

# Data Catalog

## Overview

The root URL for the HPS data catalog at JLAB is

<http://hpsweb.jlab.org/datacat>

The primary web interface can be found at

<http://hpsweb.jlab.org/datacat/display/browser>

The REST API can be accessed by appending 'r' to the root URL as in

<http://hpsweb.jlab.org/datacat/r>

These URLs will load over SSL/HTTPS if you [install the JLAB root certificate](#).

## Terminology

Here are some terms to be familiar with in data catalog lexicon.

### Dataset

A **dataset** is the data catalog's model of a resource, which is usually a file on disk. In the case of HPS, this could be individual EVIO, LCIO or ROOT files.

### Folder

A folder represents a logical container in the data catalog. Folders may contain datasets or other folders as children, similar to a file system.

The **root folder** is the container of all data HPS.

It may come as some surprise that this root folder is set to */HPS*.

### Path

As in file system terminology, the **path** will generally be the folder and the name of the dataset which denotes its unique identity in the data catalog.

All paths referring to HPS data should start with */HPS* as this is the root folder of the data catalog.

### Name

The **name** is the path without the folder of a dataset.

### Resource

The resource is the dataset's physical location on disk, which should be a valid file system path.

HPS uses MSS (tape storage) paths starting with */mss/hallb/hps* even though the file are actually accessed from the cache disk.

### Site

A dataset may have multiple **sites** representing different replicas of a physical file.

Only certain values for site will be accepted by the datacat API. For HPS, these include "JLAB", "SLAC" or "TEST".

### Format

When referring to a dataset, the **format** denotes the file format of the resource (physical file on disk).

This is constrained to a limited set of values for the HPS data catalog: "EVIO", "LCIO", "ROOT", "AIDA" or "TEST".

### Type

When referring to a dataset, the **type** denotes the logical contents of its resource.

For HPS, the following values are valid for the type: "RAW", "RECON", "DST", "DQM" or "TEST".

**RAW** means raw data from the DAQ (EVIO format).

**RECON** is physics reconstruction data in LCIO.

**DST** is a ROOT based representation of the physics data.

**DQM** are data quality plots in either AIDA or ROOT format.

## More Information

[Data Catalog](#) [Python Client](#)