

HAPI Overview

HAPI = Heliophysics Application Programmer's Interface

Jon Vandegriff for the HAPI Specification Team

Robert Weigel (George Mason Univ.), Jeremy Faden (Cottage Systems), Sandy Antunes (JHU/APL), Bernard Harris (GSFC), Robert Candey (GSFC), Aaron Roberts (GSFC), Eric Grimes (UCLA)

What is HAPI?

Why should I care?

How can I ...

use HAPI data?

provide HAPI data?

The future

What is HAPI?

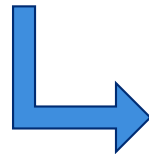
Prelude: What is timeseries data?

Time	data1	scalar2	array	multiDimArray
t0	d0	s0	a0[11]	m0[3,8]
t1	d1	s1	a1[11]	m1[3,8]
t1	d2	s2	a2[11]	m2[3,8]
t2	d3	s3	a3[11]	m3[3,8]
t4	d4	s4	a4[11]	m4[3,8]
t5	d5	s5	a5[11]	m5[3,8]
t6	d6	s6	a6[11]	m6[3,8]

- conceptually, it is a table, like a spreadsheet
- time column followed by data columns (variables)
- each variable can be multidimensional

What does HAPI enable?

It greatly simplifies the **data access problem** for time series data.

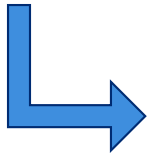


Where is the data?
What specific files do I need for my study?
What format?
What arrangement within that format?
How is time represented in the files?

No science in
these answers!

Time consuming
to find out!

HAPI offers **interoperability**.

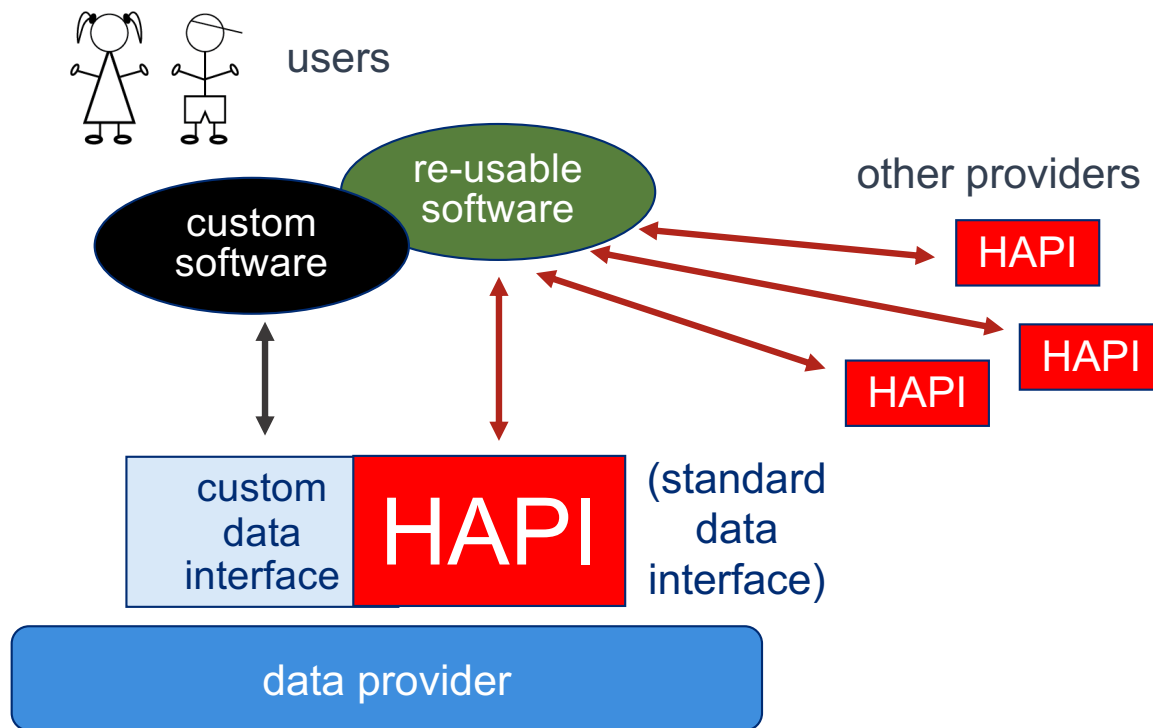


The same, exact access mechanism (same code) for reading any time series data.
No longer need to care about formats, files, and data arrangements.
The data you want, shows up inside your own application in data structures that you recognize.

