JWST Calibration Pipeline

H. Bushouse & K. Gordon

14 Sep 2017

Learning From Past Efforts

- Previous Missions
  - Hubble, Spitzer, Herschel, etc.
  - Ground-based observatories and instruments
    - Especially important for integral field spectroscopy (IFU), coronagraphy, and multi-object spectroscopy (MOS)
- What to do
- What not to do

Pipeline Philosophy

- Algorithms based on community best
  - Input from instrument teams and mode-specific expert teams
  - Overall goal is best justified algorithms
- Use the same code for different instruments
  - Where possible
  - Easier to maintain
  - Takes advantage of strengths of all teams
- Provide pipeline directly to community

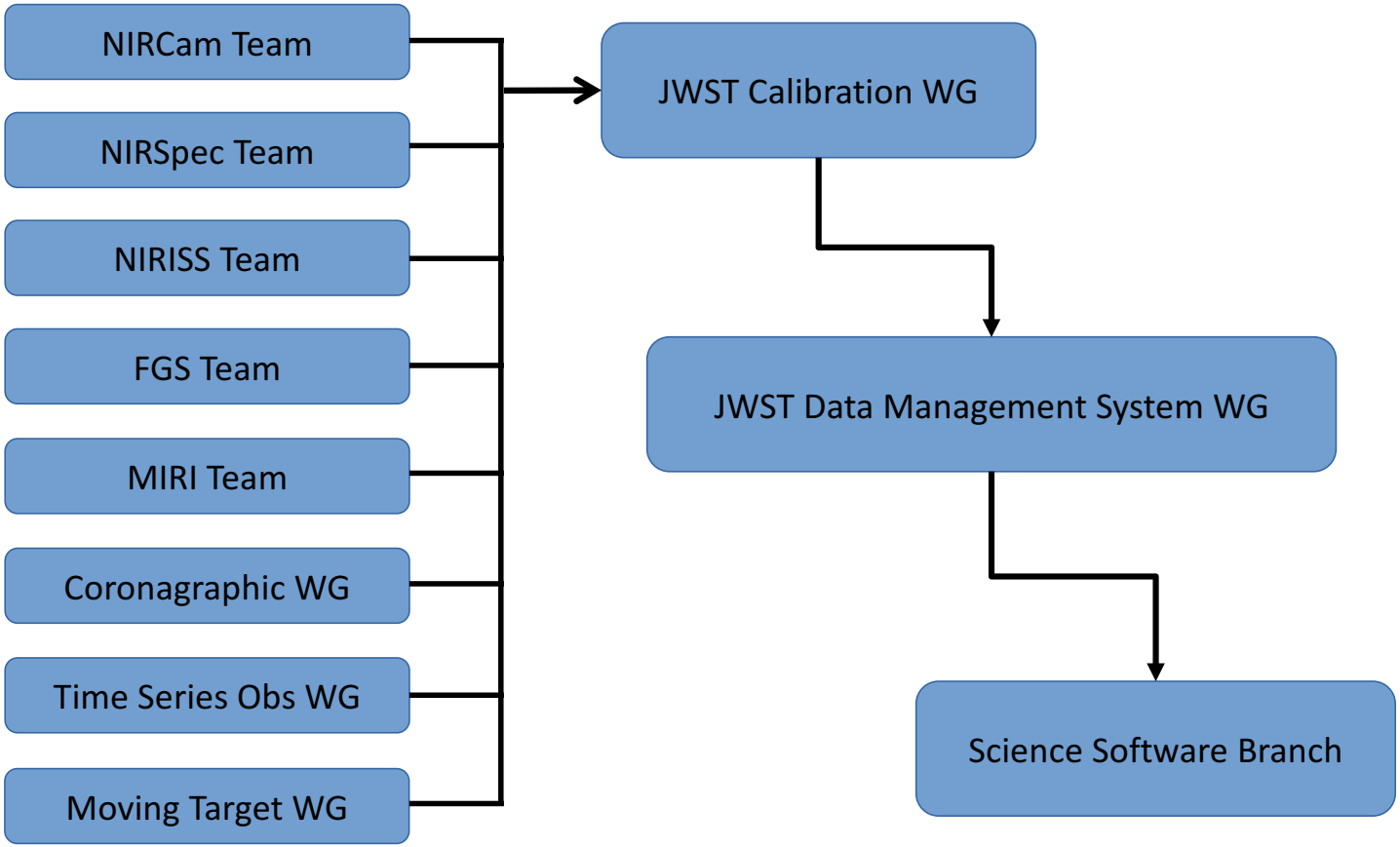
User Experience

- Pipeline automatically run on all data
- Default parameters for all pipeline steps
- Pipeline products produced and archived
  - Final as well as raw and intermediate products
- User can install and run pipeline locally
  - Change defaults
  - Add customized reduction steps
  - Access to or downloading of calibration reference files from STScI via CRDS web server

