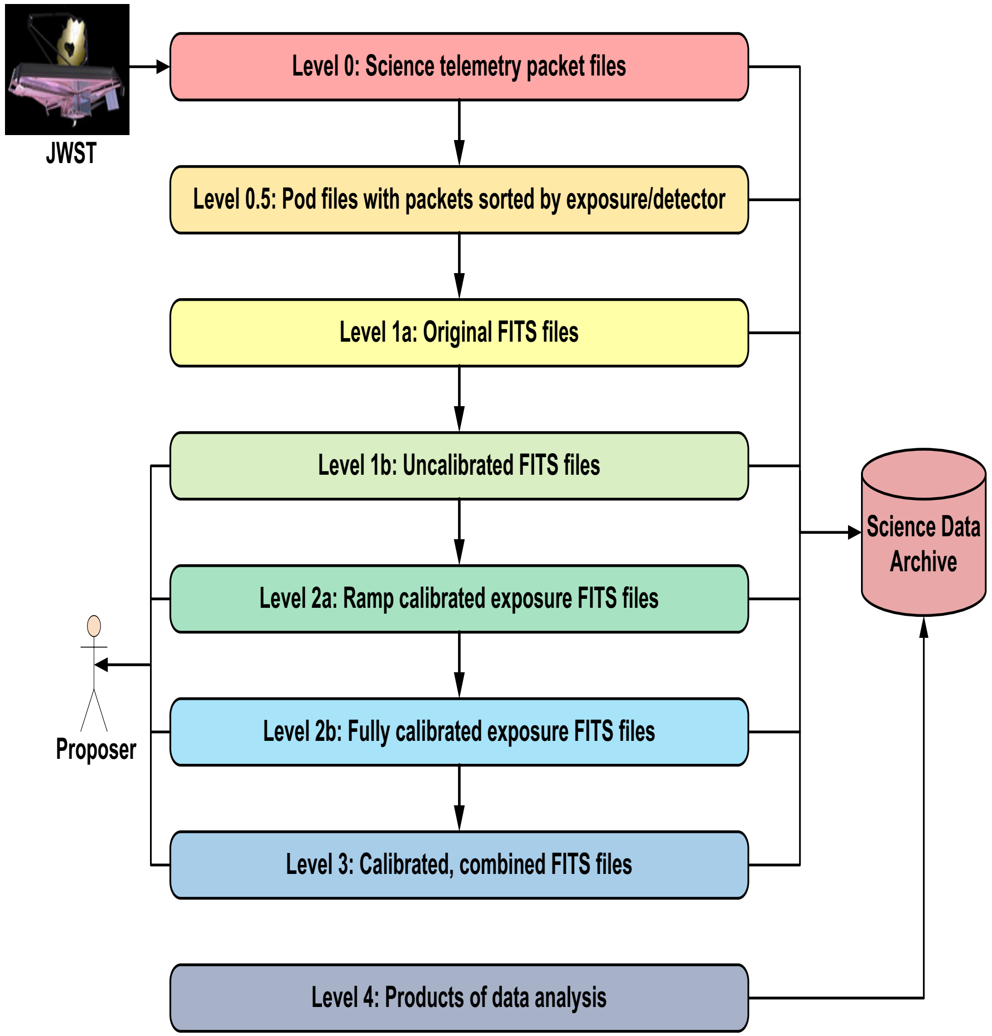
JWST Data Processing Workflow



From DPQueue to HTCondor



Data Processing Levels translate to Data Processing Job Steps



Science Data Processing Job Steps

CREATE_POD		INGEST_05
LEVEL_1	PREVIEW_1A	INGEST_1A
	PREVIEW_1B	INGEST_1B
LEVEL_2A	PREVIEW_2A	INGEST_2A
LEVEL_2B	PREVIEW_2B	INGEST_2B
LEVEL_3	PREVIEW_3	INGEST_3

James Webb Space Telescope

James Webb Space Telescope (JWST)

- Predicted to generate about:
  - 114 TB of science data per year
  - 650 TB total during the primary mission