Impedance matching of diverse data
sources to scientists workflows.

How does HAPI simplify data access?

Establish a common interface that ALL providers of time series data can implement.



Lowest Common Denominator

HAPI requests are very, very simple:

dataset name
start time
end time
(optional) list of parameters

HAPI is a *specification* (language independent).

Requests and responses are well defined and simple.

- ➔ *Implication of simple request format: final data (no dataset specific options)*
- ➔ *No access restriction mechanism (this is outside HAPI; discouraged by NASA)*

HAPI **servers implement the specification and provide data.**

Generic servers available in Python, Java, and JavaScript.

HAPI **clients understand the specification and grab data.**

Clients are available in many languages (Python, IDL, Java, Matlab, R)

Reference page: Where to find more info

- <http://hapi-server.org>
 - one-page summary, video intro, links to Github repos, mailing lists, etc
- R. Weigel et al, HAPI: An API Standard for Accessing Heliophysics Time Series Data
 - JGR Space Physics: <https://doi.org/10.1029/2021JA029534>



- **The HAPI Specification**
 - Version 3.0 <https://doi.org/10.5281/zenodo.4757597>
 - The development location for the latest spec and drafts and open issues, etc: <https://github.com/hapi-server/data-specification>
- Lots of other HAPI projects:
 - <https://github.com/hapi-server>



- **Try it out with this interactive way to explore known HAPI servers!**
 - <http://hapi-server.org/servers/>
- Tutorial Notebooks for using HAPI in Python:
 - <https://github.com/heliophysicsPy/summer-school/tree/main/hapi-tutorial>

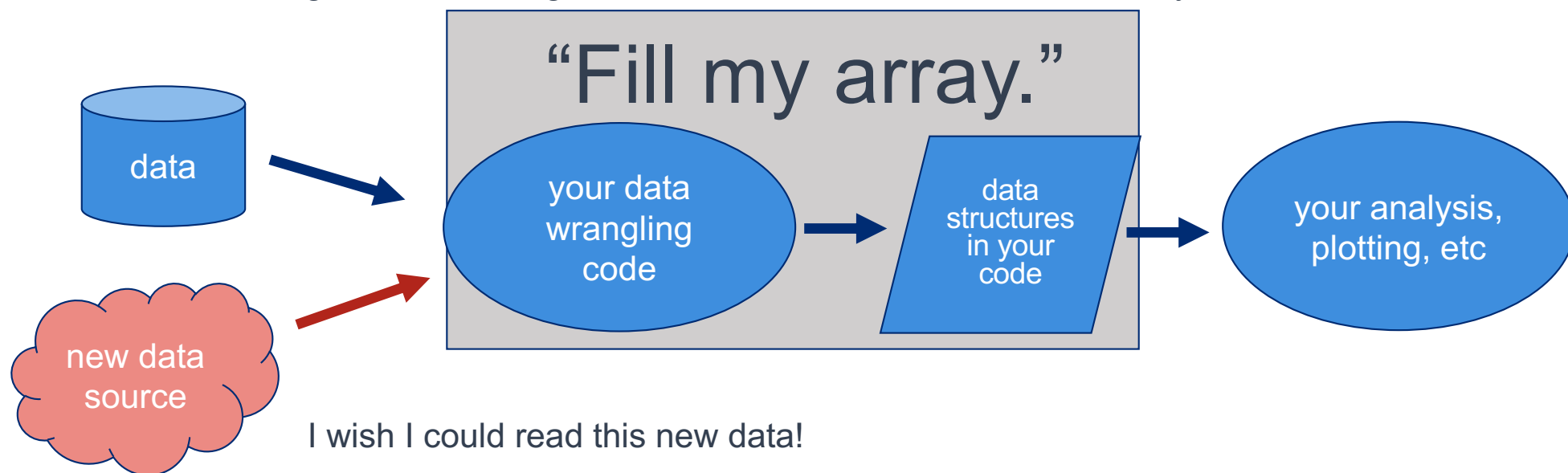
Why should I care?

