

# The Purpose and Benefits of HAPI for Interoperability in a Heliophysics Data Environment

- Outline:**
- What is HAPI (brief version)
  - Adoption so far
  - Discussion about interest levels
  - Recent additions to the specification (if time)

HAPI = Heliophysics Application Programmer's Interface

Jon Vandegriff

IHDEA Fall Meeting 2019

# HAPI is a specification for access to Time Series data

Time Series dataset: a collection of parameters; tabular data indexed primarily by time

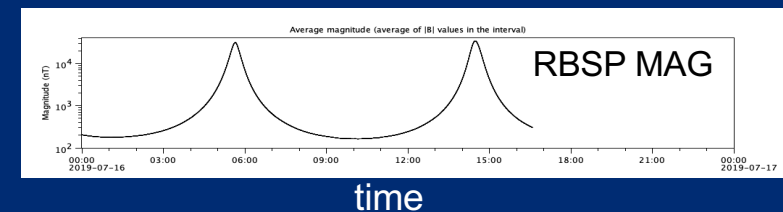
one **dataset** has many parameters with one time column

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]
...	...	...	...	...

Examples:

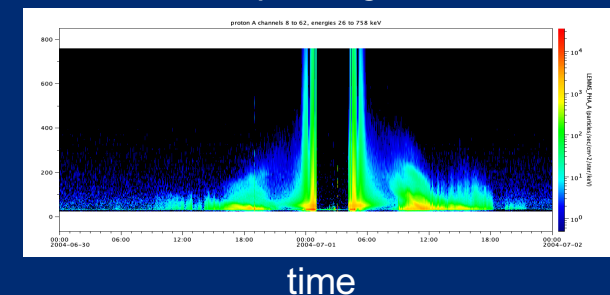
Line Plot

scalar



Spectrogram

spectrum  
(i.e., array  
in each cell)



multi-dimensional array example:  
pitch angle spectrogram for multiple energies

# The Motivation: Standardizing the Uniformizers

Examples of "Uniformizers" at this meeting:

- CDAWeb (GSFC)
- Planetary Science Archive (ESA)
- AMDA (CDPP)
- ERG Science Center (Nagoya University)
- SuperMAG (JHU/APL)
- CCMC (GSFC)
- VSO (GSFC)
- Planetary Data System (PDS/PPI node at UCLA)



Each offers time series data, and each has worked hard to offer a single interface to diverse resources.

*But each institution has a unique query mechanisms to get to their time series data resources.*

# The Assertion:

**existing data centers' time series interfaces are similar enough  
that a new standard could capture a useful lowest common denominator**

What if everyone just agreed to do this: (i.e., HAPI in a nutshell):

INPUT: dataset name and time range

```
http://datashop.elasticbeanstalk.com/hapi/  
data?id=CASSINI_MAG_HI_RES&  
time.min=2004-183T00:00:00&time.max=2004-184T00:00:00
```



OUTPUT: comma-separated ASCII values

```
2004-183T00:00:03.403Z, 1.0724e+02, -6.8993e+01, -5.1978e+02  
2004-183T00:00:07.153Z, 1.0842e+02, -6.8956e+01, -5.1962e+02  
2004-183T00:00:10.907Z, 1.0855e+02, -6.9063e+01, -5.2084e+02  
2004-183T00:00:14.653Z, 1.0852e+02, -6.9049e+01, -5.2085e+02  
2004-183T00:00:18.403Z, 1.0849e+02, -6.9035e+01, -5.2085e+02  
2004-183T00:00:22.153Z, 1.0862e+02, -6.9142e+01, -5.2207e+02  
2004-183T00:00:25.903Z, 1.0859e+02, -6.9128e+01, -5.2208e+02
```

*note: basic metadata is part of HAPI  
but is not shown here!*

# The HAPI Specification

Many more details I've not told you – things implementors will need to know!

## HAPI Data Access Specification

Version 2.0.0 | Heliophysics Data and Model Consortium (HDMC) |

### Table of Contents

- [HAPI Data Access Specification](#)
- [Introduction](#)
- [Endpoints](#)
- [hapi](#)
- [capabilities](#)
- [catalog](#)
- [info](#)
- [data](#)
  - [Data Stream Content](#)
- [Implications of the HAPI data model](#)
- [Cross Origin Resource Sharing](#)
- [HAPI Status Codes](#)
- [HAPI Client Error Handling](#)
- [Representation of Time](#)
- [Incoming time values](#)
- [Outgoing time values](#)
- [Additional Keyword / Value Pairs](#)
- [More About](#)
- [Data Types](#)
- [The "size" Attribute](#)
- ["fill" Values](#)
- [Examples](#)
- [Security Notes](#)
- [Adoption](#)
- [References](#)
- [Contact](#)
- [Appendix A: Sample Landing Page](#)
- [Appendix B: JSON Object of HAPI Error Codes](#)