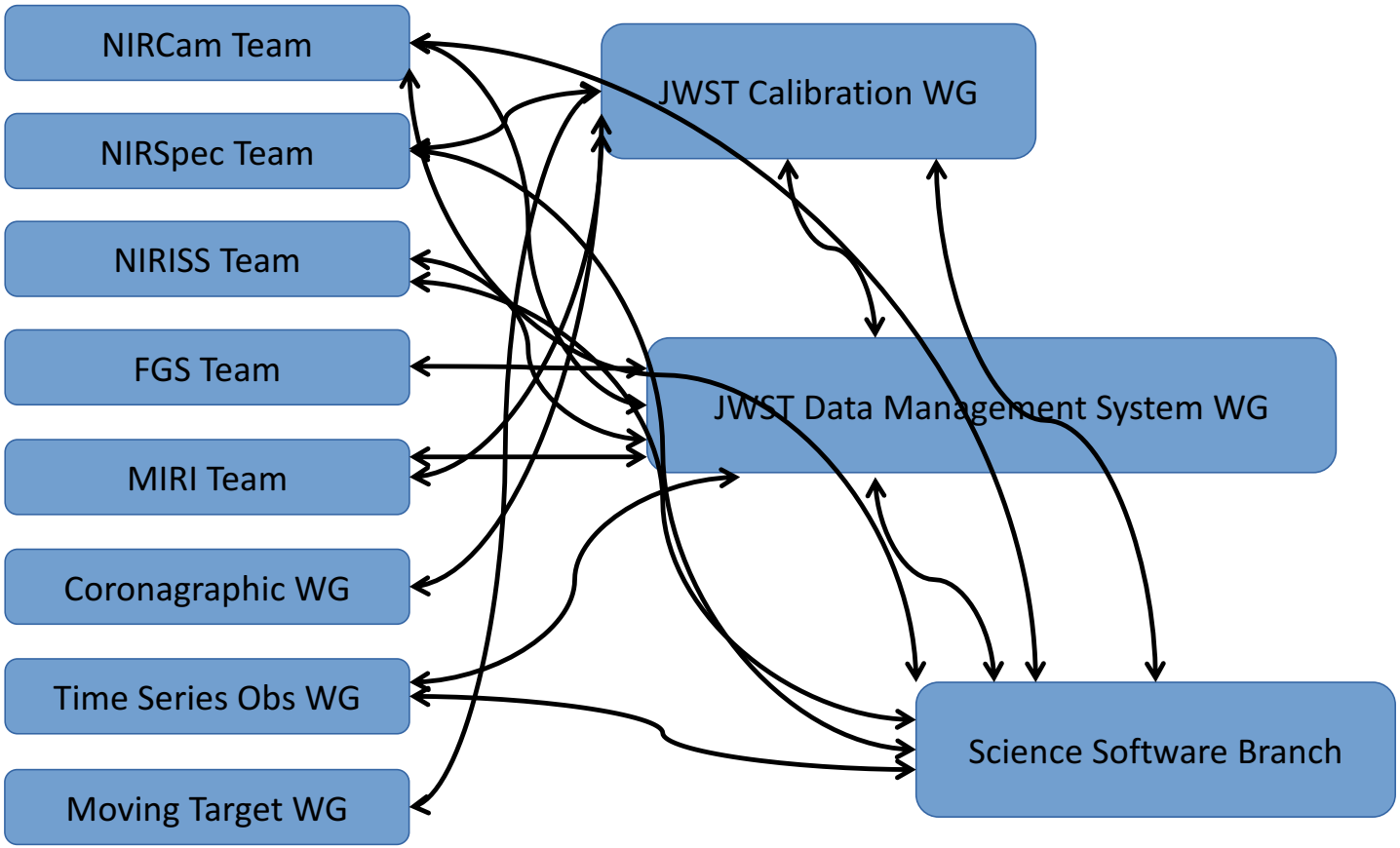
Pipeline versus Data Analysis Tools

- Pipeline
  - Automatically runs on all data, using default or best-guess parameters
  - Requires no human interaction
- Data Analysis Tools
  - Requires science decisions – human interaction
- Overlaps
  - For example, parts of the pipeline can be re-run interactively with non-default options

Official Work Flow



Real Work Flow



Development Plan

- Baseline Pipeline
  - All instruments and all modes
  - Provides good science products
  - Meets project-level requirements
  - Algorithms defined – implementation in progress now
  - Implementation done by Dec 2017
    - Final testing to happen in early 2018
- Optimal Pipeline
  - Best possible reductions; Highest quality science data
  - This is the final goal (launch + many years)
  - Start work after baseline pipeline done and continue for the mission lifetime and beyond
    - Will need to prioritize effort

