



STScI | SPACE TELESCOPE
SCIENCE INSTITUTE

EXPANDING THE FRONTIERS OF SPACE ASTRONOMY

JWST Data Status and New Search Option

Faith Abney and Scott W. Fleming
(on behalf of the DMS Team)

JSTUC Meeting, 1 Mar. 2023



Discussion Topics

- Plans To Retire Shippable Media Requests – 1 minute
- JWST Reprocessing Statistics – 10 minutes
- JWST Data Status (Notifications) – 10 minutes
- Demo of New JWST-Specific Search Form – 15 minutes

Acronym NOTE: “DMS”= “Data Management Subsystem”, the team spread across multiple Branches within STScI’s Data Management Division responsible for pipeline processing and data archiving.



Plans To Retire Shippable Media Requests



Plans To Retire Shippable Media Requests

- The mission still has a requirement to provide shippable media upon request. The cost of keeping this capability is not free.
 - The last request for shippable media was three years ago.
- Our plan is to start decommissioning the hardware used to support shippable media requests now.
 - If we receive a request for shippable media, we will respond to those requests in an ad hoc fashion using existing resources at STScI.
- If we continue to receive zero shippable media requests for the next two years, we will revisit that mission requirement and consider a request to remove it as a requirement for the mission.



JWST Reprocessing Statistics



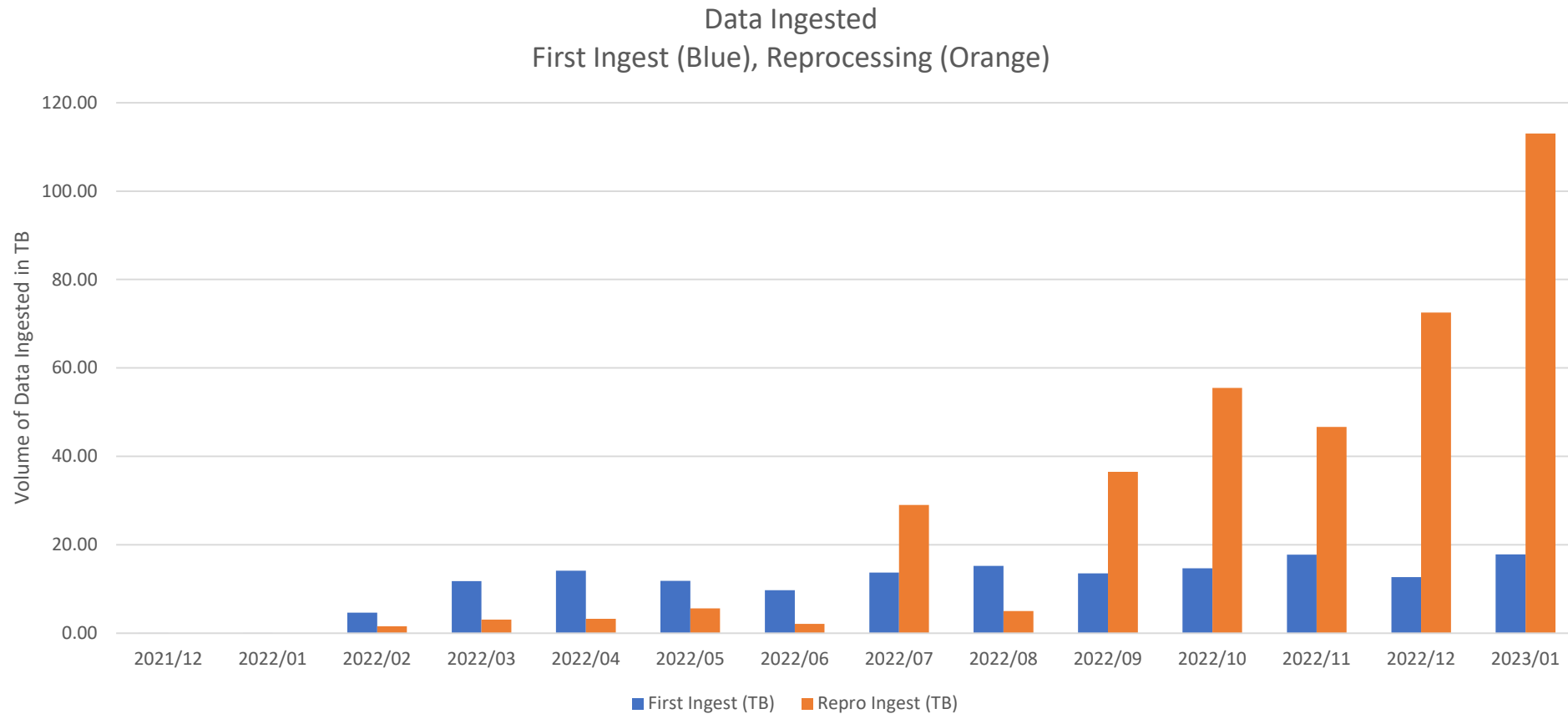
Current State of Reprocessing

- DMS updates the data in the MAST archive when new reference files are delivered or when new software is installed.
- The time to reprocess all affected data depends on the quantity of data affected.
 - Sometimes a reference file delivery affects a few files, sometimes all of the data for that instrument
- DMS works with the Instruments Division to identify the correct list of data affected and the general priority of reprocessing
 - Generally, reprocessing commissioning data is lower priority than other data
- Most all of the data in the archive was reprocessed to the best calibration and software level as of early February.
- A build was delivered yesterday (Feb 28), so most of the data in the archive is now queued up for reprocessing again.
 - Expect it take about a month to reprocess all data



