SOFIA Data Pipelines

Ed Chambers
on behalf of the DPS Team

November 4, 2019

FORCAST 19µm, 37µm; Herschel 70µm; Spitzer 3µm
USRA | NASA | SOFIA | Spitzer | Herschel

NASA/SOFIA/Lopez-Rodriguez;
NASA/Spitzer/Moustakas

Outline

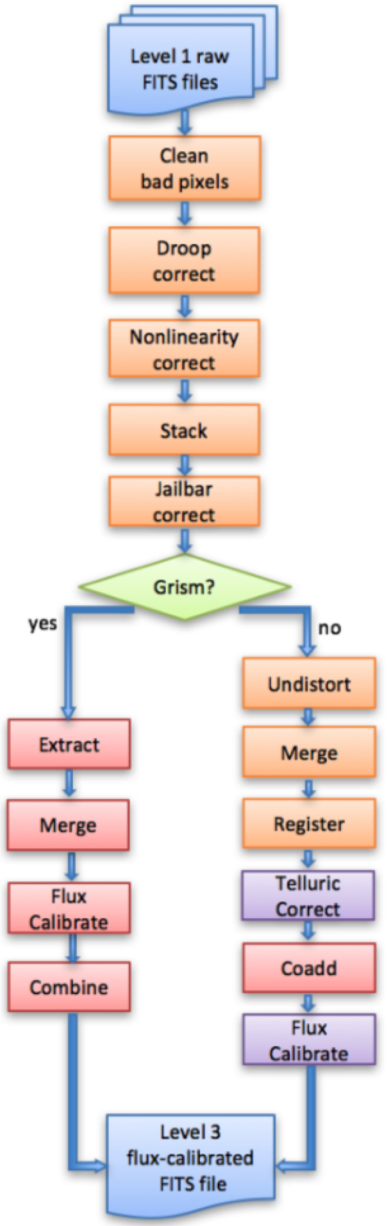
- Instrument Pipelines
- Pipeline Infrastructure
- Data Flow
- Data Products
- Future Plans

FORCAST Imaging

Original Delivery: From instrument team, in IDL

SMO modifications: Added non-linearity corrections, distortion corrections, telluric corrections, and flux calibration. Developed infrastructure to run pipeline automatically, plus interface for manual reductions, and data viewer (Redux).

Current Status: IDL in production. Imaging pipeline has been translated to Python.

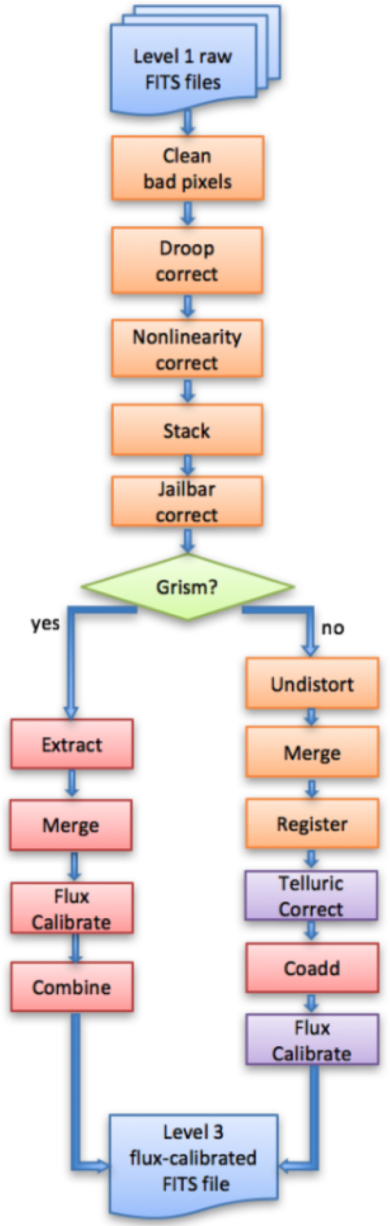


FORCAST Spectroscopy

Original Delivery: From instrument team, in IDL

SMO modifications: non-linearity corrections, wavelength calibrations, telluric corrections, and flux calibration. SMO replaced the spectral tracing and extraction algorithms with a modified version of Spextool, which incorporates optimal extraction and many other improvements. SMO developed infrastructure to run pipeline automatically, plus interface for manual reductions, and data viewer (Redux)

Current Status: IDL in production. Conversion of the spectroscopy pipeline to Python is underway.

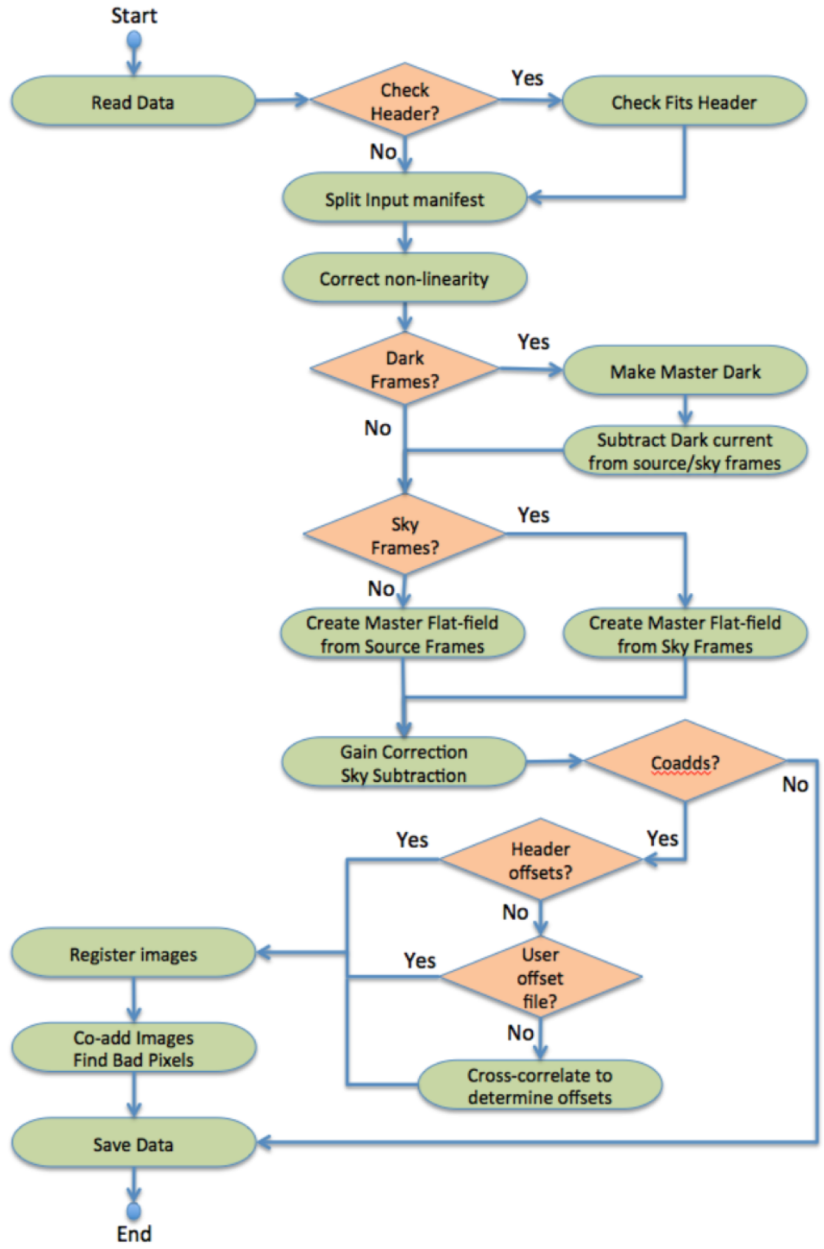


FLITECAM Imaging

Original Delivery: From instrument team, in IDL

SMO modifications: Added non-linearity corrections, telluric corrections, and flux calibration. Developed several improvements to algorithms for bad pixels and darks. Modified to run under Redux environment

Current Status: IDL in production. No Python conversion of the pipeline has started.

