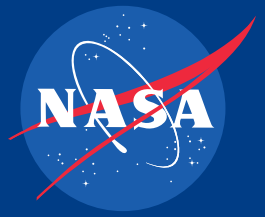
JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Interoperable Data Access via HAPI

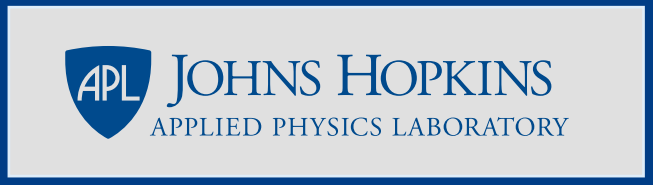
(Heliophysics Application Programmers Interface)

Jon Vandegriff, Johns Hopkins Applied Physics Lab
Todd King, University of California Los Angeles
Robert Weigel, George Mason University
Jeremy Faden, Cottage Systems / University of Iowa
Aaron Roberts, NASA Goddard Space Flight Center
Bernard Harris, NASA Goddard Space Flight Center
Robert Candey, NASA Goddard Space Flight Center
Nand Lal, NASA Goddard Space Flight Center
Scott Boardsen, NASA Goddard Space Flight Center
Doug Lindholm, Colorado University / LASP
Larry Brown, Johns Hopkins Applied Physics Lab
Eric Grimes, University of California Los Angeles


NASA

UCLA


LASP
Laboratory for Atmospheric and Space Physics


APL
JOHNS HOPKINS
APPLIED PHYSICS LABORATORY


THE UNIVERSITY OF IOWA


GEORGE
MASON
UNIVERSITY


l'Observatoire
de Paris


esa

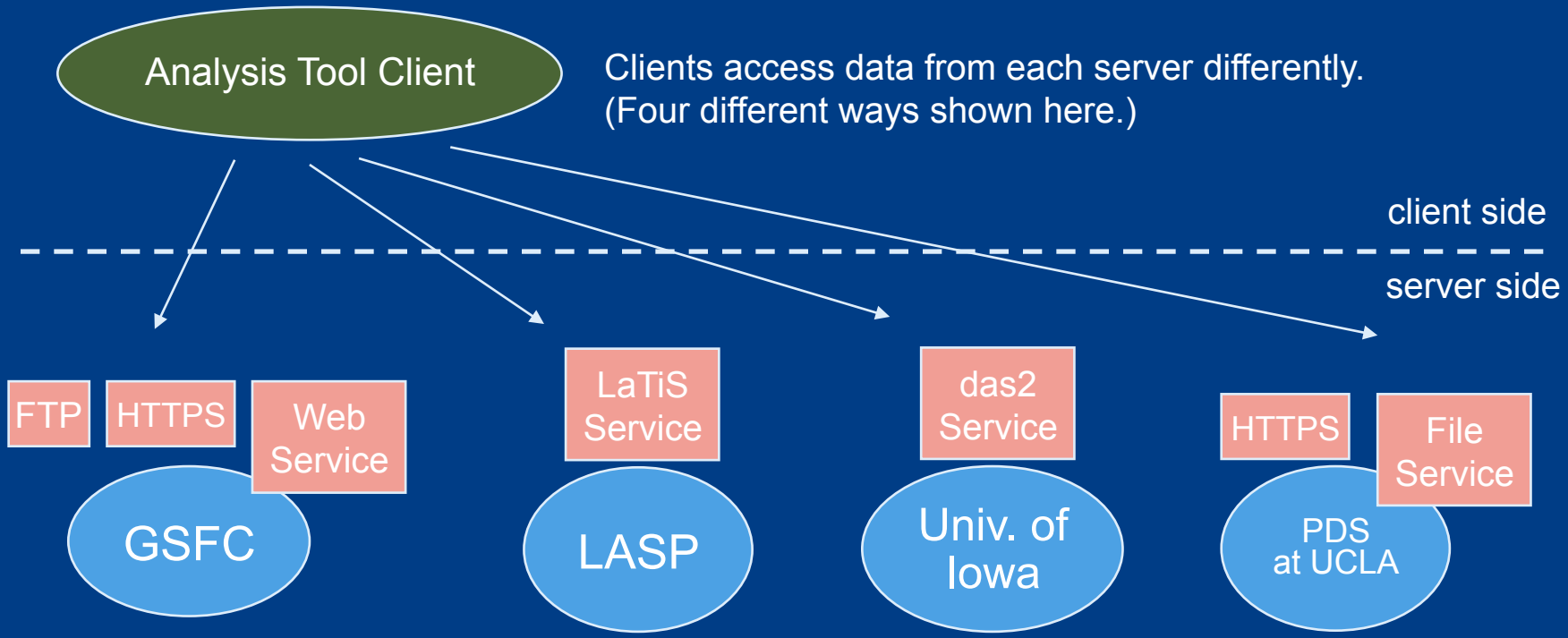
Motivating the need for interoperability