Wide adoption of a common access interface could greatly simplify your workflow.

I wish every dataset I ever wanted appeared in Madrigal.



The next best thing – I wish Madrigal data access worked the same as every other data center.



Not just line plots: higher dimensional data is supported by HAPI

ESAC HAPI Server
Cluster-3 Electric Wave Form Power Density

CDAWeb HAPI Server
Van Allen Probes, RBSPICE Ion Energy Spectrogram

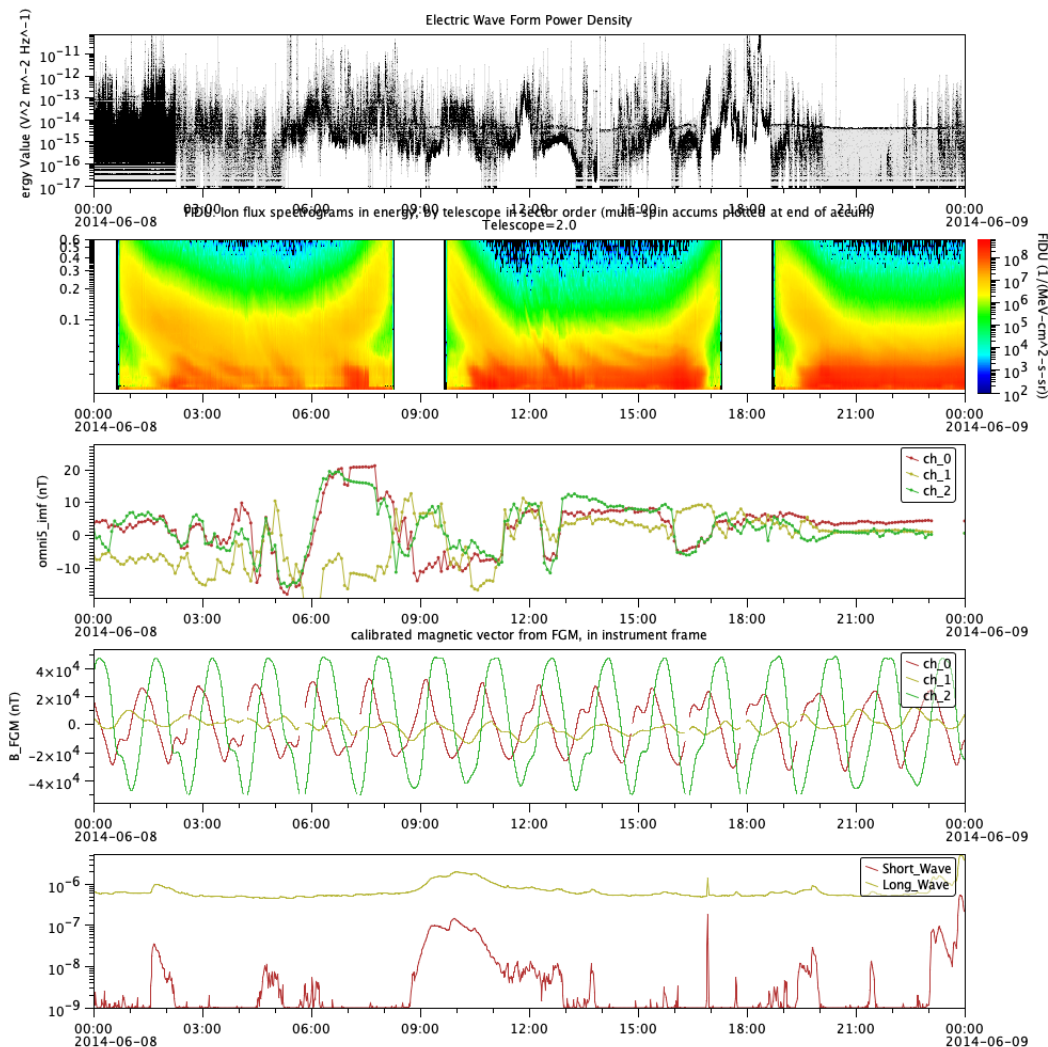
Made with Autoplot
autoplot.org
talking to 5 different servers!