Current State of Reprocessing

We are increasing the amount of data we are reprocessing

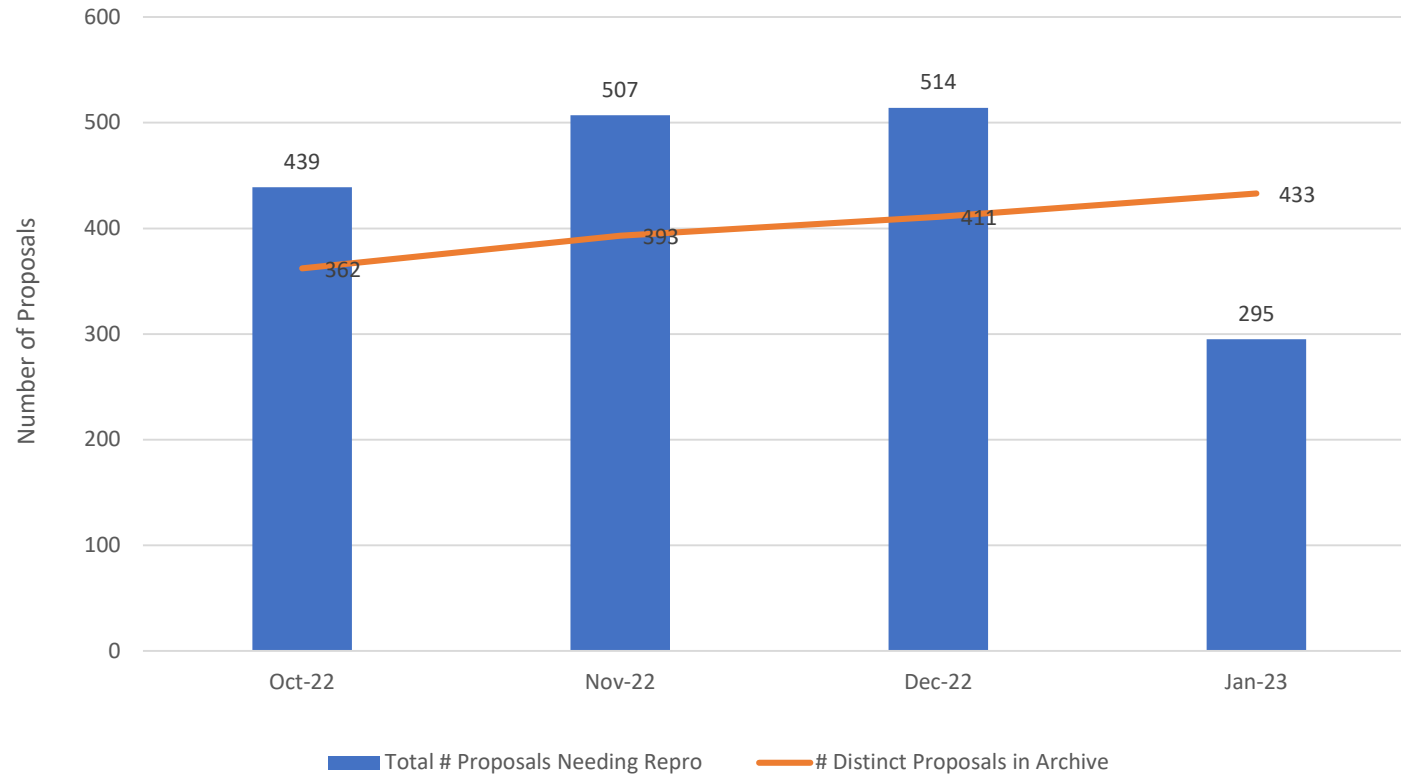




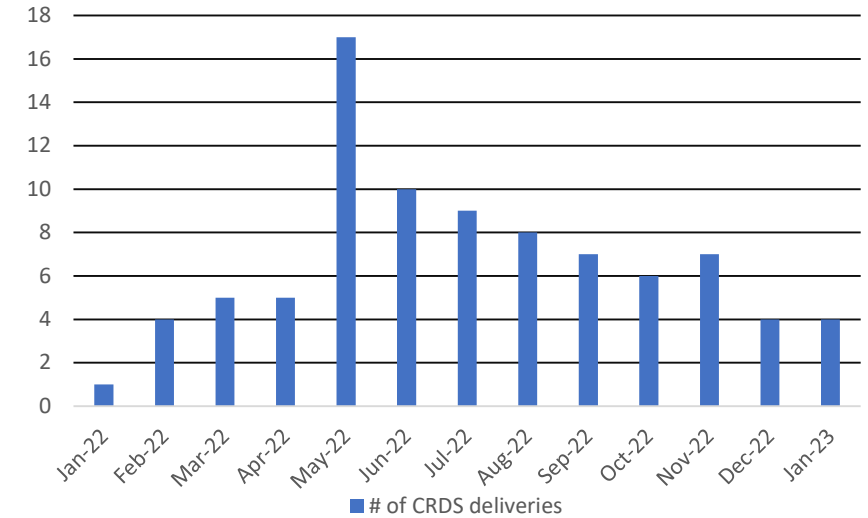
Current State of Reprocessing

