

Hit the Ground Running: Data Management for JWST

Anastasia Alexov,¹ Mark Kyrianou² and Catherine Kaleida³

¹*Space Telescope Science Institute, Institution Baltimore, MD, USA;*
alexov@stsci.edu

²*Space Telescope Science Institute, Institution Baltimore, MD, USA;*
kyp@stsci.edu

³*Space Telescope Science Institute, Institution Baltimore, MD, USA;*
ckaleida@stsci.edu

Abstract.

As the launch of James Webb Space Telescope (JWST) approaches a team of engineers and scientist is hard at work developing the Data Management Subsystem (DMS) for JWST with its cadre of complex imaging and spectral instruments. DMS will perform receipt of science and engineering telemetry data; will perform reformatting, quality checking, calibration, data processing; will archive the data; will have tools for retrieving the data; will have the capacities for reprocessing the data; will have external/public calibration tools; will provide user notification, search, and access tools for JWST science and engineering data; will distribute data to the end user; provide extensive user analysis/visualization tools; and, will provide support for contributed data products from the community. We will give an overview of the software components, the hardware they run on, the programming languages/systems used, the complexity of the tested end to end science data flow, the current functionality of the system and what's to come for the JWST Data Management Subsystem in preparation for launch.

1. Who Are JWST DMS?

Relationships between multiple exposures are captured in an association. An association is a means of identifying a set of exposures that belong together and may be dependent upon one another. The association concept permits exposures to be calibrated, archived, retrieved, and reprocessed as a set rather than as individual objects. For each association, DMS will generate the most combined and least combined data products. Intermediate data products will not be saved. For imaging associations, the most combined will be either a mosaic or a dither product.

2. Explain the SOC

Automated Processing of Level 3 “combined” products is REALLY DIFFICULT! This complexity was realized with HST, which, after decades of experience was implemented only in a semi-automated way via the Hubble Legacy Archive (“HLA”). For JWST we have completely automated this procedure and are including combined prod-

ucts as part of the standard pipeline processing step. The resulting associations/combined products will be made available in the Archive User Interface (AUI). The different types of Association Products which will be made available to users are enumerated below.

- Association Pools - Contain a list of all potential exposures that may form an association.
- Association Tables - Created by the Association Generator from Association Pool information; each table lists exposures to be processed into Level 3 Data Products.
- Level 3 Data Products - Formed in the calibration pipeline from exposures listed in an Association Table. All exposures, even single, unassociated exposures, will have a Level 3 data product. All Level 3 data will get a source/target based file name.

3. End to End Data Flow

There are three main types of Associations. Only the first will be made part of automated processing in JWST DMS. The Level 4 type of Associations may be produced by Archive operators and work will begin on an overall automated design after launch.

- **Observation or Entire Proposal Associations (Level 3)** are exposures executed in the same observation, or in a set of linked observations. Examples of these are: dithers, mosaics, contemporaneous calibration exposures and on-board target acquisitions. Related observations in the same proposal, such as an image used to determine where to extract spectra, may also be included.
- **Multiple Proposal Associations (Level 4)** are mosaics taken of the same region of the sky in different, unrelated proposals with the same instruments and filters.
- **Open-ended Associations (Level 4)** are not bound by an observation or by a single proposal. These are independent images with spatial overlap, taken at very different times with different scientific intent. An example is an HLA-like ‘super-mosaic’ image of a region of the sky. These types will evolve as more data are added to the JWST Data Archive.

4. Operations

Figure ?? shows the overall data flow from the Proposal Planning System to receiving the data from the spacecraft, in order to perform full automated processing, ending in Associations. Each subsection number below matches the number scheme encircled in the figure. Additional information regarding JWST Science Data Processing can be found in (?), and on JWST Calibration pipelines in (?).

4.1. Lessons Learned

JWST proposals capture the Association “intent”; the astronomer decides how to group their observations (dithers, mosaics, background source for spectra, etc.). Proposals are sent to the Proposal Planning System Database (PPS DB). DMS replicates PPS DB

which contains Associations Candidates based on rules derived from the astronomers' intent. The design of the rules have been coordinated between the Proposal Planning system and DMS.

4.2. TBD

The Flight Operations Subsystem (FOS) creates Observatory Status Files (OSFs) denoting what happened onboard. DMS ingests the OSFs into a Metrics Database (DB). Science data are sent from Flight Operations Subsystem (FOS) to DMS, where automated science data processing starts. OSF and PPS DB information is then used to populate FITS header keywords in science data processing.

4.3. Another section

Automated science data processing takes place for every exposure. FITS files are created with valid headers. Products are ingested into the archive and made available through the Archive User Interface (AUI).

4.4. Initial Steps in Organizing Associations

The Association Pool Maker creates an Association Pool file for every proposal. The Association Generator is run when all exposures are processed through Level 1b for: Association Candidates and observations and for entire proposals. The Association Generator creates Association Tables; these tables contain: a ruleset for an Association, a list of input file names, and an output science file name/product.

4.5. Creating Associations

Science calibration pipelines are run using the Association Generator tables to create association products. Automatically combined Association products are generated using combined science products and auto-calibrated exposures, which comes with many challenges. Association products are then ingested into the archive and are finally searchable in the AUI.

5. Summary

Automated combined JWST Association products are generated using calibrated exposures. Defining what is in an Association and how to combine them without human intervention has proven very challenging. End to end data processing of Associations encompasses 14 different JWST teams, using a variety of software tools and languages to build Associations: Python 2.7 & 3.X, Java, HTCondor, ASP, Javascript, AstroPy, AstroConda, Shibboleth, MS SQL Server, SQLite, Travis, Jenkins, Jira, Confluence. A team of 15 STScI staff (developers, systems engineers, scientist, managers) worked for over 4 years on the Association design and basic implementation. JWST automated Associations will help the science users better understand their data more quickly, leading to (hopefully) faster publications. The science-oriented nomenclature of Association products will better the user experience.

6. References

This template has no bibtex file. Look for the larger template and Makefile how to do this. By default the Makefile will create an empty O2-1.bib. When you add references to this, uncomment the line `\bibliography` below, make use “make” to create your beautifully looking PDF.

Acknowledgments. I would like to thank the Association Working Group. These 15 people met weekly for over 4 years to come up with a basic design for Associations. This design has gone through a number of de-scope changes and has been approved by the JWST instrument teams.