AMDA HAPI Server
Omni MAG data

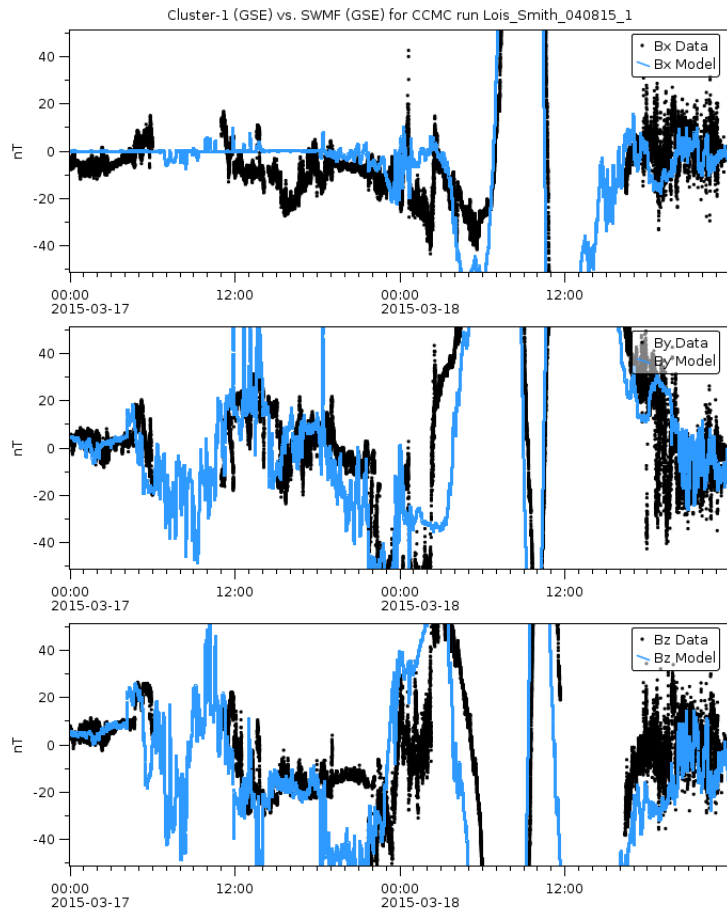
ViRES HAPI Server
GRACE-A MAG Data

CCMC ISWAT HAPI Server
GOES X-ray flux

2014-06-08



CCMC example: Data-Model Comparisons



Steps:

- Use model data from SWMF.
- Get Cluster-1 **ephemeris** from **SSCWeb HAPI server**.
- Fly the Cluster-1 trajectory through the model output to get mag data along this trajectory
- Get Cluster-1 mag **data** from **CDAWeb HAPI server**.

This example is from the CCMC web page:

https://ccmc.gsfc.nasa.gov/ungrouped/GM_IM/GM_analysis.php?Pid=12870&Pt=BO&Ps=Cluster-1

Work is by Darren De Zeeuw.

How can I (get to | provide) HAPI data?

If you are a data provider

You can expose your data to many users through a standard API with very little effort.
(Most users nowadays want a computer API to access data.)

Approach:
Read the HAPI spec

Join the hapi-dev@groups.io mailing list

Choose an existing HAPI server to put on top of your data (Python or Java or node.js)

--OR--

Adapt your current data interface to emit HAPI responses.

Test your server using the online HAPI server checking tool (verifies compliance with HAPI spec)

Issue a pull request to add your tested server to the official list of HAPI servers.

<https://github.com/hapi-server/servers>

Use the HAPI verifier to check a server

- Runs a battery of tests to see if the server meets the spec
 - hits all the endpoints
 - requests data
 - tries bogus inputs
- Run locally (recommended)
 - code is here: <https://github.com/hapi-server/verifier-nodejs>
- Web version (comprehensive but slow – lots of requests!)
 - <http://hapi-server.org/verify/>

How to access HAPI data?