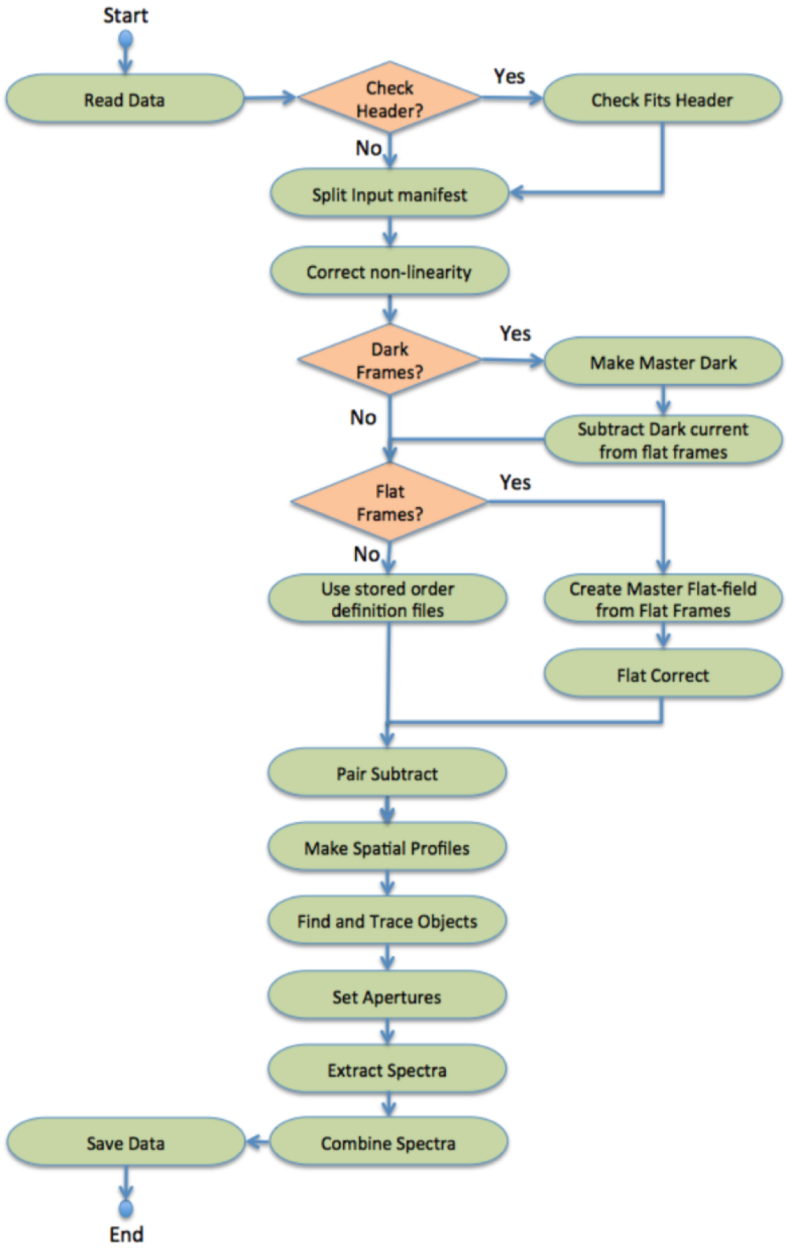


FLITECAM Spectroscopy

Original Delivery: None

SMO modifications: SMO created a pipeline by modifying Spextool to reduce, extract, telluric-correct, and flux calibrate spectra

Current Status: IDL in production. No Python conversion of the pipeline has started.

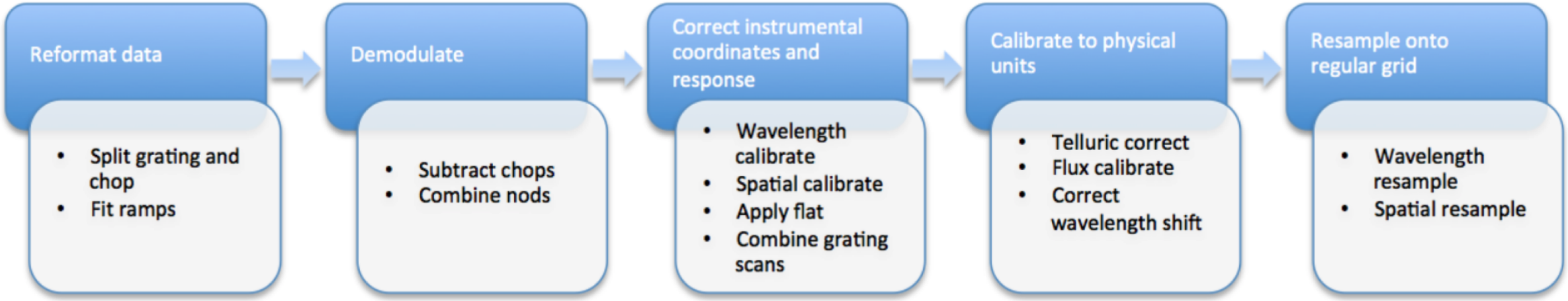


FIFI-LS

Original Delivery: Fragments of pipeline from instrument team, in IDL

SMO modifications: Developed entire IDL pipeline from the ground up, including resampling algorithms, telluric corrections, flat fielding, flux calibration, and adaptive smoothing

Current Status: Fully converted to Python.



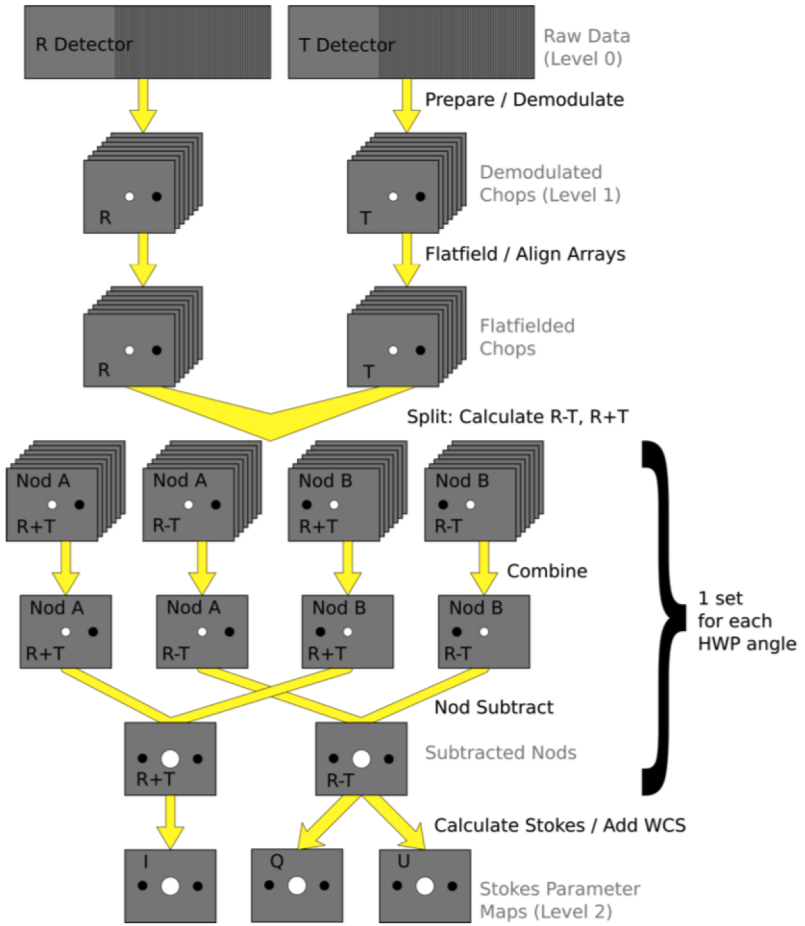
HAWC+ Chop-Nod and Nod-Pol

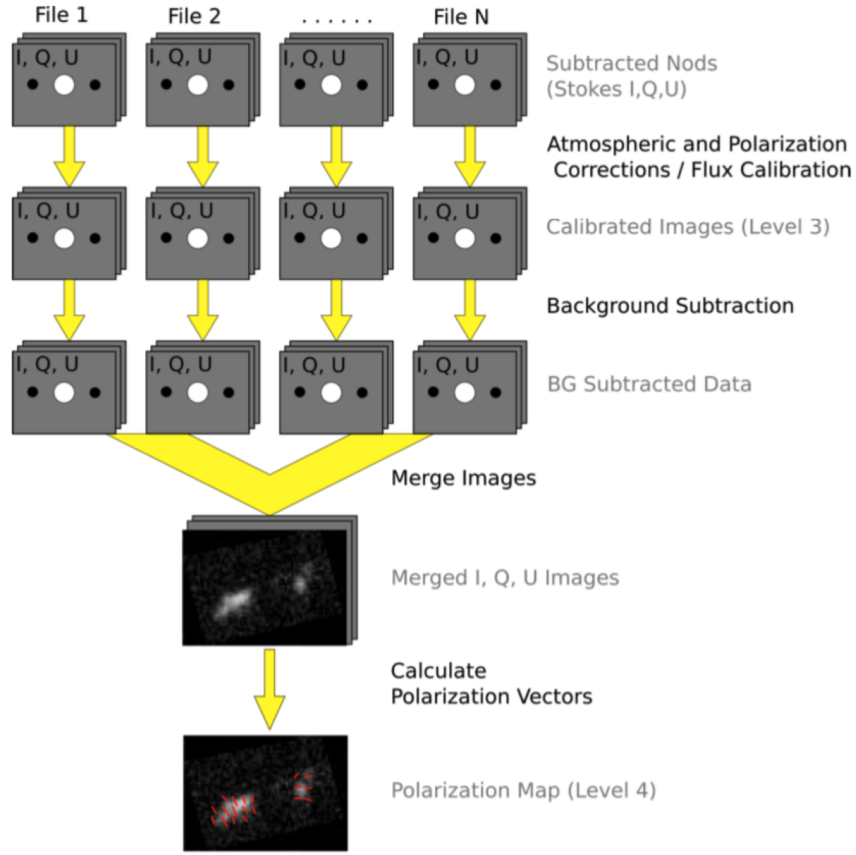
Original Delivery: From instrument team, written in C, IDL, and Python

SMO modifications:

Added telluric correction, flux calibration, and uncertainty calculation and propagation. Revised pipeline to fit within the Redux infrastructure.

Current Status: Fully converted to Python.