We have been asked to reprocess a large amount of data each month

Number of Proposals Requesting Reprocessing



Number of CRDS Deliveries



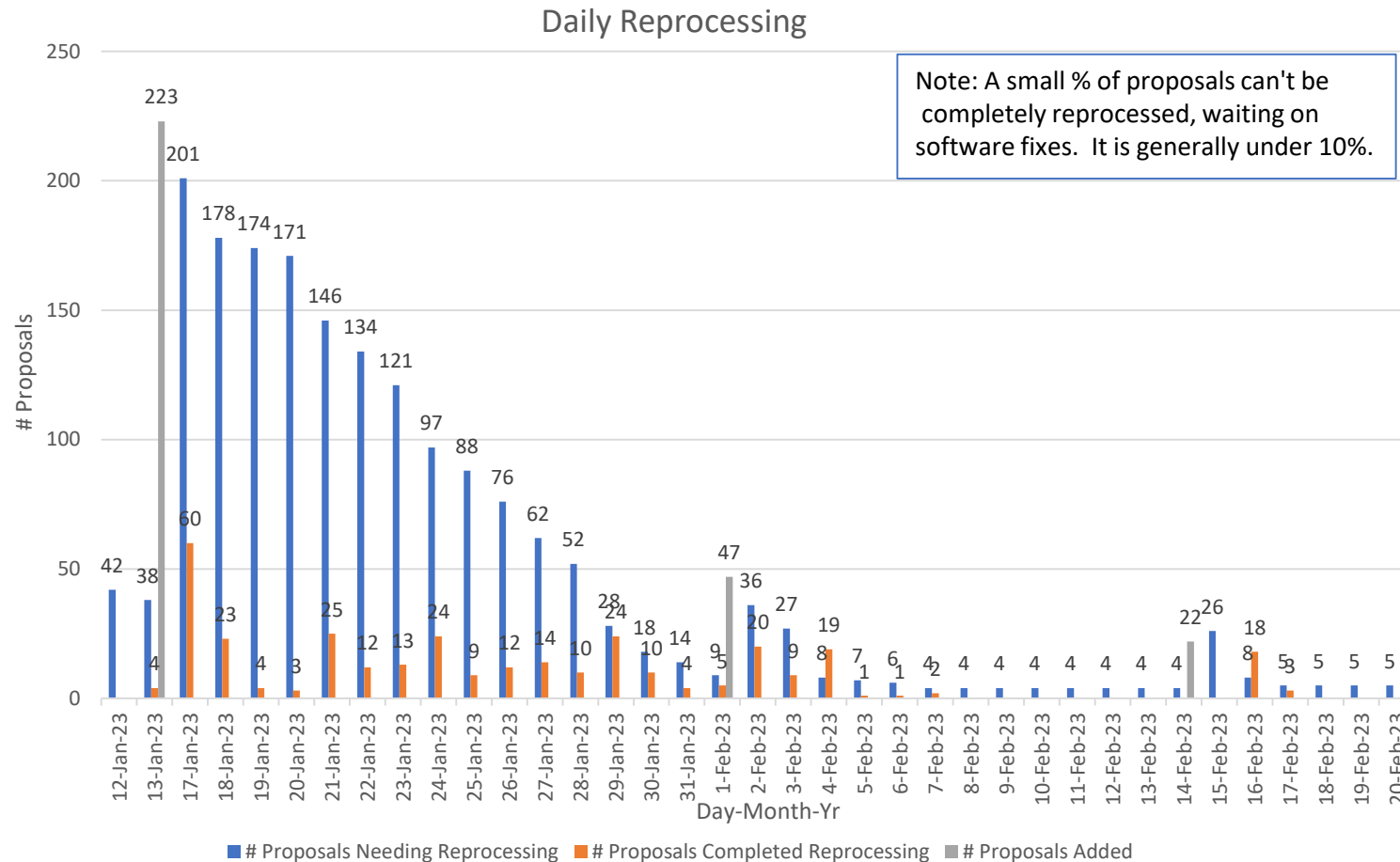
Notes:

- Oct and Nov numbers are estimated
- Proposal requests are higher than # Distinct Proposals because we are requested to reprocess some proposals multiple times within a month



Current State of Reprocessing

Starting mid-January, we are tracking the number of proposals that we reprocess each day and the number of proposals yet to be reprocessed





Current State of Reprocessing

- Even though we can get reference file deliveries at any time, we only refresh our list of what to reprocess twice a month (around the 1st and 15th), so that we can complete the high priority and work on the low priority before getting more high priority data
- There are a few proposals that can't be completely reprocessed, as they are waiting on software fixes. It is generally under 10% of the data, ~ 30 proposals.
 - The DMS build yesterday included fixes for most of these issues



Plans Forward to Increase Reprocessing Speed

- Continue to make updates to the DMS Software to improve performance
 - Have issues with reprocessing taking priority over new data processing
 - Have operational workarounds
 - Software fixes to address this
 - The current reprocessing system is very manually driven
 - Generating the list of data to be reprocessed
 - Verifying that reprocessing has completed successfully
 - Cleaning up the system so that it can handle more load
 - *All of the above are manual processes which lead to inefficiencies*
 - Better handling of lots of small files to improve performance
 - Enhancements so we can reprocess at higher levels and reprocess only the affected portions of proposals, so less reprocessing will be needed



Plans Forward to Increase Reprocessing Speed

- Continue to work with ITSD to explore how to improve system performance:
 - Adding more compute nodes
 - Grown the cluster from 10 machines to 60
 - Adding more memory and other system changes
 - Get more write throughput to storage
- Gather more metrics
 - Determine where bottlenecks are and the most time consuming processing steps so we can concentrate development on those areas
- Upgrade key parts of the system
 - HTCondor update (workflow management software)
 - Python update
 - Replace the storage (Isilon – FY24)
- Investigate cloud reprocessing



JWST Data Status (Notifications)



Warning Messages In Portal UI – Current Behavior

- Observations enter a reprocessing queue. This table records JWST observations that are in the queue and *scheduled to be reprocessed*, but are still awaiting reprocessing to begin.
- If a user attempts to download JWST observations in the Portal UI that are in this table, a message will pop up alerting them, and ask if they want to still proceed to download the data (screenshot on next slide).
- An estimate is also provided as to when they will be reprocessed (a mission requirement) however, the current estimate is not very reliable. This is difficult to estimate accurately due to varying processing times across instrument configurations and product levels (3, 2b, 2a, etc.).