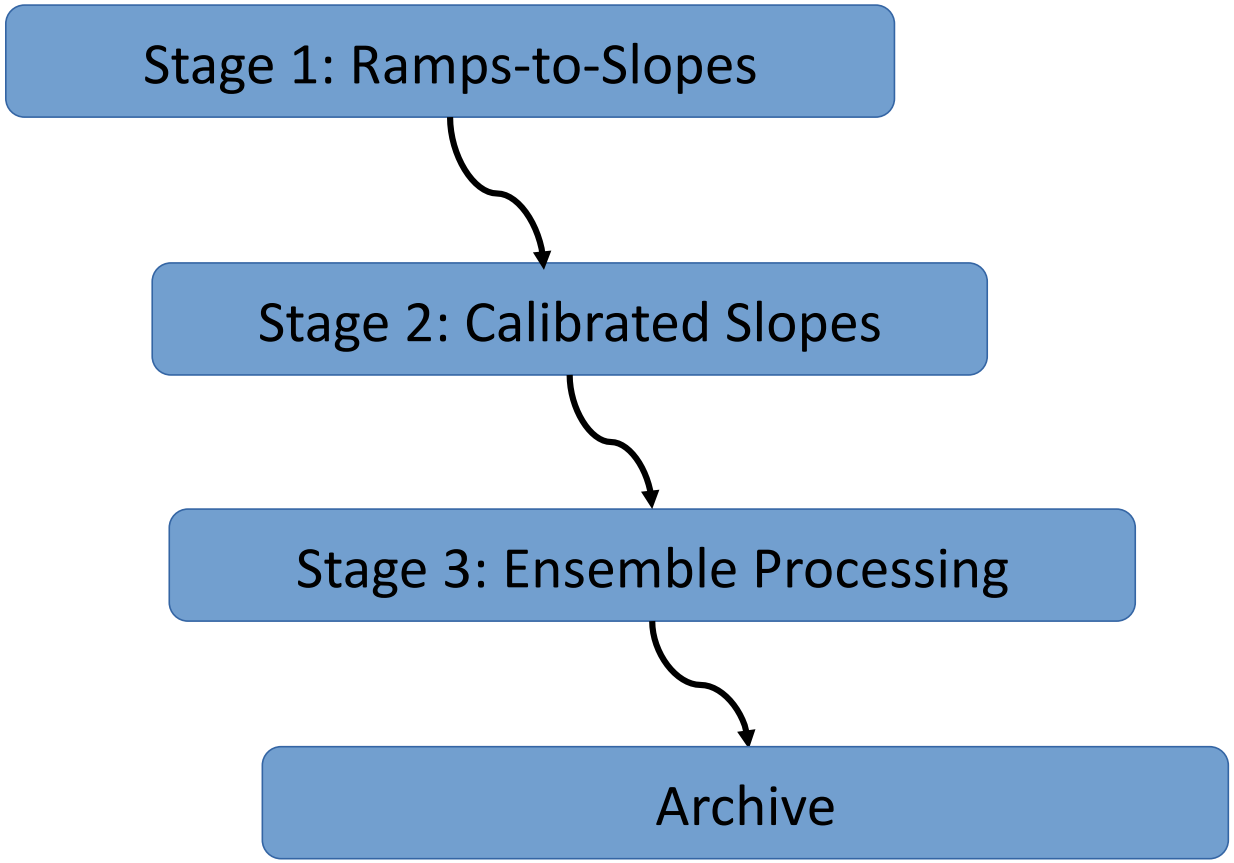
Pipeline Organization

- By observing modes
  - Imaging
  - Spectroscopy
    - Slit spectroscopy
    - Wide Field Slitless Spectroscopy
    - Integral Field Spectroscopy
  - Coronagraphy
  - Aperture Masking Interferometry