- lots of data sources being considered at this meeting
  - model input (indices, solar wind properties), model output, spacecraft trajectories, sensor measurements (for comparison with models)
- these usually come from disparate sources

- example: Omniweb data as input to a magnetic field model, whose output should be compared to GOES magnetometer data along the spacecraft trajectory
  - all of these could have a different access method and format

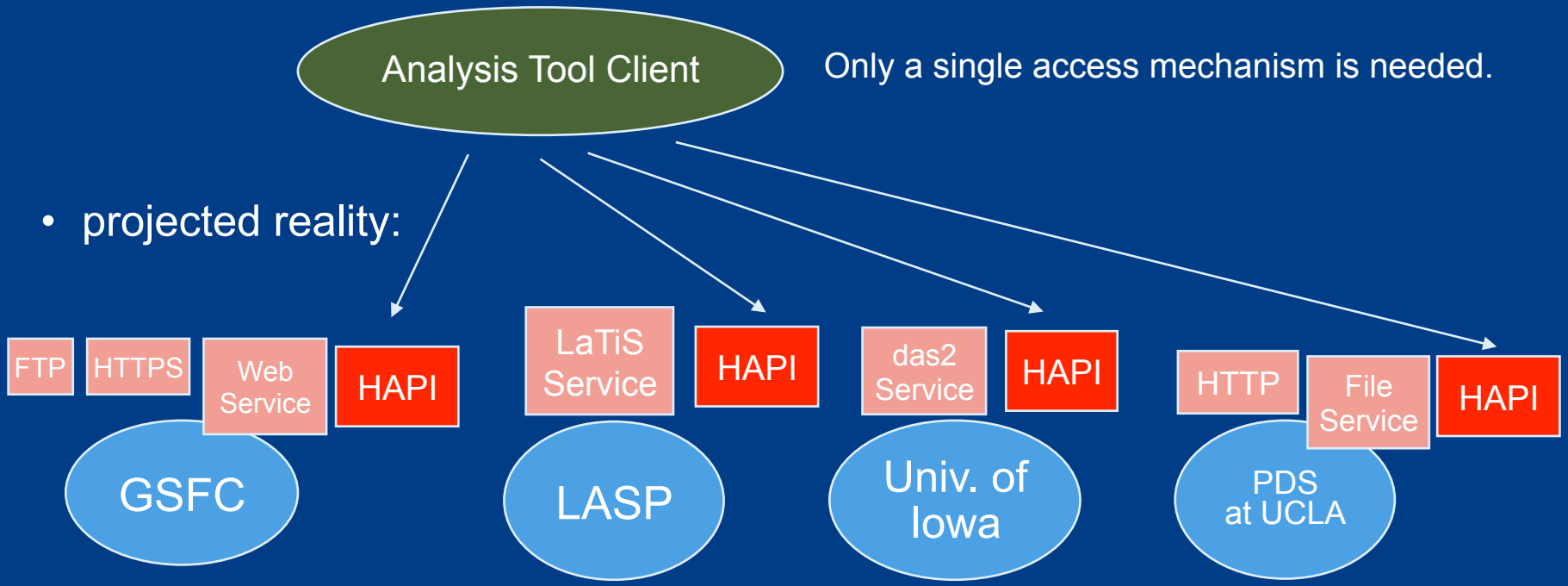
Putting data online is easy, so diverse options exist