Warning Messages In Portal UI - Example Not Signed In

Select a collection... and enter target: anonymous Login...

MAST Observations by Object Name or RA/Dec Enter object name or RA and Dec to cone search Search

Download Manager

Download Basket Download History

Remove Selected Remove All

Displaying 3 of 12 Total Files 226.90 MB Selected / 51200 MB Max

Download Batch Retrieval

Home 3 Total Filters Clear Filters

Recommended Products

Uncheck Minimum Recommended Products to see all files. [Learn more.](#)

Minimum Recommended Products (3 of 3)

Product Category

SCIENCE (3 of 3)

AUXILIARY (0 of 3)

Extension

fits (3 of 6)

FITS (0 of 6)

Group

UNCAL (3 of 3)

DARK (0 of 3)

Mission > Observation > File	Files	Actions	File Size	Product Type	Batch	Description	Project
JWST	3						
jw01011002001_02101_00001_mirf...	1		75.633 MB	image	Yes	exposure (L1b): Uncalibrated 4D expos...	CALJWST
jw01011002001_02101_00001_mirf...	1		75.633 MB	image	Yes	exposure (L1b): Uncalibrated 4D expos...	CALJWST
jw01011002001_02101_00001_mirf...	1		75.633 MB	image	Yes	exposure (L1b): Uncalibrated 4D expos...	CALJWST

Files May Not Be up to Date

The following fileset is being calibrated, estimated wait time is 300 minutes:

jw01011002001_02101_00001_mirimage

Would you like to proceed with your download?

OK Cancel

Details

Summary

intentType: science

Mission (obs_collection): JWST

provenance_name: CALJWST

instrument_name: MIRI

Project (project): JWST

filters: OPAQUE

wavelength_region: Infrared

target_name: UNKNOWN

target_classification:

Observation ID (obs_id): jw01011002001_02101_00001

s_ra: 0

s_dec: 0

Product Type (dataprodtype): image

proposal_pi: Ressler, Michael E.

calib_level: 2

t_min: 59679.3971255764

t_max: 59679.3980891319

t_exptime: 83.251

em_min: NaN

em_max: NaN

obs_title: MIRI Focal Plane System Cold Check-Out

t_obs_release: 59774.8541666

Proposal ID (proposal_id): 1011

proposal_type: COM

sequence_number: -999

s_region: POLYGON 122.926243081 19.567213101 122.935911729 19.597582024 122.967673943 19.588290953 122.958036523 19.558215992

jpegURL: mast:JWST/product/jw0101100

dataURL: mast:JWST/product/jw0101100



Warning Messages In Portal UI - Example Signed In

The screenshot displays the JWST Data Portal's Download Manager interface. At the top, there is a search bar with the text "Select a collection... and enter target:" and a dropdown menu set to "MAST Observations by Object Name or RA/Dec". The main area is titled "Download Manager" and shows a "Download Basket" with "6 of 41 Total Files" selected, totaling 777.37 MB. A table of files is visible, including JWST observations with various file sizes and product types. A modal dialog box titled "Files May Not Be up to Date" is overlaid on the table, displaying a warning message:

Files May Not Be up to Date

The following fileset is being calibrated, estimated wait time is 300 minutes:

jw01022012001_02102_00001_nrs2

Would you like to proceed with your download?

Buttons: OK, Notify Me..., Cancel

The interface also includes a left sidebar with filters for "Recommended Products", "Product Category", "Extension", and "Group". The right sidebar shows a "Details" section for the selected file.



Warning Messages In Portal UI - Example New Subscription

Select a collection... and enter target:

MAST Observations by Object Name or RA/Dec

Download Manager

Download Basket | Download History

Remove Selected | Remove All

Displaying 6 of 41 Total Files 777.37 MB Selected / 51200 MB Max

Filters: Clear Filters

Recommended Products

Uncheck Minimum Recommended Products to see all files. [Learn more.](#)

Minimum Recommended Products (6 of 6)

Product Category

AUXILIARY (3 of 35)

SCIENCE (3 of 3)

Extension

fits (6 of 38)

Group

GS-ID (1 of 70)

GS-ACQ2 (2 of 6)

GS-ACQ1 (2 of 6)

GS-TRACK (3 of 3)

UNCAI (1 of 1)

Mission	Observation	File	Actions	File Size	Product Type	Batch	Description	Project
JWST		6						
JWST	jw01022012001_02102_00001_nrs2	6						
JWST	jw01022012001_gs-track_202218...							
JWST	jw01022012001_gs-track_202218...							
JWST	jw01022012001_gs-track_202218...							
JWST	jw01022012001_02102_00001_nr...							
JWST	jw01022012001_02102_00001_nr...							
JWST	jw01022012001_02102_00001_nr...							

