# **Automated Detection of Hazards Challenge**

# Phenomena Detection Challenge Resources

In this section, you will find resources for the phenomena detection challenge, including satellite imagery, labeled data, and labeling tools.

If you visit our [GitHub](https://github.com/nasa/spaceapps-phenomena_detection) you can find all this information, as well as additional resources such as example code for imagery downloading, processing, and machine learning.

# Pre-Labeled Datasets

The following is a list of example phenomena for which we have provided labeled datasets. You may choose to use any of the provided phenomena for your detection model.

## Transverse Cirrus Bands (TCB)

TCBs are bands of clouds oriented perpendicular to the atmospheric flow in which they are embedded. TCBs are often an indicator of strong turbulence and often **associated with severe weather such as hurricanes and thunderstorms or atmospheric jets.**

* Transverse Cirrus Bands Data: s3://impact-datashare/transverse\_bands/

## Cloud Streets

Cloud streets are extended lines of cumulus clouds oriented parallel to the atmospheric flow. Cloud streets are known to be **significant to the vertical transport of heat, moisture, air pollutants, and momentum within the atmosphere.**

* Cloud Streets Data: s3://impact-datashare/cloudstreet

## High Latitude Dust

High Latitude Dust (HLD) consists of a mixture of solid and liquid particles suspended in the atmosphere varying in composition, source and size. **Dust storms and transport can have extensive climate and societal impacts.**

* HLD Data: s3://impact-datashare/hld-labeled

## Biomass Burning Smoke

Smoke is the mixture of gas and particulates in the atmosphere resulting from the combustion of biomass. **Smoke from events such as wildfires and agricultural burning may be transported thousands of kilometers and causes numerous health impacts.** The training and validation folders can be found using the link provided. The files are in the .tif and .bmp formats.

* Smoke Data: s3://impact-datashare/smoke-labeled

## Air Quality Data

Monitoring surface air quality helps ensure the protection of human health and property. However, surface air quality data is sparsely monitored. To help fill in these gaps, inferences are derived from other sources including remote sensing.

* Air Quality Data: s3://impact-datashare/pm2.5-labeled

The CSV formatted data and labels are provided below. It contains the following fields which have been described.

1. station\_id: Unique identifier of the PM 2.5 monitors stationed across US
2. stime: Time and date of sample recorded
3. air\_data\_value: EPA air data PM2.5 readings
4. RH: relative humidity from HRRR
5. UGRD, VGRD: Wind speed vectors from HRRR
6. HPBL: Height of Planetary Boundary Layer from HRRR
7. TMP: Temperature recorded from HRRR
8. goes\_measurement: AOD reading from GOES R

## How to download data from aws s3 buckets

1. Download and install AWS Command Line Interface (CLI). More information can be found [here](https://docs.aws.amazon.com/cli/latest/userguide/install-cliv2.html).
2. use aws s3 cp <s3://link/to/folder/> <path/to/save> --recursive --no-sign-request

# Unlabeled Data

If you decide to use raw data, you will need to first choose and download your data, and then you will need to label it. Here we have provided links to data sources as well as instructions for using the tool ImageLabeler for labeling your data.

## NASA CMR

NASA's Common Metadata Repository (CMR) is a metadata system that catalogs all data and service metadata records for NASA's Earth Observing System Data and Information System (EOSDIS). You can obtain any NASA owned, publicly available dataset through CMR or its API.

* [NASA CMR Website](https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr)
* [CMR API Endpoint](https://cmr.earthdata.nasa.gov/search)
* [NASA CMR API Documentation](https://cmr.earthdata.nasa.gov/search/site/docs/search/api.html)
* [Example Notebook](https://github.com/nasa/spaceapps-phenomena_detection/blob/dev/data/raw/cmr_search_example.ipynb)

## Earthdata Search

Earthdata Search is the user-friendly frontend to CMR. Using this service, you can find satellite datasets and filter them by parameters such as science keywords, dates, platforms and instruments, data format, and location.

