

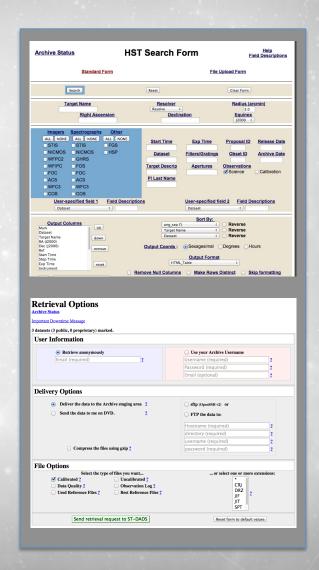
The HST Online Cache: Changes to the User Experience

Karen Levay
Faith Abney
Mark Kyprianou



Current Process for HST data

- User submits request for data
 - On the Fly Reprocessing for active instruments
 - "Static" archive for legacy instruments
- All requests submitted are batch requests via "DADS"
 - Requests for legacy instrument data is quick as the request is just copying data to stage or to user's platform
 - All other data is completely reprocessed even if only flt files requested and delivered. All local copies of the processed data are deleted.





Future HST Online Cache

- All data will be stored online.
- Public data will be url accessible.
- Proprietary data will be protected. Initially not available via browser/ftp or through the portal.
 - Later maybe next year, as the archive becomes comfortable with SSO and data protections, the proprietary data may become available through the portal and via proxy with proper authentication.
- Data for active instruments will be reprocessed as new calibrations / software are introduced.



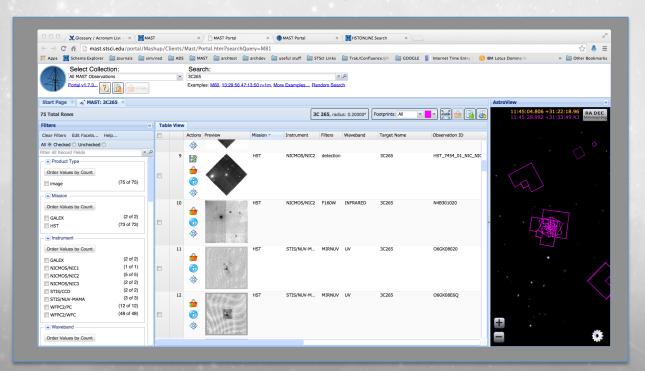
Classic HST search and DADS retrieval behavior remain the same

- Users will still be able to use the familiar HST search and batch retrieval interfaces, but should have faster response time for active instruments, as the data will be copied from the online-cache.
- The batch request will appear to be the same but retrieve all data from the "online cache" and deliver to stage or to the users platform as designated. Hard media (DVDs, CDs) will remain an option for HST retrievals.



... but there will be new options

 Data that are in the HST public online cache will be directly available via the MAST Discovery Portal and the download basket as the Common Archive Observation Model (CAOM) database is updated.



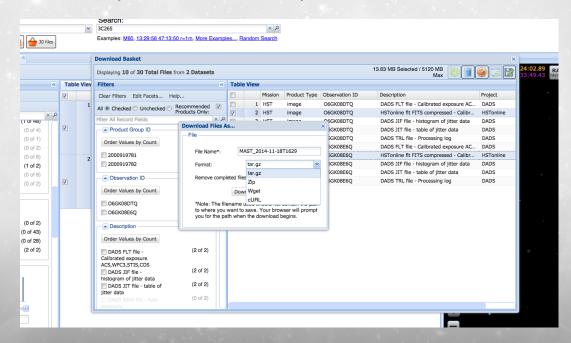


MAST Users Group Meeting

> Dec 2, 2014

... but there will be new options

- Download basket options include:
 - Downloading a tar or zip files of data through the browser (must be < 5 GB)
 - Creation of cURL and wget scripts so users can download to platform of choice in the background
- If a user knows the specific dataset of interest, the data are easy to find in the public online cache. Users will be able to easily build programmatic scripts to download public data from the cache.





Over the next 2-3 years some new features may be introduced.

- Some of these features are primarily being developed for future JWST use and the HST/MAST data and interfaces will be used as a test bed. This is the reason for the longer development time-line.
- Logged-in users may be able to use a set of "subscription" services, e.g. notification when data become public, notification when data are reprocessed.....



Over the next 2-3 years some new features will be introduced.

- Users will be told when datasets of interest are in the queue for reprocessing.
- Users who are logged in and authorized to view specific proprietary datasets, will be able to see previews of proprietary data and to request them through the download basket



Timeline (1/3)

- Dec 2014
 - Population of the online cache begins with static, legacy instrument data
 - CAOM updates required for access to the online cache through the MAST Discovery portal
 - Implementation of the Owl/Condor workflow replacing OPUS pipelines for several functions



Timeline (2/3)

- Jan May 2015
 - Processing to fill cache for active instruments begins.
 - CAOM population and preview population done via operational tools.
 - As CAOM is updated, the newly processed data will become available through the MAST discovery portal.
 - New previews will be created. HLA preview software being used for image data. New spectral plotting software for spectral data.
 - MAST discovery portal search options expand beyond spatial searches.
 - OTFR remains in place for data not yet in the onlinecache.



Timeline (3/3)

March 2015

 STScI Single Sign-on (SSO) implemented for archive users. Initially will affect the standard retrievals. Planning very extensive public announcement and mailings about how the changes will affect users.

• June 2015

- Replacing the OPUS "pre-archive" pipelines with Condor/Owl workflows.
- CAOM updates and preview creation part of the Data Processing workflow.