1. ready-made analysis tools

SPEDAS / PySPEDAS
<http://spedas.org>

Autoplot
<http://autoplot.org>

2. write your own code using libraries that read HAPI data

Python, Matlab, IDL, R, Java

<https://github.com/hapi-server>

3. Explore what HAPI data is available

<http://hapi-server.org/servers/>

This site can generate a script stub for grabbing data:

```
from hapiclient import hapi
from hapiplot import hapiplot

server      = 'https://cdaweb.gsfc.nasa.gov/hapi'
dataset     = 'AC_K1_SWE'
parameters  = 'Vp'
start       = '2017-07-01T00:00:00Z'
stop        = '2017-07-03T00:00:00.000Z'

data, meta  = hapi(server, dataset, parameters, start, stop)

hapiplot(data, meta)
```

Python example.

Reference page highlights for HAPI clients

- Lots of other HAPI projects:
 - <https://github.com/hapi-server>
- **Try it out with this interactive way to explore known HAPI servers!**
 - <http://hapi-server.org/servers/>
- Tutorial Notebooks for using HAPI in Python:
 - <https://github.com/heliophysicsPy/summer-school/tree/main/hapi-tutorial>

If you want to do machine learning with Heliophysics data

Data wrangling is still the most time consuming part of ML!

HAPI simplifies this. You can get to many different datasets with a single API.

Approach:

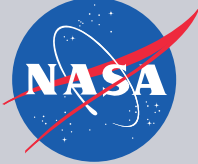

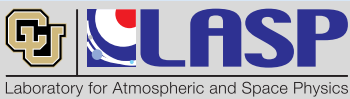
See if the data you want is available via HAPI

Look for a HAPI reader library in your favorite language.

Write one piece of code to grab data from different places and recast it into your ML-ready content.


Note: project coming to connect HAPI data to ML tools like TensorFlow, PyTorch.


What data is available now with HAPI?

Institution	Server	Type of Data	Number of Datasets
	CDAWeb	Heliophysics	2800
	SSCWeb	Ephemeris	250
	CCMC	Space Weather	250
	AMDA	Heliophysics and Planetary Data and Ephemeris	500
University of Iowa	DAs2 Server	Helio. & Planetary	30
 <small>Laboratory for Atmospheric and Space Physics</small>	LISIRD	Solar Irradiance	40
SWARM Mission	ViRES Data Server	Space Mag Data	14

Note: Both European and US buy-in.

Coming soon

Institution	Server	Type of Data	Number of Datasets
	ESAC	Heliophysics	2073
JHU / APL	SuperMAG	global ground mag	~500
JHU / APL	TIMED / GUVI	ionospheric images	~10

Eventually			
	PDS PPI Node	Planetary Plasma, Particle, and Fields	~1000

CEDAR / NSF	Madrigal	Space Weather	1000+ (??)
-------------	----------	---------------	------------

Note: Heliophysics, Planetary Science, Geospace.

Future Plans

- explore the use of HAPI for some kinds of model output
- use HAPI in cloud-based analysis (data in the cloud and analysis in the cloud)
 - working with HelioCloud effort – an emerging cloud-based analysis platform in Heliophysics.
- promulgation: help more data centers deploy HAPI
- more client tools that can read HAPI
- easier on-ramp for smaller data providers

HAPI is a COSPAR endorsed standard for Space Weather data access

<https://doi.org/10.1016/j.srt.2021.11.014>



Space Research Today
Volume 212, December 2021, Page 19



COSPAR Panel on Space Weather Resolutions on Metadata Standards and Data Access

Show more ▾

Share Cite

<https://doi.org/10.1016/j.srt.2021.11.014>

Get rights and content

COSPAR Panel on Space Weather Resolution on Metadata Standards

Accepted at COSPAR Panel on Space Weather (PSW) Business Meeting on 18 July 2018 (revised 6 December 2018; updated 15 October 2021)

Taking into account that:

1. It is in the general interest of the international space physics and space weather community that data be made as widely accessible as possible,
2. The open exchange of data benefits from accurate, complete, and standardized metadata,
3. The ILWS-COSPAR Roadmap recommends to standardize metadata and harmonize access to data and model archives,
4. There are a variety of metadata standards, few of which meet the full needs of the community to document and enable discovery of available diverse data,
5. The SPASE (Space Physics Archive Search and Extract) metadata standard has demonstrated that it is comprehensive and mature and can meet the needs of the community, and
6. That sufficiently detailed metadata enhances data access and discovery by applications and through services like HAPI (Heliophysics Application Programmer's Interface) as well as bulk downloads,

The COSPAR PSW resolves that there is a need for a common metadata standard to facilitate and enhance international access to data acquired by space- and ground-based instruments.