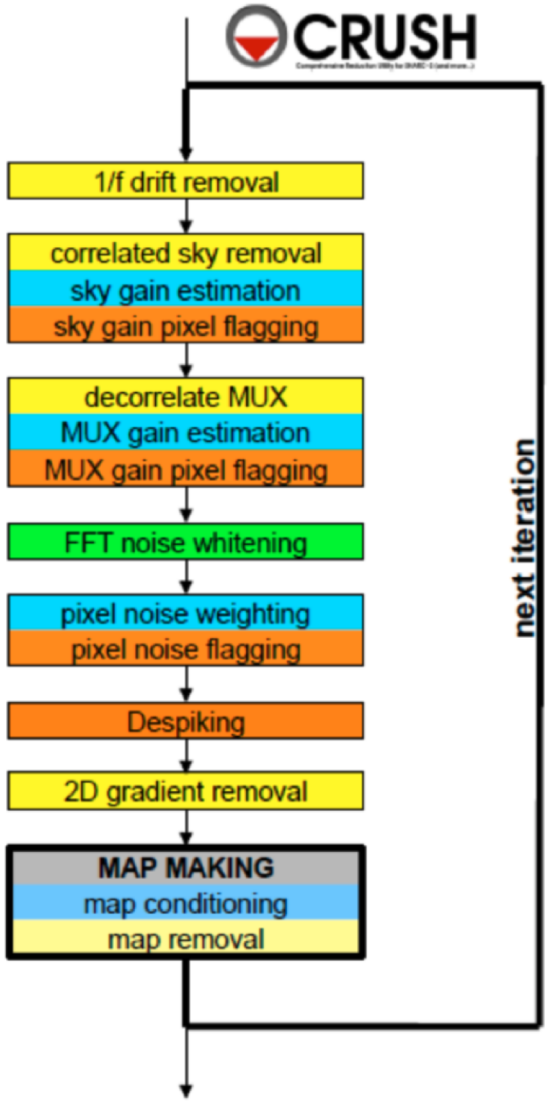


HAWC+ Scanning and Scan-pol

Original Delivery: From instrument team, written in Java(CRUSH), IDL, and Python

SMO modifications: Added telluric correction, flux calibration, and uncertainty calculation and propagation. Revised pipeline to fit within the Redux infrastructure. Modified CRUSH to permit scan-pol mode data to be reduced.

Current Status: All except CRUSH fully converted to Python. Conversion of CRUSH to Python is ongoing.

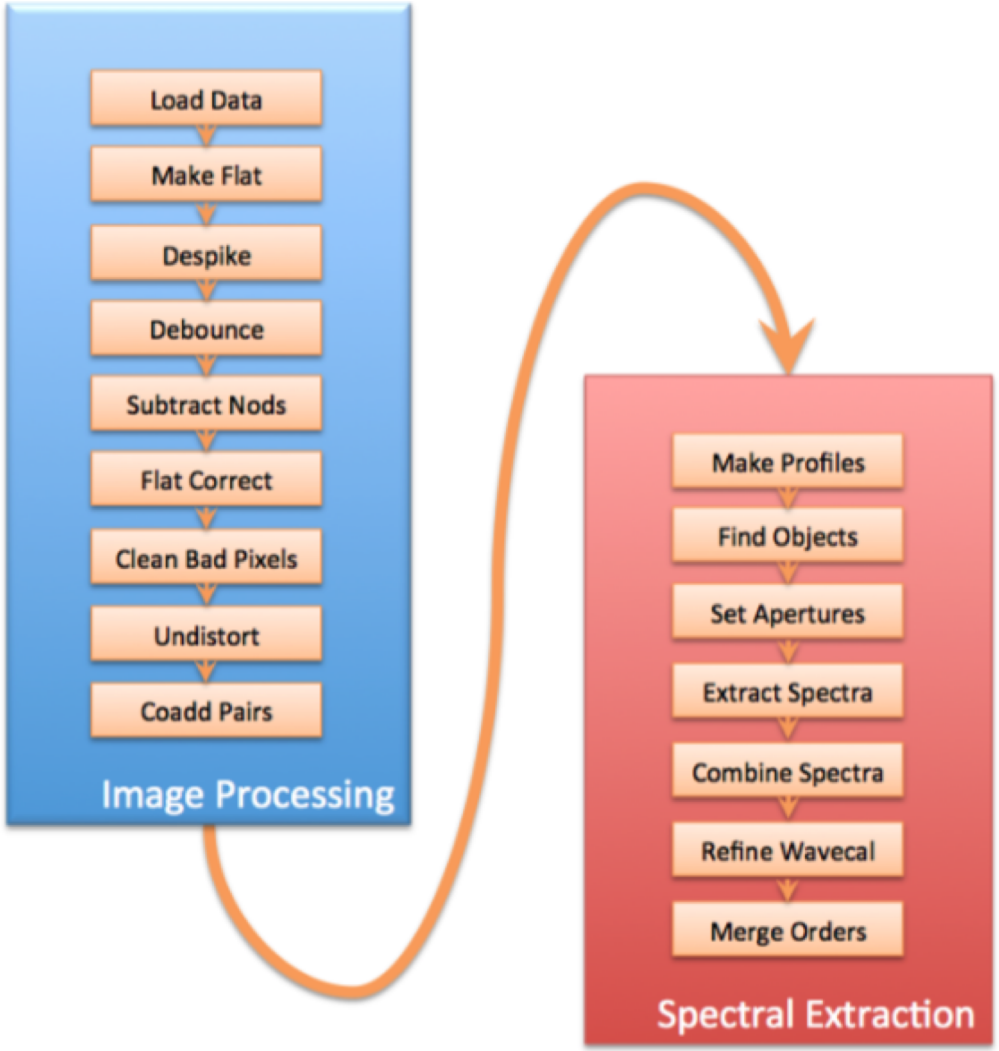


EXES

Original Delivery: FORTRAN TEXES pipeline from instrument team

SMO modifications: Translated it to IDL, modified it to work correctly for EXES, and adapted it to run with Spextool and Redux. Added non-linearity corrections.

Current Status: IDL in production, instrument team supports their own (forked) version of the pipeline. No Python conversion of the pipeline has started. SMO not currently authorized for EXES pipeline development.