  - Time Series Observations
  - Moving Targets

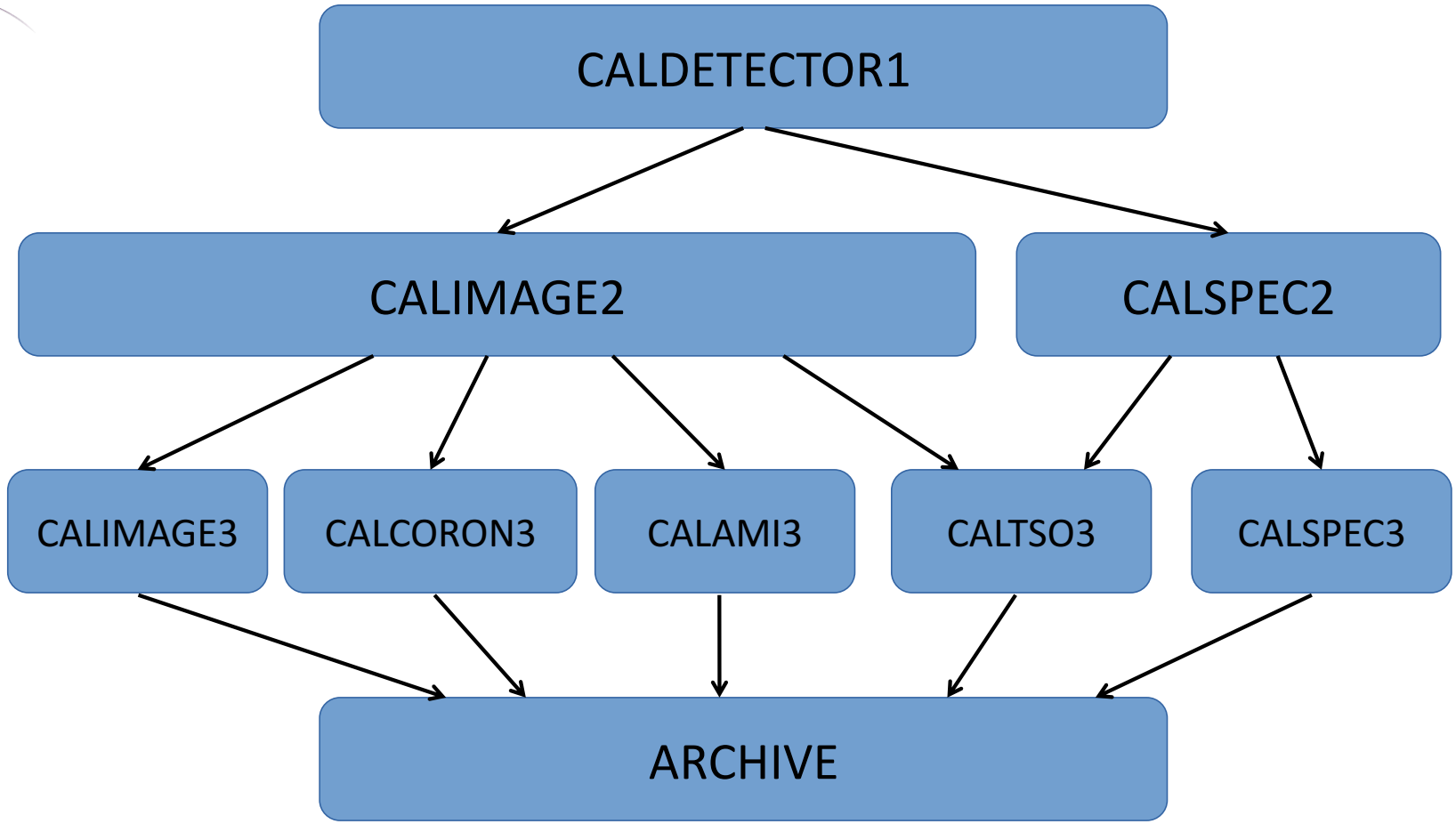
- By calibration activity
  - Detector signatures
  - Calibration
  - Background subtraction
  - Data combination & resampling
  - Spectral extraction