Therefore, it is recommended that:

1. SPASE (<https://doi.org/10.48322/E72C-5Y75>) along with its simulations extensions (<https://doi.org/10.48322/TXC-A-X050>) be the metadata standards to describe space physics and space weather resources.
2. Funding agencies provide encouragement and adequate support to enable data

produced by projects to be described with SPASE metadata.

COSPAR Panel on Space Weather Resolution on Data Access

Accepted at COSPAR PSW Business Meeting on 18 July 2018 (updated 15 October 2021).

Taking into account that:

1. It is in the general interest of the international heliophysics and space weather community that data be made as widely accessible as possible,
2. The open exchange of data benefits from well-defined and standardized methods of access,
3. The ILWS-COSPAR Roadmap has recommended to standardize metadata and harmonize access to data and model archives, and
4. The Heliophysics Application Programmer's Interface (HAPI) specification has demonstrated that it is comprehensive and can meet the needs of the community,

The COSPAR PSW resolves that there is a need for at least one common data access API to facilitate and enhance international access to data.

Therefore, it is recommended that:

1. HAPI (<https://doi.org/10.5281/zenodo.4757597>) be the common data access API for space science and space weather data.
2. Funding agencies provide encouragement and adequate support to enable data produced by projects to be accessed by using HAPI compliant services.

COSPAR Constellation of Small Spacecraft: Implementation Phase

[Daniel N. Baker (LASP, University of Colorado Boulder, USA), Amal Chandran (Satellite Research Center, Nanyang Technological University, Singapore), Loren Chang (National Central University, China: Academy of Sciences located in Taipei), Malcolm Macdonald (University of Strathclyde, UK),

hapi-help@groups.io

hapi-server.org



JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

HAPI defines 5 URL endpoints every server must have

Endpoints must be directly below a URL that ends with 'hapi'

- `http://example.com/hapi/about`
- `http://example.com/hapi/capabilities`
 - describes options implemented by the server
- `http://example.com/hapi/catalog`
 - list of datasets at the server
- `http://example.com/hapi/info`
 - show metadata for one dataset at a time (basically a data header)
- `http://example.com/hapi/data`
 - retrieve a stream of data content for one dataset over a specific time range

Note: The intent is for computers to read from these endpoints, but humans can look at them easily too (web browser, curl, etc)

HAPI defines the format of what each endpoint returns

Endpoints must be directly below a URL that ends with 'hapi'

- `http://example.com/hapi/about`
- `http://example.com/hapi/capabilities`
- `http://example.com/hapi/catalog`
- `http://example.com/hapi/info`

These are various types of metadata.

- `http://example.com/hapi/data`

Data is streamed as an immediate response to the request.

HAPI defines the format of what each endpoint returns

Endpoints must be directly below a URL that ends with 'hapi'

- `http://example.com/hapi/about`
- `http://example.com/hapi/capabilities`
- `http://example.com/hapi/catalog`
- `http://example.com/hapi/info`

These are various types of metadata.

JSON metadata

↑
Takes dataset ID as an argument

- `http://example.com/hapi/data`

↓
Data is streamed as an immediate response to the request.

↑
Takes dataset ID and time range as args.
Returns stream as CSV, JSON, or binary.

The responses from a HAPI server are defined

- **catalog** (lists the datasets available) is JSON
- **info** (metadata for each dataset as a list of variables with types, dimensions, fill values, "bins" for spectral data) is also JSON
- **data** returns digital content as a stream of numbers
 - CSV is a required format
 - JSON (optional)
 - binary (optional)

Key standardizations in the data include a common format for the time values (ISO 8601 strings)