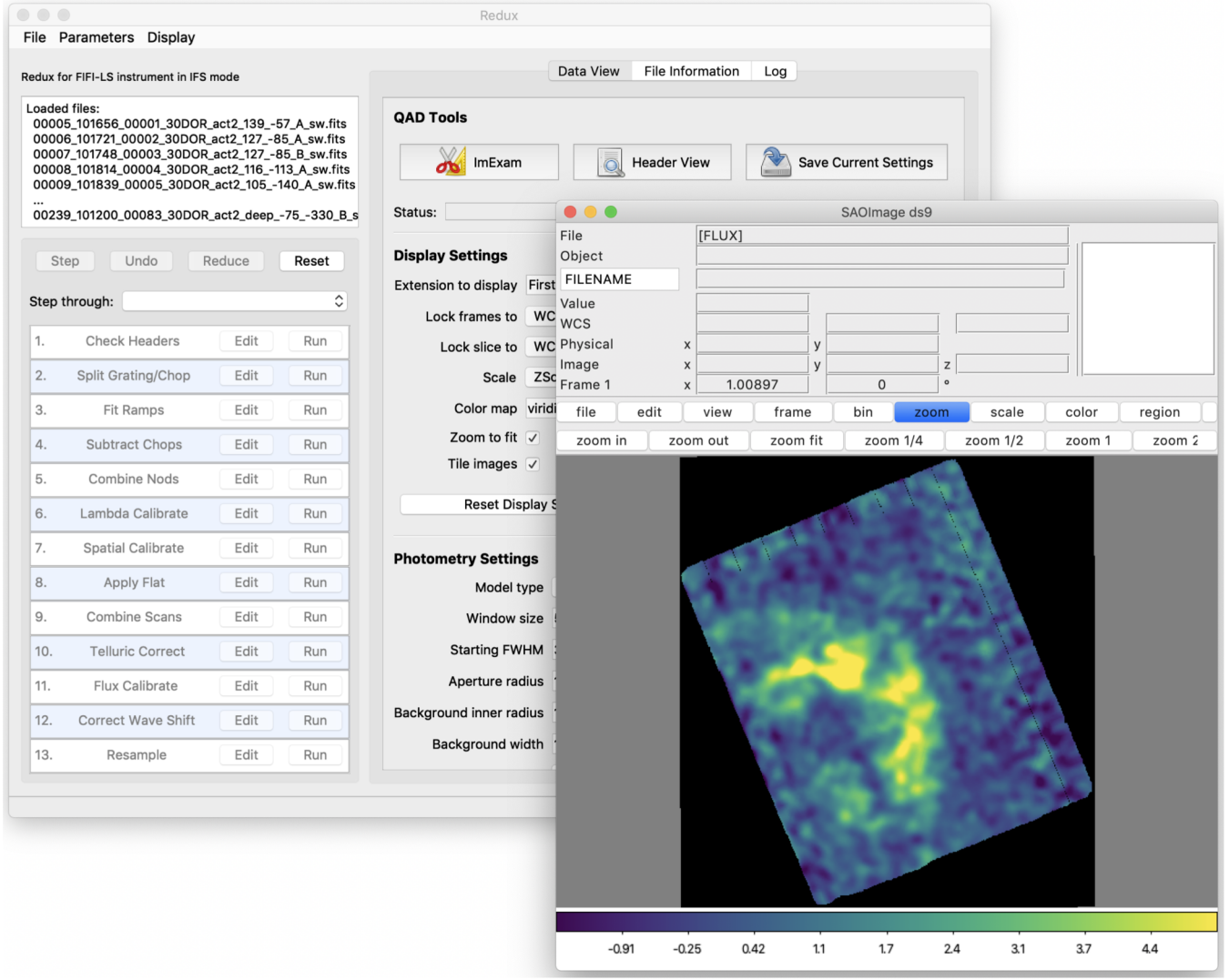


Pipeline Infrastructure

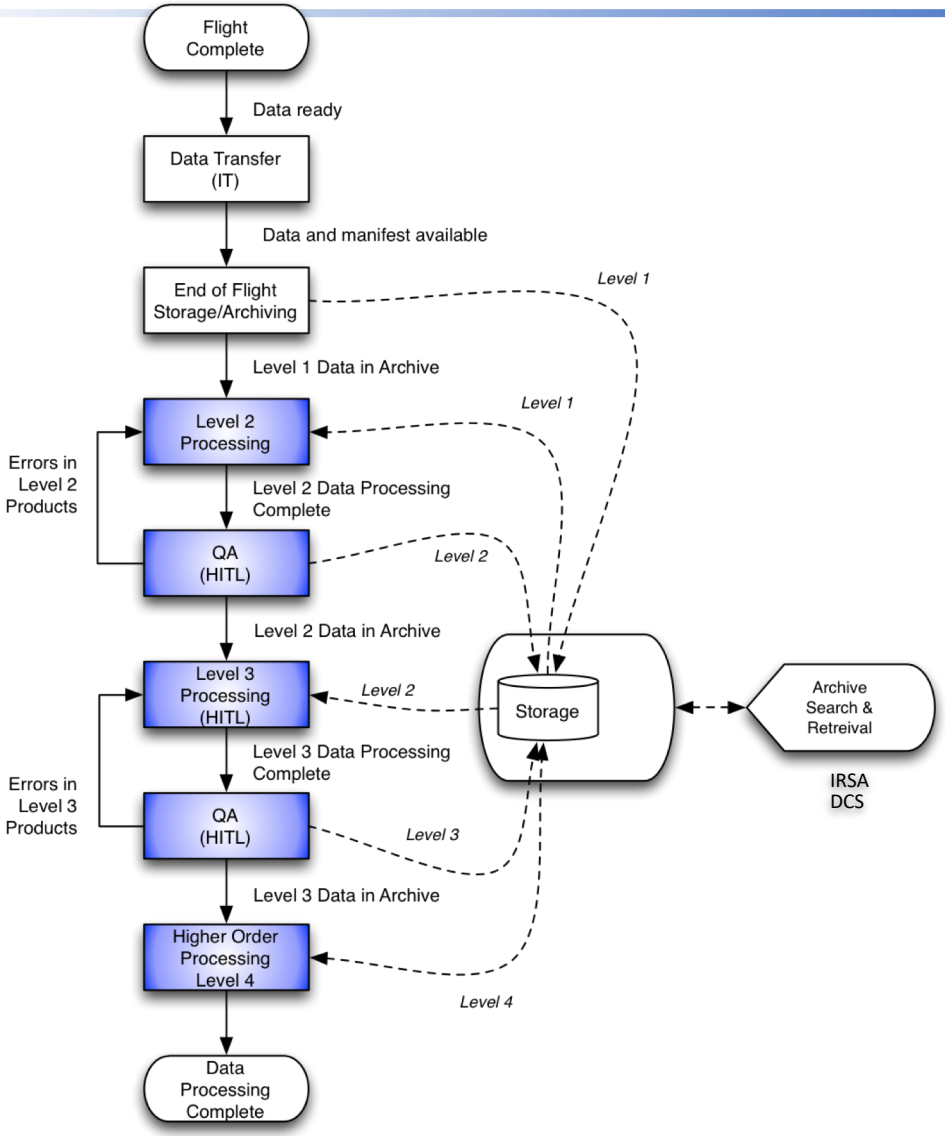
pipe-tools is a set of python-based utilities that group and stage the data in preparation for pipelining.

(py)redux (*right*) is a flexible, extensible framework for performing astronomical data reduction. Two versions of redux are currently supported--IDL (FORCAST, FLITECAM) and python (HAWC+, FIFI-LS).

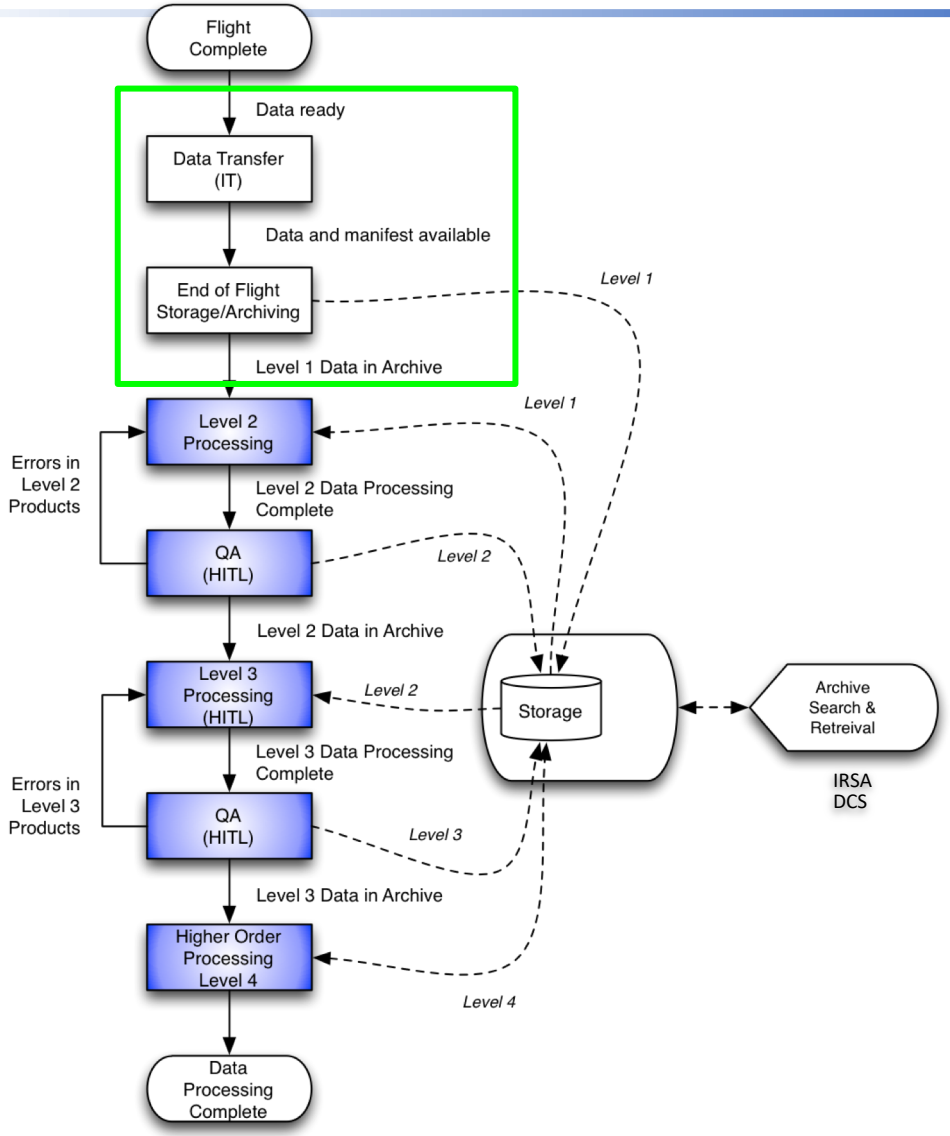
This infrastructure is maintained by the DPS team.



Data Flow


Flight Complete
Data ready
Data Transfer (IT)
Data and manifest available
End of Flight Storage/Archiving
Level 1
Level 1 Data in Archive
Level 2 Processing
Level 1
Errors in Level 2 Products
Level 2 Data Processing Complete
QA (HITL)
Level 2
Level 2 Data in Archive
Level 3 Processing (HITL)
Level 2
Storage
Archive Search & Retreival
IRSA DCS
Level 3 Data Processing Complete
Errors in Level 3 Products
QA (HITL)
Level 3
Level 3 Data in Archive
Higher Order Processing Level 4
Level 4
Data Processing Complete

Data Flow -- Level 1

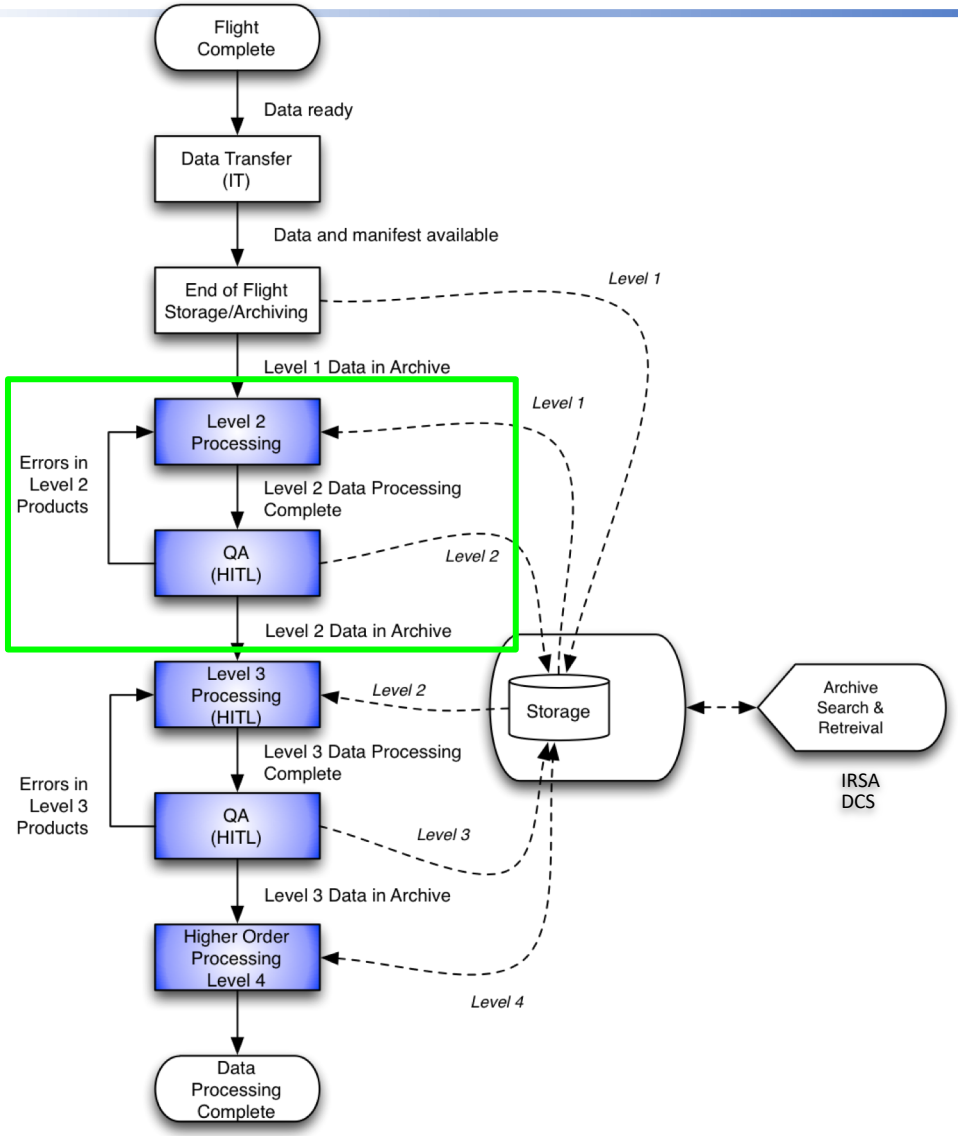


Level 1:

Processing happens after each flight, once the data are transferred from the plane. As part of this process, the DPS:

- checks data integrity
- updates FITS headers (header_checker, Python code supported by DPS team)
- archives the data

Data Flow -- Level 2



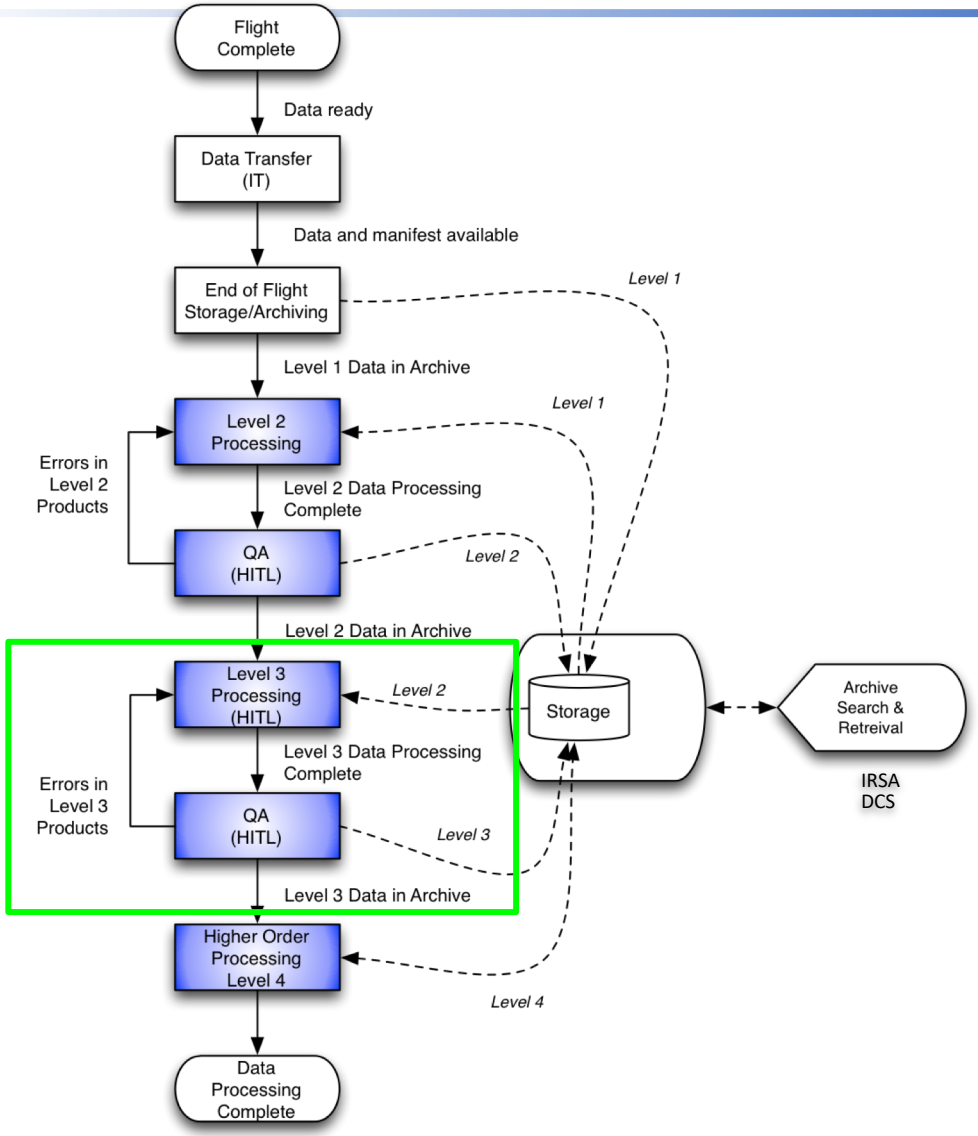
Inputs for Level 2 Processing

- bad pixel masks
- non-linearity correction factors (can be 3-D data files)
- flat fields
- wavelength calibration files
- slit correction functions
- telluric correction files

These inputs change over time.

Initial pipelining for each flight is performed automatically (using pipe-tools, redux). QA is done by instrument scientists, who update headers, adjust coordinates, remove bad files, then re-run the pipeline manually.

Data Flow -- Level 3



Inputs for Level 3 Processing

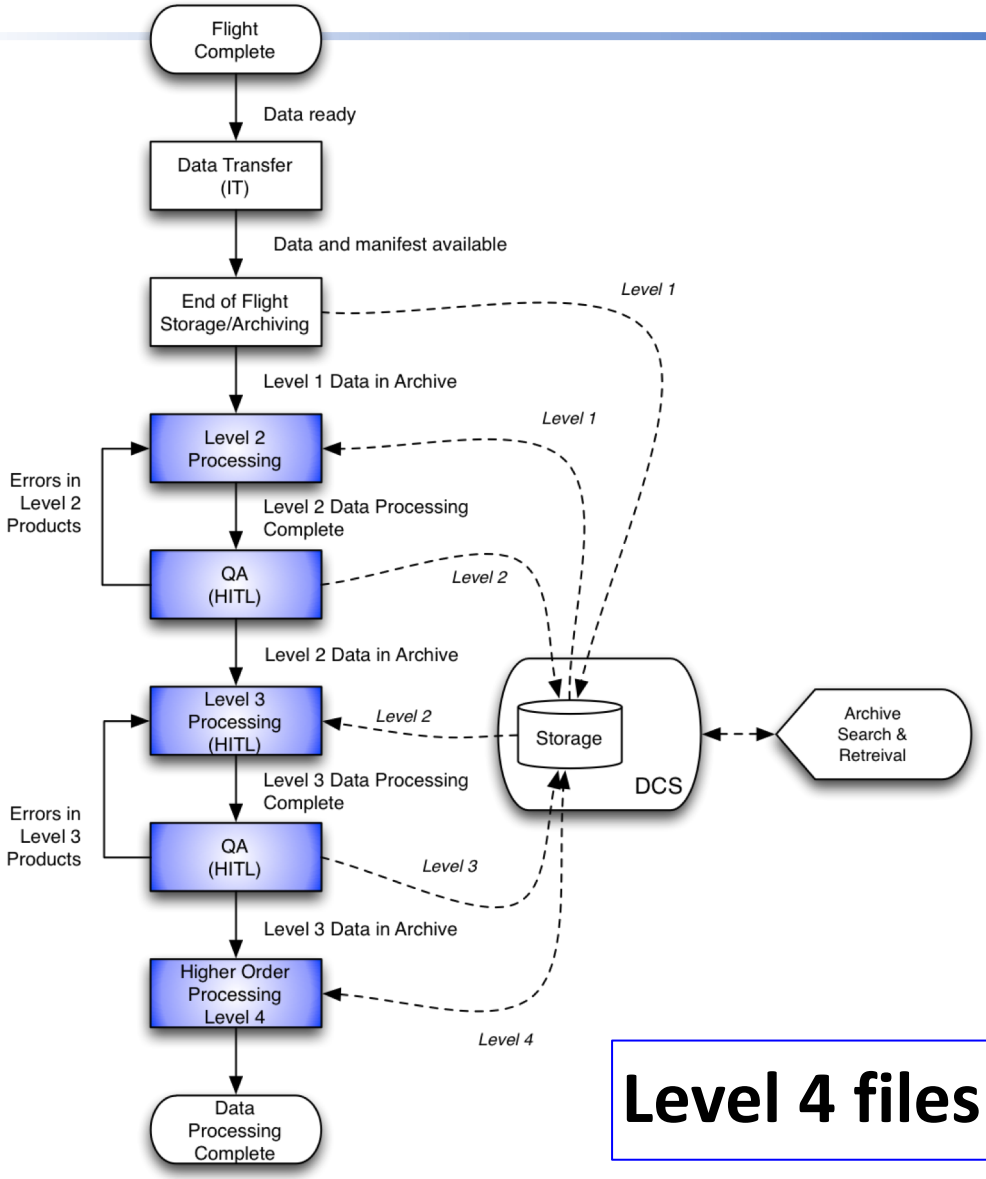
- observations of calibrators

- modules to generate fluxes of calibration standards

Calibration observations are made during every flight series

Calibrator data are reduced to Level 2 using the pipeline. These L2 data, along with instrument profiles and calibrator information, are used by the pipe-cal tool to generate data calibration factors for each filter/wavelength. Calibration factors are reviewed and analyzed by instrument scientists, and then applied to the appropriate Level 2 data to create Level 3 data.