- Introduces one time critical processing path:
  - 90 minute processing time for mirror focusing data
  - can be delivered mixed in with regular science data
  - Once identified it must move to the top of the queue

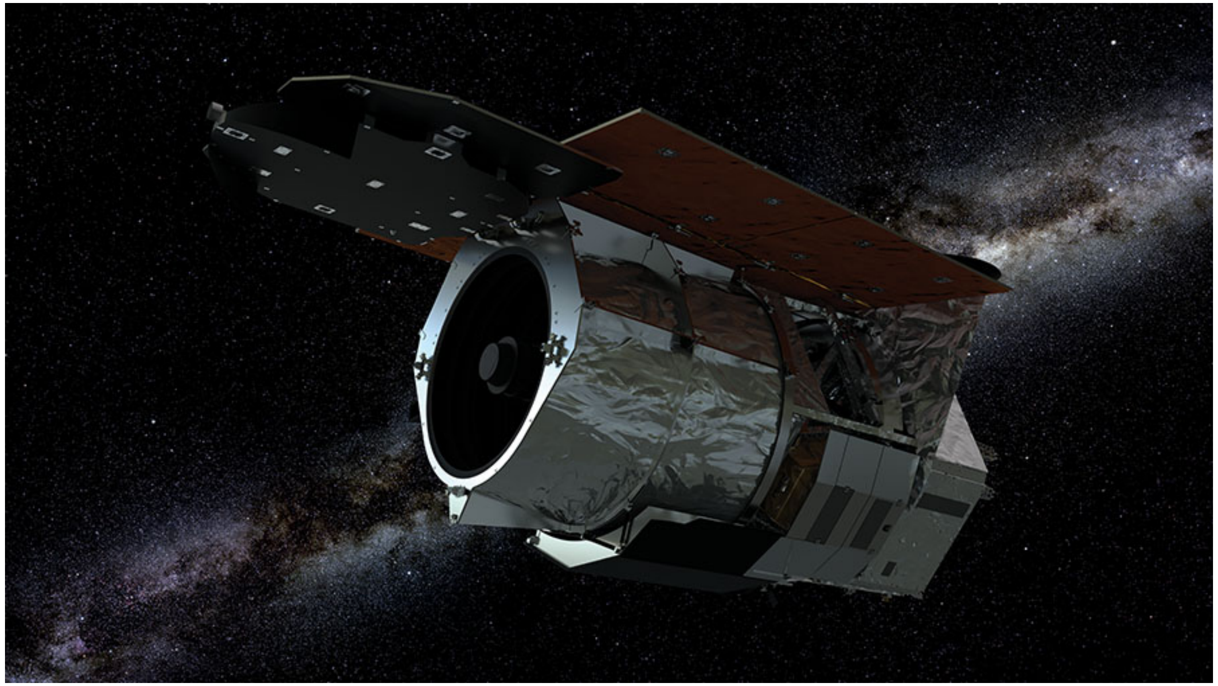
- Launch Delay brings time for additional new tasks.
  - Explore cloud storage options for some JWST data
  - Explore Jupyter notebooks for new data processing ideas.
  - Explore making those Jupyter notebooks available to all users of JWST data.

Wide Field Infrared Space Telescope

Wide Field Infrared Space Telescope (WFIRST)

Wide Field Infrared Space Telescope (WFIRST)