14 pages total

## Introduction

This document describes the Heliophysics Application Programmer's Interface (HAPI) specification, which is an API and streaming format specification for delivering digital time series data. The intent of HAPI is to enhance interoperability among time series data providers. The HAPI specification describes a lowest common denominator of services that any provider of time series data could implement. In fact, many providers already offer access to their data holdings through some kind of API. The hope is that this specification captures what many providers are already doing, but just codifies the specific details so that providers could use the same exact API. This would make it possible to obtain time series science data content seamlessly from many sources.

This document is intended to be used by two groups of people: first by data providers who want to make time series data available through a HAPI server, and second by data users who want to understand how data is made available from a HAPI server, or perhaps to write client software to obtain data from an existing HAPI server.

HAPI constitutes a minimum but complete set of capabilities needed for a server to allow access to the time series data values within one or more data collections. Because of this focus on access to data content, HAPI is very light on metadata and data discovery. Within the metadata offered by HAPI are optional ways to indicate where further descriptive details for any dataset could be found.

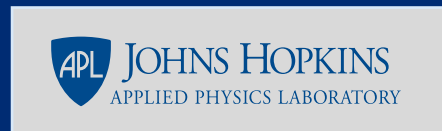
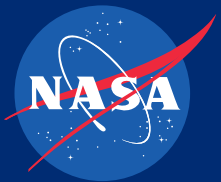
All HAPI-related projects are online at:  
<https://hapi-server.github.io/>



# HAPI Adoption

hapi-news@hapi-server.org

## Existing servers



(not part of  
specification  
team!)



## Existing clients



Autoplot



a basic  
Python  
library

*an IDL  
example*

*a basic  
Java  
library*

# HAPI Adoption – The Verifier

A way to test if your server meets the specification.

- green, yellow, red indicators for basic functionality and most corner cases
- can test one endpoint, or one dataset, or the whole site (every endpoint and dataset)

Two ways to use the verifier:

1. Download the verifier code and run it yourself:

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

2. Submit your HAPI URL to a service.

<https://hapi-server.org/verify>

# What is needed next:

The specification is in good shape, but for adoption, you need more servers to justify more clients...

The tautological expectation: if everyone would adopt it, HAPI would enable interoperability:  
more data centers add HAPI access method  
more clients to support HAPI

To enable this the HAPI team needs to:

- create a semi-generic "drop-in" server for exposing existing data via HAPI (this is hard!)
- develop better client libraries for all major languages (easier)
- sponsor workshops / tutorials / implementation assistance (would do this if asked)

Eventually, HAPI becomes invisible infrastructure that just allows data to flow easily.



# Interest level discussion

Possible discussion questions:

Is your institution planning on implementing a HAPI server?

How much assistance would you want in this process?

What aspects of HAPI make you hesitate to adopt it?

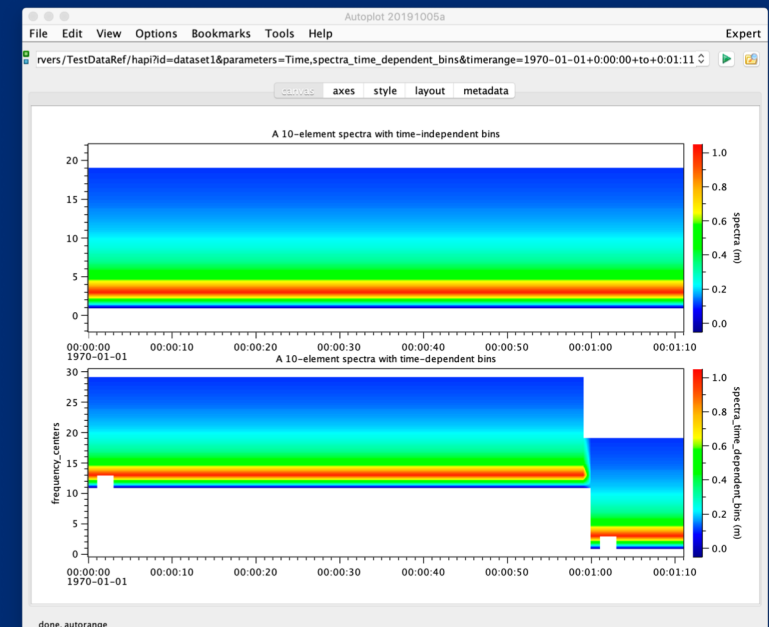
How involved do you want to be in controlling the evolution of the specification?