Data Flow -- Level 4



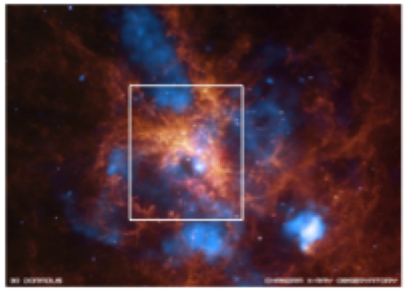
- 'Enhanced' Data Products
  - HAWC+ polarization maps and region files
  - Slit scan data cubes from FORCAST and EXES
  - FIFI-LS data cubes
- 'Combined' Data Products
  - Maps generated from observations carried out over multiple flights and/or observing cycles
  - Large area maps generated by FORCAST, FIFI-LS, and HAWC+
- 'Combined' + 'Enhanced' Products:
  - HAWC+ large area polarization maps
  - FIFI-LS large area data cubes.

Level 4 files are science-ready data products

Level 4 ‘Combined’ + ‘Enhanced’ Data Products

DDT Science Target: 30 Dor (LMC)

Program: 76_0001 PI: H. Yorke

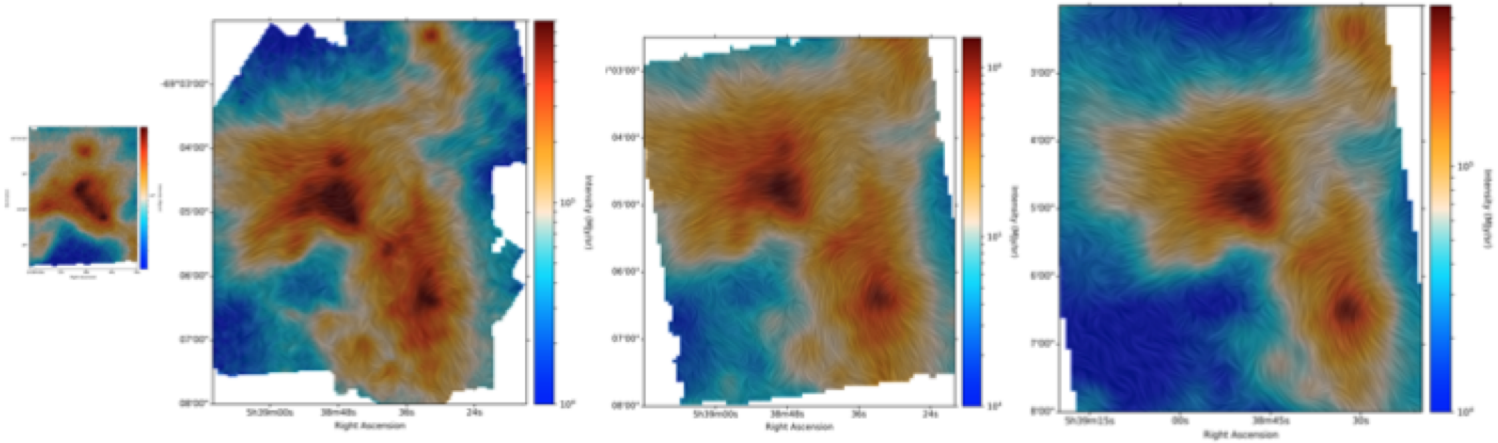


HAWC+ Band A (53 microns)

HAWC+ Band C (89 microns)

HAWC+ Band D (154 microns)

HAWC+ Band E (214 microns)

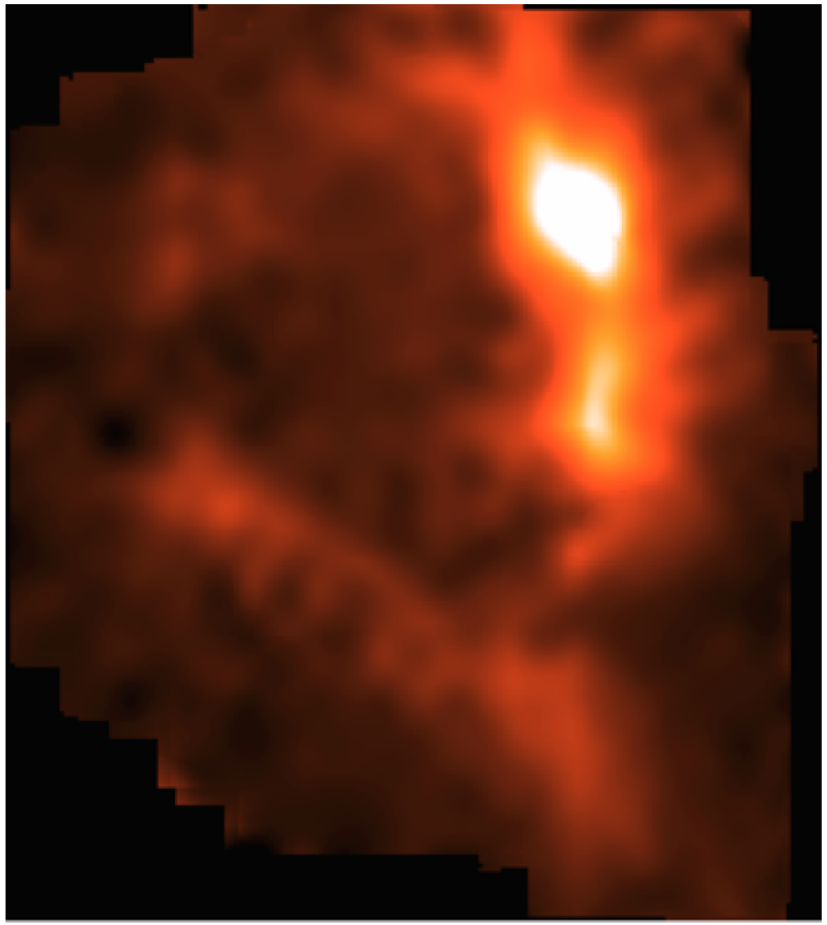


HAWC+ observations from F481, F483, and F484!