Much more ambitious than most previous HST pipelines

Pipeline Stages



Overall Architecture & Flow



CALDETECTOR1

- Bad pixel flagging
- Saturation flagging
- Superbias subtraction
- Reference pixel corrections
- Linearity correction
- Dark subtraction
- Persistence correction
- Jump/CR detection
- Ramp fitting

Stage 2 Calibration

CALIMAGE2

- Build WCS object
- Background subtraction
- Flat-field correction
- Flux calibration
- Image rectification

All processing up through this point applied to individual exposures

CALSPEC2

- Build WCS object
- Background subtraction
- 2D cutout extraction
  - Mode dependent
- Flat-field correction
- Point vs extended decision
- Pathloss/aperture correction
- Flux calibration
- Image rectification
- 1D spectral extraction

Stage 3 Calibration

CALIMAGE3

- Refine image offsets
- Background matching
- Outlier detection
- Image combination
- Source catalog

CALSPEC3

- Background matching
- Outlier detection
- Image combination
  - 2D or 3D rectified data
- 1D spectral extraction

All level 3 (and some level 2) processing relies on data Associations

- Background exposures to subtract
- Multiple exposures to combine
- PSF/reference target exposures for coronagraphy and AMI

Imaging Data Products

- Mosaic of all images with the same filter
  - In sky coordinates (RA, Dec)
- Catalog of sources
  - automated generation
  - source parameters (size, shape, photometry)
- Exposure level products
  - All images fully corrected with flagging of CRs, etc.
    - In detector & sky (rectified) coordinates