Do you think a streaming format will be a viable way to serve your data going forward?

# Next features to be added to the spec

- currently at HAPI version 2.1.0
- version 3.0.0 will be next
- backwards compatible additions:
  - a. support for "time-varying bins"  
(specify a parameter that has the bin ranges)
  - b. support for some linkages in metadata  
(references allow bins to be identical for data and uncertainties)
  - c. a way to indicate which units framework you use

some datasets, the spectrogram bin ranges change over time



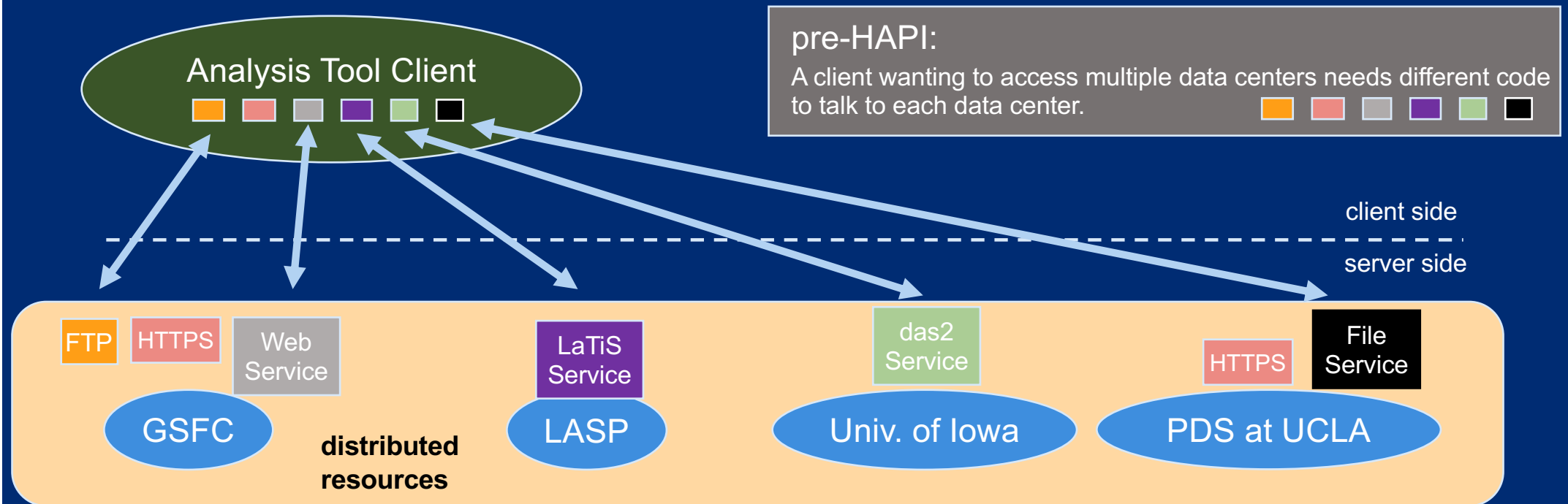


JOHNS HOPKINS  
APPLIED PHYSICS LABORATORY

# Backup Slides

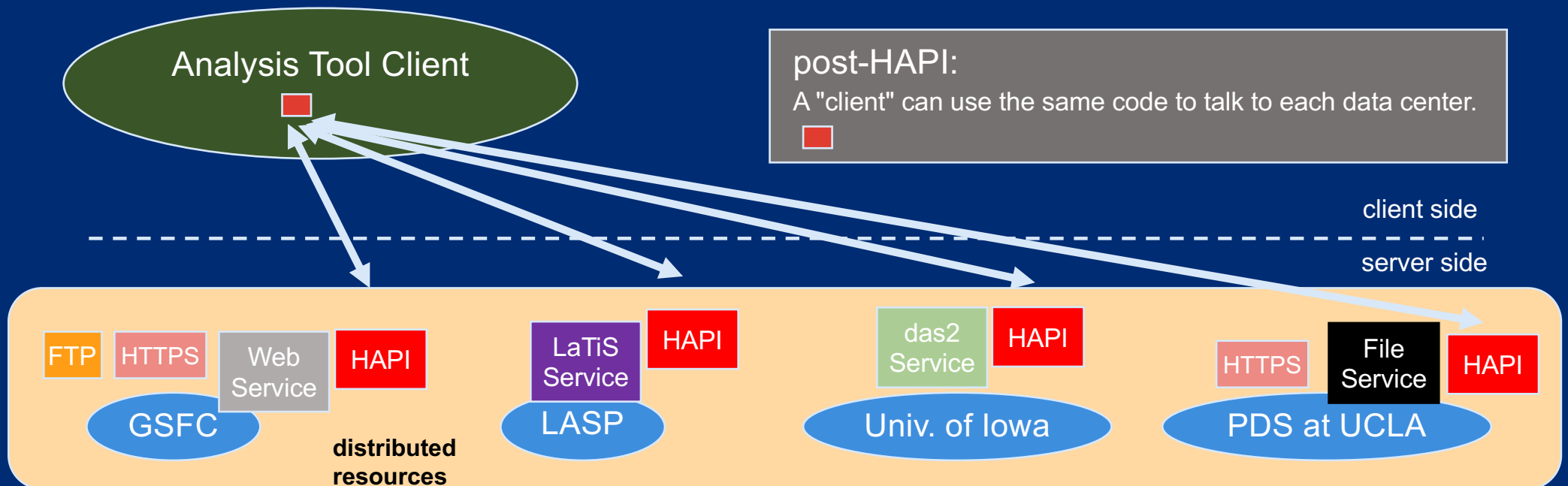
# Current data systems are similar but not identical