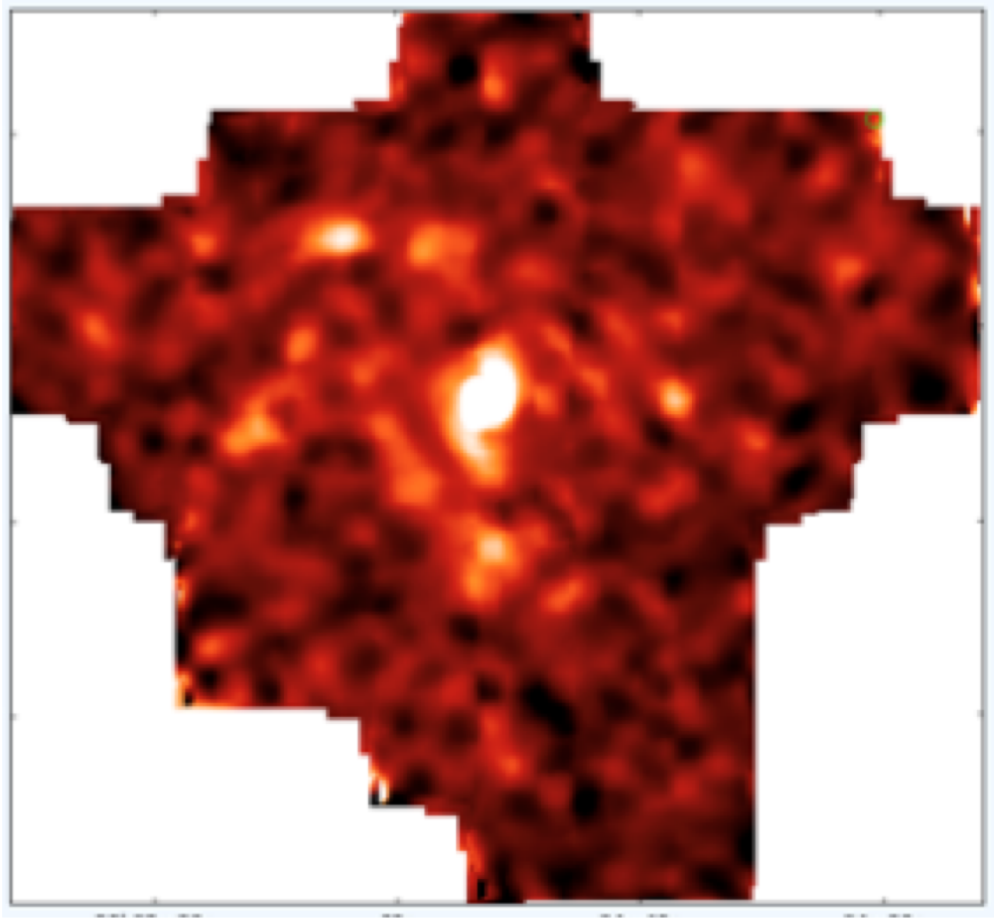
Level 4 'Combined' Data Products

FIFI-LS Observations

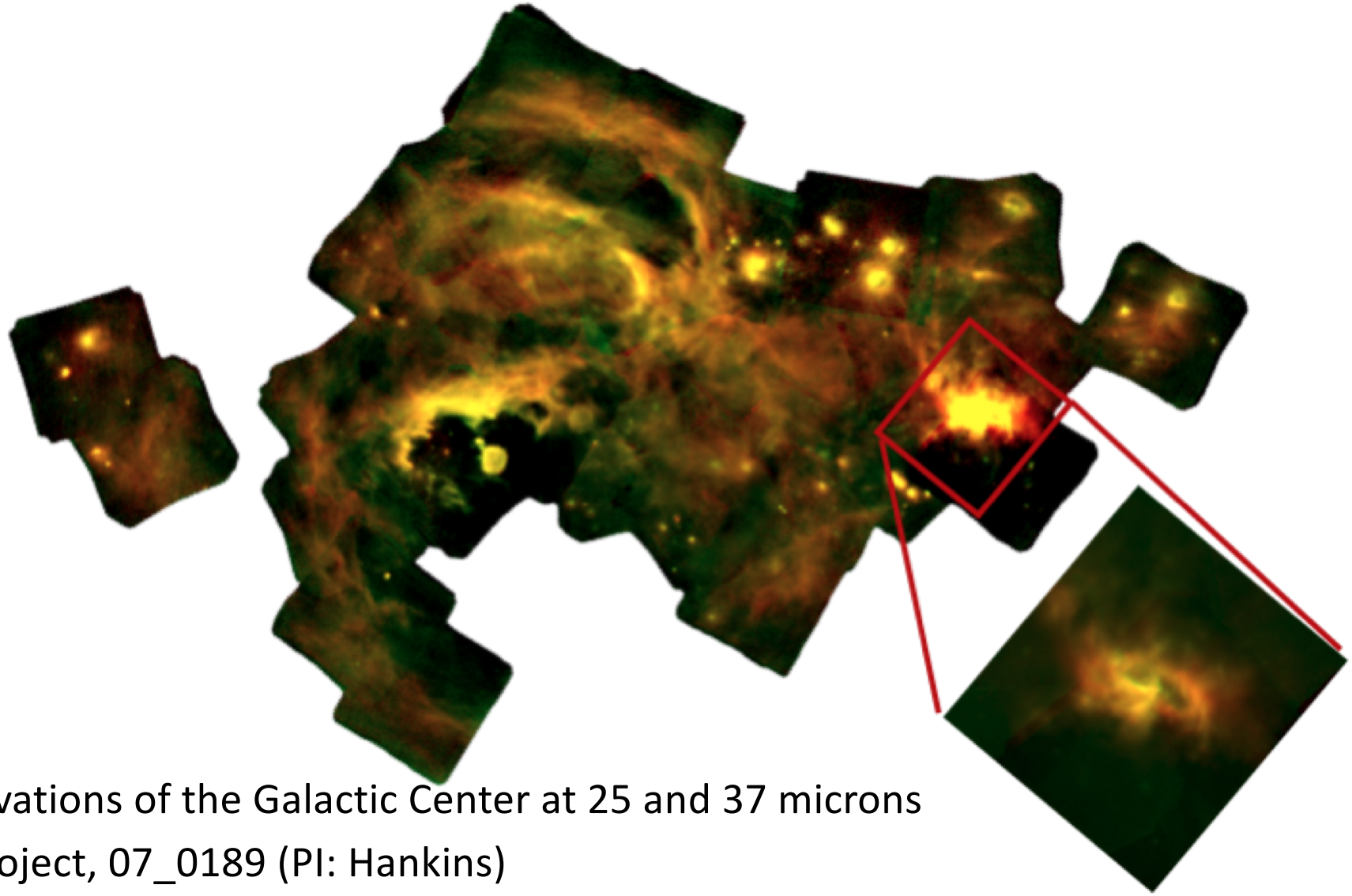
Orion, 145 micron



NGC 4946, 158 micron



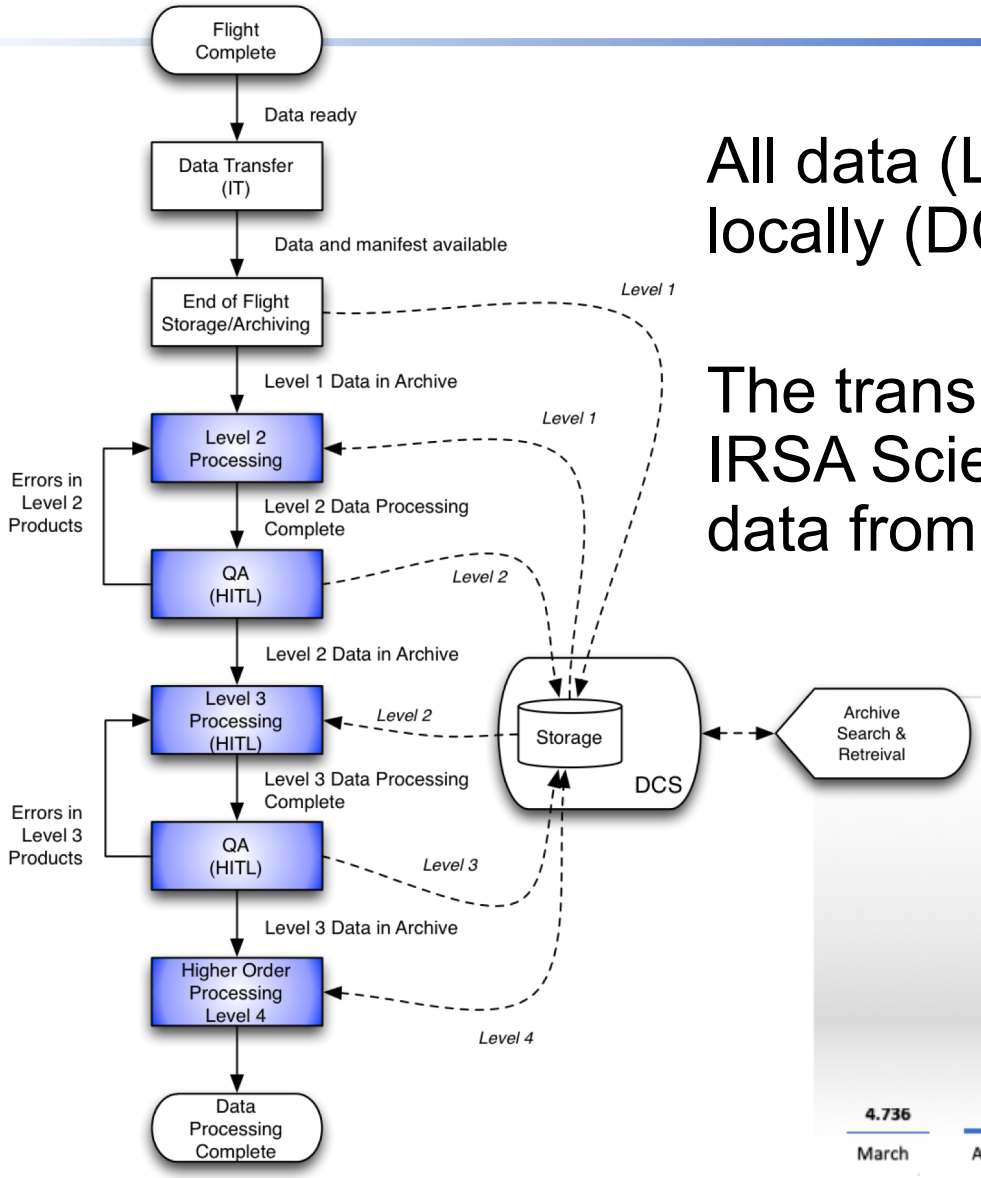
Level 4 ‘Combined’ Data Product



FORCAST Observations of the Galactic Center at 25 and 37 microns

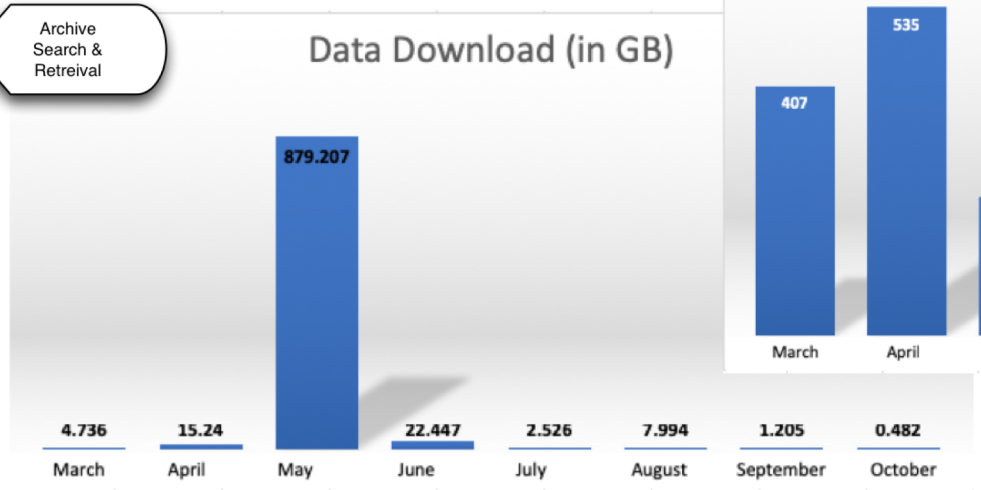
SOFIA Legacy Project, 07_0189 (PI: Hankins)

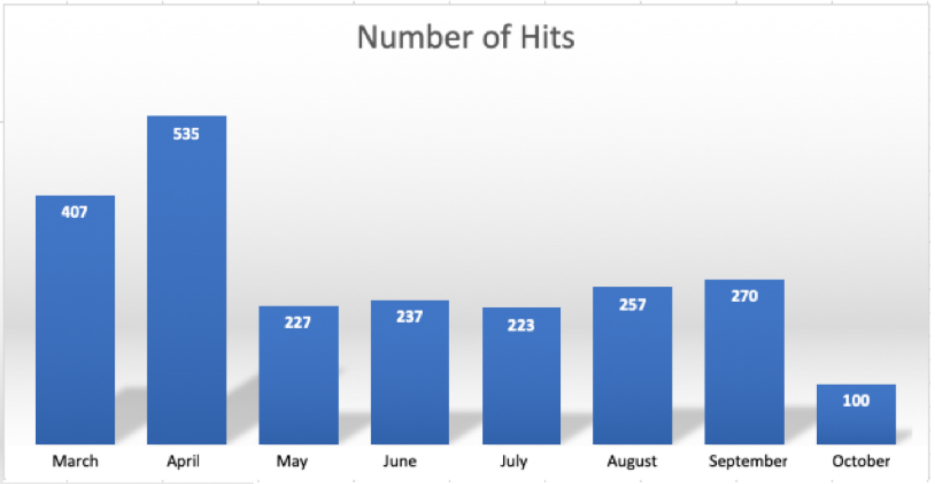
Data Flow -- Archive



All data (Levels 1 through 4) are stored and archived locally (DCS).

