

Public AWS Data

On this page...

- [AWS Open Data Program](#)
- [Current MAST AWS Datasets](#)
 - [Completed Missions](#)
 - [Active Missions](#)
- [Cloud Data Access via Astroquery](#)
- [Bringing New Missions Online](#)

AWS Open Data Program

The Amazon Web Services (AWS) Open Data Program currently hosts hundreds of datasets on various topics, ranging from astronomy to biology. To sort through these datasets, you can [search for astronomy on the registry of open data](#). In addition to data from MAST, you'll find data from other NASA missions and a few ground-based telescopes.

Current MAST AWS Datasets

First and foremost, an important note: the AWS Public Datasets are not a replacement of the MAST Archive. Data are, *and will always be*, available free of charge from MAST. By distributing this copy of data on AWS, we're exploring a new kind of archive service: one where the data are **highly available** through bulk, high-speed access in proximity to the vast computational resources of AWS. This enables us to create new services, like [Cloud Science Platforms](#), and gives users a chance to access "big data" far more quickly and reliably than can be done through conventional downloads.

See our [STScI Homepage on AWS](#) for the most up-to-date information about which datasets are available in the cloud. A small subset of that information is reproduced here for convenience.

Completed Missions

Completed missions are no longer observing or producing data. On AWS, these are marked with an update frequency of "Never." The AWS and MAST servers contain identical sets of data.

- GALEX
- Kepler and K2

Active Missions

Active Missions are still observing, with data regularly ingested to MAST. There is a delay between when the data arrives at MAST and when it becomes available on AWS. This depends on the mission in question; see the table below for specifics on the update frequency. Only public data is uploaded to the cloud; PIs looking for their proprietary data must use MAST.

Mission	AWS Update Frequency
Hubble	Hourly
TESS	Monthly

Cloud Data Access via Astroquery

Outside of the STScI-developed platforms, you still have access to the data on the cloud. The [cloud access features of astroquery.mast](#) allow you to pull data from the cloud to your local machine. Note that this is not the same as working on the cloud; instead, downloads will originate from the cloud when possible. If your data is not from one of the missions listed above, then it will be sourced from the original, on-premises MAST server.

one-line cloud access

```
from astroquery.mast import Observations

Observations.enable_cloud_dataset()
```

Bringing New Missions Online

We plan to continue adding new missions to the cloud. A particular focus is on frequently accessed missions with large data volumes. As the cloud collection grows, more wavelength domains will become available, enabling diverse and varied research and creating the potential for multi-mission research.

To those curious about what it takes to migrate missions from the MAST Servers in Baltimore to Amazon's in Virginia: the best way to transport large quantities of data is still via mail. For Hubble, we transferred 110 TB of data using the AWS [Snowball](#) service. We received two 80 TB banks of hard drives, rsync-ed the data, then mailed them back. After the initial set-up, we modified our pipelines so that internal changes (i.e., new or reprocessed data) are reflected in AWS within 10-20 minutes.