Spectroscopic Data Products

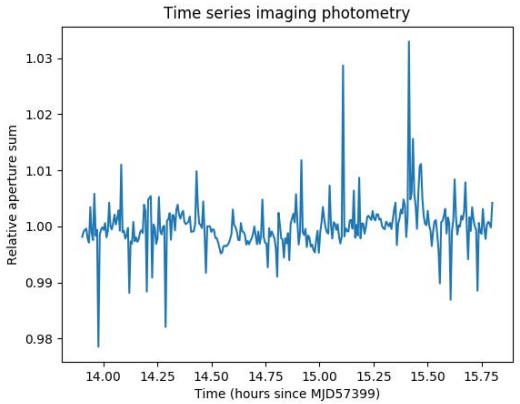
- Slit & Slitless: 2D spectral image(s)
  - Individual source products for multi-object modes (NIRSpec MSA, NIRCam & NIRISS Wide-Field Slitless)
- IFU: 3D spectral cube
  - RA, Dec, and wavelength
- Extracted spectrum for each source
  - Advice on point/extended type from APT, MPT, or WFSS source catalog
- Exposure level products
  - All images fully corrected with flagging of CRs, etc.
  - 3D cubes or 2D images depending on observation type

Coronagraphic and AMI Data Products

- Coadded PSF-subtracted images
- Integration level products
  - All images fully calibrated with flagging of CRs, etc.
    - In detector & sky (rectified) coordinates
- Tables of
  - Closure amplitudes and phases
  - Fringe amplitudes and phases
- Reconstructed images and fit residuals

Time Series Observation Products

- Integration-level data
  - Fully corrected with flagging of CRs, etc.
  - In detector frame with embedded sky transforms
- Imaging
  - Aperture photometry per integration
- Spectroscopy
  - Extracted spectrum per integration
  - White-light photometry per integration



MJD	aperture_sum	aperture_sum_err	annulus_sum	...	aperture_bkg_err	net_aperture_sum	net_aperture_sum_err
				...			
float64	float64	float64	float64	...	float64	float64	float64
57399.5792803	2646453.38093	724066.772192	238079.605507	...	942748.625804	2188607.98573	1188716.81323
57399.5795394	2649678.59194	724066.772192	238682.680756	...	942748.625804	2190673.43664	1188716.81323
57399.5797984	2663962.29593	724066.772192	245752.006275	...	942748.625804	2191362.28386	1188716.81323
57399.5800574	2650620.26893	724066.772192	238565.036121	...	942748.625804	2191841.35331	1188716.81323
57399.5803165	2648207.11774	724066.772192	239399.326949	...	942748.625804	2187823.79668	1188716.81323
57399.5805755	2648483.38487	724066.772192	240313.886387	...	942748.625804	2186341.29567	1188716.81323
57399.5808345	2657080.97989	724066.772192	237544.986117	...	942748.625804	2200263.6989	1188716.81323
57399.5810935	2651862.76529	724066.772192	240121.080814	...	942748.625804	2190091.45604	1188716.81323
57399.5813526	2652115.51585	724066.772192	241668.468849	...	942748.625804	2187368.46037	1188716.81323
...	...	...	... ...		...	...	...

Pipeline Availability

- Written in python
- Makes use, where feasible, of existing astropy libraries
- Freely available
- Configurable (highly modular)
- Users can rerun all or part of the pipeline(s)
- Users can replace specific modules
- Code repository hosted on github
- Installable using “conda”
  - Will be available via STScI “astroconda” channel
- First public release early- to mid-2018
  - Will include high-level and detailed code documentation

A (light) Taste of Details

- Each calibration step implemented as a python class
- Pipelines are lists of steps
- Callable from command line or from python shell/notebook
- Uses software data models for science and calibration reference products
  - All actual file I/O done by the data model, not the user
  - Keeps details of on-disk file formats at arms length from user/developer
    - Can change file format without having to update calibration step code
  - Defined by a model schema that allows for validation

Status and Further Reading

- DMS Build 7.1 delivery to I&T mid-Nov 2017
- DMS Build 7.2 delivery to I&T 1-Apr-2018
  - Last pre-flight build for use in the SOC
  - Will contain all “baseline” calibrations specified by JCCWG
- Documentation
  - High level
    - http://jwst-docs.stsci.edu
    - Click on “JWST Data Reduction Pipeline”
  - Lower level (details of individual pipelines and steps)
    - http://ssb.stsci.edu/doc/jwst/