The transition to the IRSA archive is ongoing. The SOFIA-IRSA Science Archive Second Release Review (including data from Cycles 4-6) is scheduled for 8 November 2019.

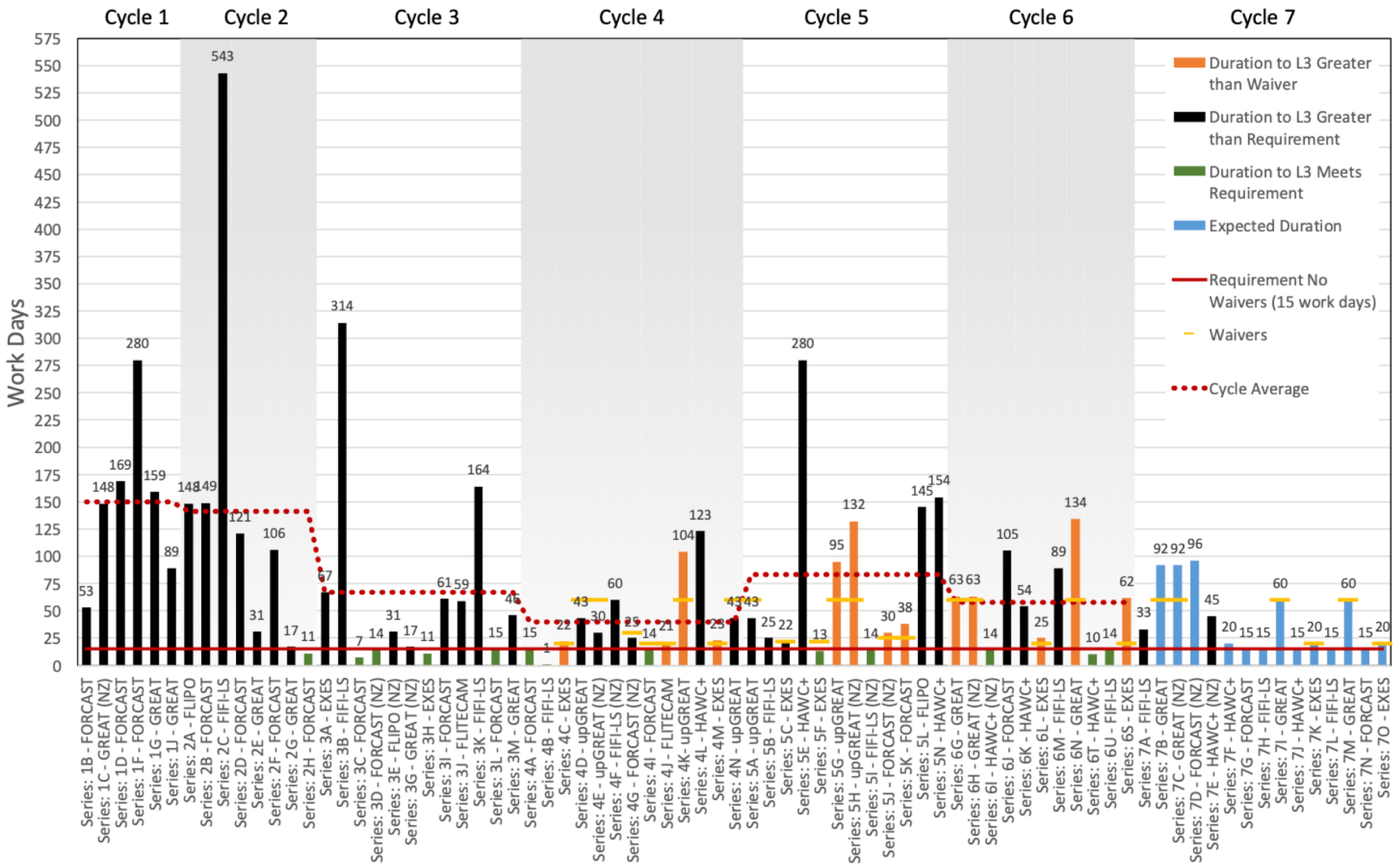




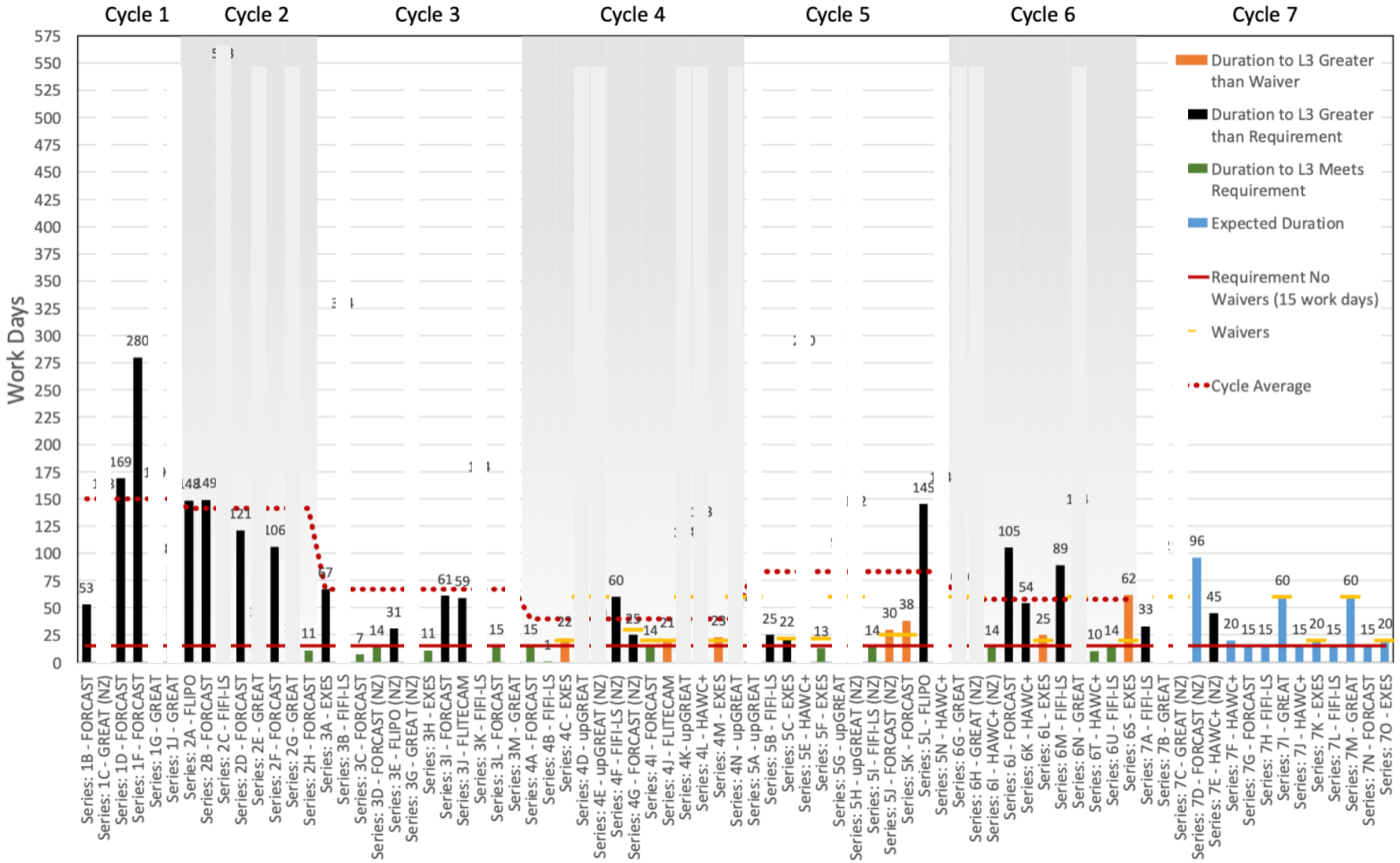
Future DPS Plans

- Continue porting all pipelines and modules to Python
  - CRUSH module for HAWC+ pipeline
  - Spextool (for FORCAST, FLITECAM, EXES pipelines)
  - FLITECAM pipeline
- Development of a universal SOFIA Data Viewer
- General maintenance of all pipelines, tools, calibration database, and other software
- **Under consideration**: Public release of data pipelines
  - Convert all pipelines to Python
  - Develop infrastructure for external release
  - Create user guides and documentation
  - Develop release schedule

Level 3 Data Archive Timeline



Level 3 Data Archive Timeline



Future DPS Plans

- Continue porting all pipelines and modules to Python
  - CRUSH module for HAWC+ pipeline
  - Spextool (for FORCAST, FLITECAM, EXES pipelines)
  - FLITECAM pipeline
- Development of a universal SOFIA Data Viewer
- General maintenance of all pipelines, tools, calibration database, and other software
- **Under consideration**: Public release of data pipelines
  - Convert all pipelines to Python
  - Develop infrastructure for external release
  - Create user guides and documentation
  - Develop release schedule
  - Create user support plan