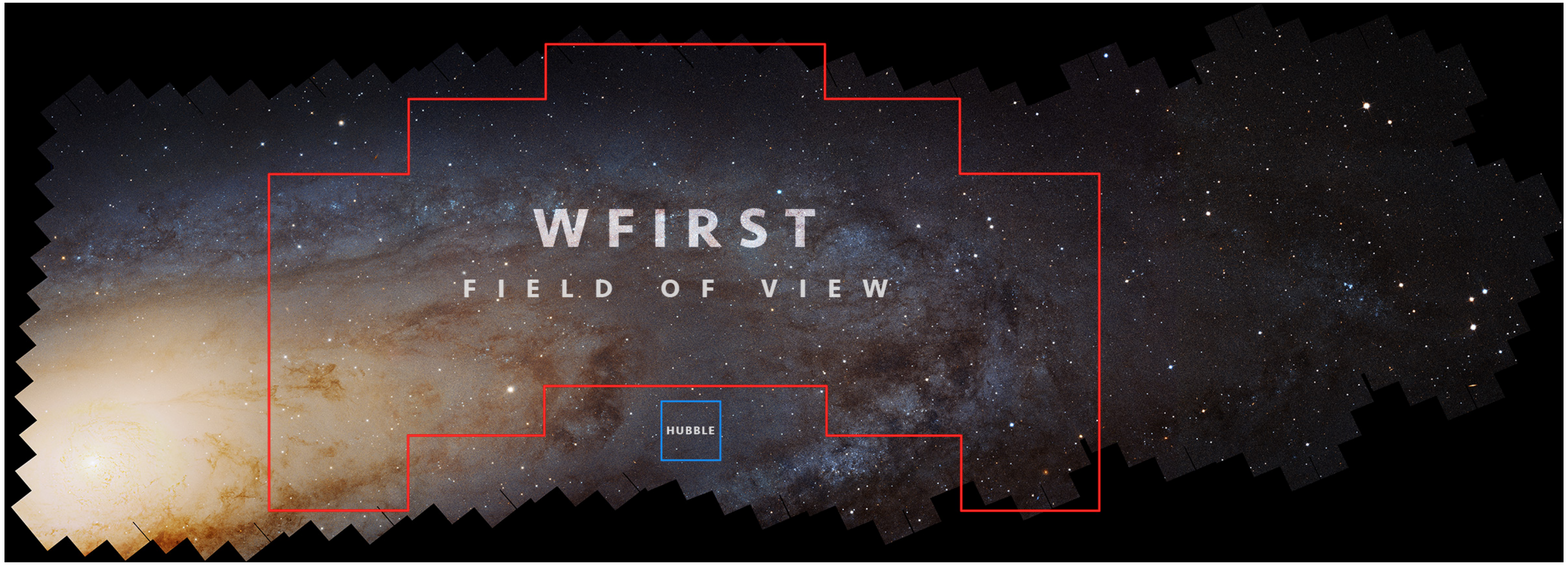
- 2.4 Meter Primary Mirror - Same size as Hubble Space Telescope's primary mirror, but one fifth the weight of the Hubble mirror
- 5 year primary mission w/possible 5 year extension
- Launch date mid 2020's
- Two Instruments:
  - Wide Field Instrument (WFI)
    - STScI responsible for data processing
    - STScI responsible for archiving
  - Coronagraph Instrument(CGI)
    - IPAC responsible for data processing
    - STSci responsible for archiving



Wide Field Infrared Space Telescope (WFIRST)

- Even though the mirror is the same size and resolution as the HST mirror, the Wide Field Instrument field of view is 100 times larger.

- 1.1 Terabyte Solid State Recorder
- 1.4 Terabytes downlinked data per day
  - 24+ Petabytes data mission lifetime
  - 32 Terabits of onboard storage



Wide Field Infrared Space Telescope

- With this field of view and this large data volume, we have to rethink our data processing methods.
- Instead of delivering data and software to the astronomer to use on their computer, we are working to bring the astronomer to the data and the software.

- Some part of the data will be stored in the cloud (probably AWS)
- Some parts of the data processing software will live in notebooks (probably Jupyter Notebooks)
- We have to figure out what preprocessing will be done in house and what will be done in the cloud.
- We have to figure out what calibration work will work on all data vs what to put in the notebooks.
- We have to determine how and where to catalog the data for the archive so users can find data.
- We have to determine how batch processing, priorities, multiple users, security will apply.
- All of these questions will determine if and how HTCondor will be part of this new process.

Thank You!!

We at Space Telescope Science Institute would like to thank the HTCondor team at UW for their continued support and advice.

We would also like to thank the members of the HTCondor users list for their invaluable collective knowledge and willingness to share that knowledge.

Questions?