* [Earthdata Search](https://search.earthdata.nasa.gov/search)
* [Getting Started Guide](https://earthdata.nasa.gov/learn/getting-started)

## GIBS Worldview

Global Imagery Browse Services (GIBS) provides over 900 datasets of globally available NASA satellite imagery in full resolution. The satellite imagery can be rendered in your own web client or GIS application. Users can also get location and time information of some Earth science events such as wildfires, volcanoes, and algal blooms. The links below provide the extended description and a sample script to download GIBS imagery.

* [Worldview Data Viewer](https://worldview.earthdata.nasa.gov/)
* [GIBS Satellite Image Search](https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/gibs)
* [MODIS Image Downloader](https://github.com/NASA-IMPACT/data_share/blob/master/examples/url_generator.ipynb)

## GOES Satellite Data

GOES satellites 16 and 17 provide atmospheric and Earth surface imagery across the Western Hemisphere every 10 minutes. Because they stay above a fixed spot on the surface, they are able to provide consistent data of the same location over time.

These files are stored in compressed NETCDF formats and can be accessed through [AWS S3 buckets](https://registry.opendata.aws/noaa-goes/). The raw files need to be processed for obtaining RGB images.

* [GOES 16/17 data on AWS](https://registry.opendata.aws/noaa-goes/)
* [GOES Documentation](https://docs.opendata.aws/noaa-goes16/cics-readme.html)

# Image Labeler Documentation and Website

The Image Labeler is an interactive application created by the development team at NASA IMPACT designed to help users generate tagged images for machine learning. Images can be uploaded directly, or users can extract images from an interactive satellite map. Labeled images can also be downloaded for training your detection model.

[Image Labeler Tool](https://labeler.nasa-impact.net/) [Image Labeler Documentation](https://nasa-impact.github.io/image_labeler_docs/html/index.html)

## Image Labeler's Data

All image data for a given event inside the Image Labeler application falls into two main classifications; *labeled* and *unlabeled*. *Unlabeled* data has been organized into an Earth Science Event category and any particular image may or may not actually show the event. This data is also referred to as *untagged*.

*Labeled* (or *tagged*) data has been manually reviewed and further refined with a status of *present* or *not present*, indicating whether the phenomenon in question is shown or not in the given image. This data is ready to be downloaded and used for analysis and training.

For this challenge, *labeled* data is available for immediate download (see below for directions). Alternatively, there is an option to upload, select, and tag your own images to create a custom dataset.

## Labeling Images

To begin labeling, you will first need to upload your own set of images. Documentation on this topic can be [found here.](https://nasa-impact.github.io/image_labeler_docs/html/sectionfour.html) The linked page explains how to populate an event by either uploading images or extracting them from satellite data.

Once you're satisfied with the *unlabeled* images contained inside an event, you can start assigning labels. A walkthrough of how to classify images as *present* or *not present* is [shown here.](https://nasa-impact.github.io/image_labeler_docs/html/sectionfive.html)

## Downloading Labeled Images

Now that you have chosen a set of *labeled* images, it is time to download this data from Image Labeler. There are a couple ways to download images depending on what exactly you are looking for.

When browsing the Earth Science Events list page, two download options are given. First is the Bulk Download button in the top right. Clicking this button will give the option to download all images(both *labeled* and *unlabeled*) for any selected events. The use of *Shift+Click* or *Ctrl+Click* allows multiple events to be downloaded at once.

A download button is also given for each event in the list. When clicked, a pop-up menu will prompt for which combination of *unlabeled*, *present*, or *not present* images you wish to download.

Lastly, when inside an event page, individual images can be selected and downloaded. This is useful for downloading a subset of images already classified into one of the aforementioned categories. Directions for performing this download can be [found here.](https://nasa-impact.github.io/image_labeler_docs/html/sectionsix.html#images)

# Contacts

Please contact the following sponsors for additional information or assistance with the challenge and the resources made available through this repository. You can contact us using [rocketchat](https://chat.spaceappschallenge.org/home).

Carson Davis

* Primary Contact
* Computer Science, Machine Learning
* rocketchat: CarsonDavis

Aaron Kaulfus

* Primary Contact
* Atmosphere and Earth Science
* rocketchat: ask08

Kumar Ramasubramanian

* Machine Learning, Computer Science
* rocketchat: mr0051

Ankur Shah

* Atmosphere and Earth Science, Machine Learning
* rocketchat: ankshah13