Create New Subscriptions

General Subscription Options

Notification Type:

Notification Frequency:

Subscriptions to Create

#1 Mission:

Subscription Type:

New Reprocessed

Public First

Select One or More to Include:

Proposal ID:

Observation ID:

Optional:

Data Level:

Details

Summary

Select an observation or file to view details

DEC

SPACE TELESCOPE SCIENCE INSTITUTE



Warning Messages In Portal UI – Plans Forward

- DMS will be working to ensure the impacted table is kept up-to-date and is as complete as possible.
- DMS will also work on providing a more conservative (and realistic) singular estimate of when reprocessing of data will likely finish, based on some upper bound across JWST reprocessing timescales.
- Possible future work would be to continue to improve the estimate, e.g., by providing upper-bound estimates per instrument/mode, or perhaps using sophisticated analyses of previous reprocessing metrics.



Questions on Portal Warnings Before Moving On To Subscriptions?



Subscriptions By Program ID or Dataset – Current Behavior

- Users can create or edit subscriptions based on program IDs or dataset IDs.
- There are multiple options for frequency of notifications: “fast” = instantly, or in summarized digests 1x per day, week, or month.
- There are multiple types of triggers that can be set for a dataset or proposal ID:
 - New (whenever data for a program appears for the first time)
 - First (only when the first data for a program appears)
 - Reprocessed (whenever existing data within a program get reprocessed)
 - Public (whenever data within a program switches from EAP to public).
- The combination of “fast” + “reprocessed” was a particularly potent mix for those with subscriptions to large or multi-object spectral programs, leading to far too many notifications being sent out.



Subscriptions by Program ID or Dataset – Existing Message Info

- JWST notification message event structure: {"mission", "propid", "obsets", "level", "type"}
 - "mission" = "JWST"
 - "propid" = Program ID.
 - "obsets" = The 3-digit observation number, can be a list of more than one.
 - "level" = The product level the message is for.
 - Events are only generated for science files at Level 2b or above.
 - "type" = "new", "repro", "public", "first"
- Example message:

```
[{"obsets":["001","002"],"mission":"JWST","propid":"96090","type":"public","level":"2b"}]
```

 - This message event informs that data from JWST Observation Number 1 and 2 under program 96090 have switched from EAP to Public.



Subscriptions By Program ID or Dataset – Recent Improvements

- As of Jan. 31st, the default frequency for new subscriptions is “daily”, no longer “fast”. For most users, this is a far better default.
- As of Feb. 28th, the email notifications that get sent contain a link at the bottom that points to instructions of how to edit existing subscriptions to change the delivery frequency.



Subscriptions By Program ID or Dataset – Plans Forward

- Short-term: DMS is investigating other small-scale changes to improve the notices that get sent out.
 - What worked before for HST and the original JWST reprocessing schedule, is not optimal for the reprocessing realities of JWST now.
- Medium-term: DMS is starting to work on a technology refresh + redesign of the subscription service itself, which may include additional information and descriptive content in the emails. Examples may include:
 - additional information on what changed with the calibration updates
 - which specific files were impacted within the observation(s)
 - easier access to download the impacted files from within the notifications
 - integration of subscriptions with the new JWST-specific search form



Demo Of A New JWST-Specific Search Form



Modern, Accessible Web Search Forms

- New HST form has been in operations for ~2 years. The older HST form permanently redirects to new HST form as of Nov. 1st 2022.
- JWST search form is in development (demo upcoming). Includes ability to search for observations using JWST-specific metadata (header info) and easier ways to select products (by Level, by file extension, file name (patterns), etc.)
- Important to note that all *existing* JWST search and download methods will remain unchanged and supported while we continue to build out the new search form's features and capabilities. Examples: Portal Cross-Mission Search, Instrument Keyword Search, [astroquery.mast](https://astroquery.mast.umd.edu/).



Modern, Accessible Web Search Forms – JWST Demo

Demo Begins