- many Heliophysics data centers have an access API
- they are similar, but not API compatible



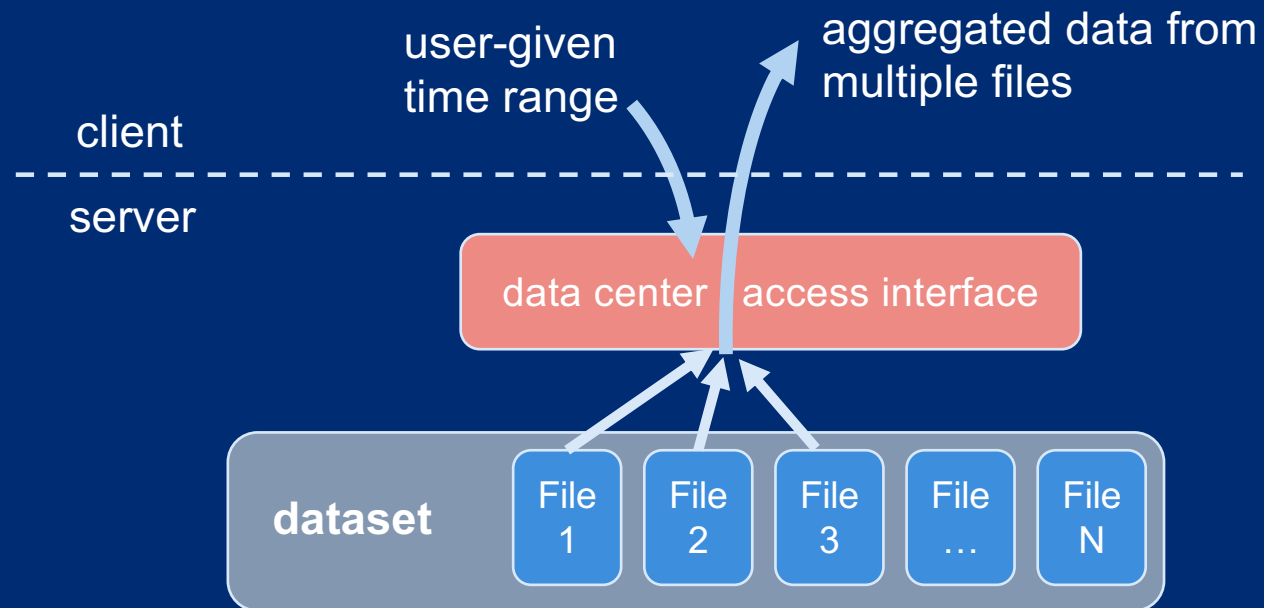
# HAPI captures the lowest common denominator

- focus of HAPI is plain access with nothing fancy (no filtering, averaging, etc)
- often, HAPI can be integrated into an existing service's code base



# Simplified Data Center View

Many data centers offer a service to request data content programmatically (computer-to-computer)



*the aggregate can be:  
a single merged file, or  
a stream;*

*for HAPI, it is a stream*

# Key Elements of the HAPI Spec

- request interface – how do clients request data (and metadata)

4 REST-ful endpoints, each with simple parameters (or none)

**capabilities**

**catalog**

**info** (one input, the dataset name)

**data** (multiple inputs: time range, dataset name, desired parameters)

- response interface – what is the format of the returned data stream (and metadata)
  - a. uniform view of the data, regardless of its storage format on the server
  - b. time column is always the first column
  - c. time format always the same
  - d. simple data layout that also allows for multi-dimensional data
  - e. contains spectral bin parameters (energy ranges, frequency ranges, etc)

HAPI is very light on metadata;

The focus is data delivery with the minimum metadata to make the data scientifically useful