- creating an online data access mechanism has been done many times
- different groups have made similar but not identical access mechanisms
- even if everyone had used the same standard access mechanism, these standards are flexible enough that they would not automatically support interoperability



Interoperability requires a simple, common access mechanism: HAPI

- nothing new in terms of technology or complexity
- just need everyone to agree to use the same lowest common denominator
- can add a HAPI mechanism alongside existing services



- projected reality:

Focus is Timeseries Data

- conceptually a table, like a spreadsheet
- time column with any number of variables (data columns)
- each variable can be multidimensional

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]

Types of Data Access

1. **direct access to files: http or ftp**
   - obvious standards needed: file format, file names, directory structure
   - more subtle standards needed for full interoperability:
     - file metadata
     - variable layout
     - time value format
2. **service-based access: CGI, web services (RESTful or custom)**
   - obvious standards needed:
     - request format (how to ask for data)
     - response format (what does the returned data look like?)
       - collection of files, OR
       - stream of numbers (JSON, XML, CSV binary, etc)
   - more subtle standards:
     - file metadata
     - variable layout
     - time value format

HAPI (Heliophysics Application Programmer's Interface)

- HAPI is primarily a specification – what to do on your server to make your data interoperable

  `https://hapi-server.github.io/`

- two key parts:
  - request interface
  - response format

The request interface: 4 HAPI Endpoints

All endpoints must be directly below a URL that ends with '**`hapi`**'

- `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

The response data from a HAPI server

- **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 and binary data are optional

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

The response data from a HAPI server

capabilities

list the optional elements implemented by the server

arguments: none; return format: JSON

catalog

list the names of the datasets the server can provide

arguments: none; return format: JSON

info

provides a description of one dataset; the metadata is minimal, but enough to configure a reader

arguments: dataset identifier; return format: JSON

data

(This provides CSV. JSON and binary are also possible.)

`json`

streams data content in a format defined by the spec; time values must all be standardized

arguments: dataset identifier, time range, and desired output format; return format: CSV, binary or JSON (details defined by spec)

`binary`

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

Sample access

http://datashop.elasticbeanstalk.com/hapi/
data?id=CASSINI_LEMMS_PHA_CHANNEL_1_MIN_AVG&
time.min=2004-05-02T00Z&time.max=2004-05-02T06Z&
include=header

Sample access

http://datashop.elasticbeanstalk.com/hapi/catalog


```
{ "HAPI": "2.0",

  "status": { "code": 1200, "message": "OK" }

  "catalog": [
    { "id": "CASSINI_LEMMS_PHA_PITCH_ANGLES_10_MIN" },
    { "id": "WEYGAND_GEOTAIL_MAG_CPI_GSE" },
    { "id": "WEYGAND_GEOTAIL_MAG_GSM" },
    { "id": "WEYGAND_WIND_MAG_GSE" },
    { "id": "Wind_EPACT_LEMT_Events_OMNI_5min_C" },
    { "id": "Wind_EPACT_LEMT_Events_OMNI_5min_FE" },
    { "id": "spase://VMO/NumericalData/Prognoz9/MAG/PT10S" },
    { "id": "spase://VSPO/NumericalData/ACE/MAG_SWEPAM/PT240S" },
    { "id": "spase://VSPO/NumericalData/Cassini/MAG/PT60S" } ]
}
```

Working examples

Servers

George Mason
GSFC
Univ. of Iowa
JHU/APL

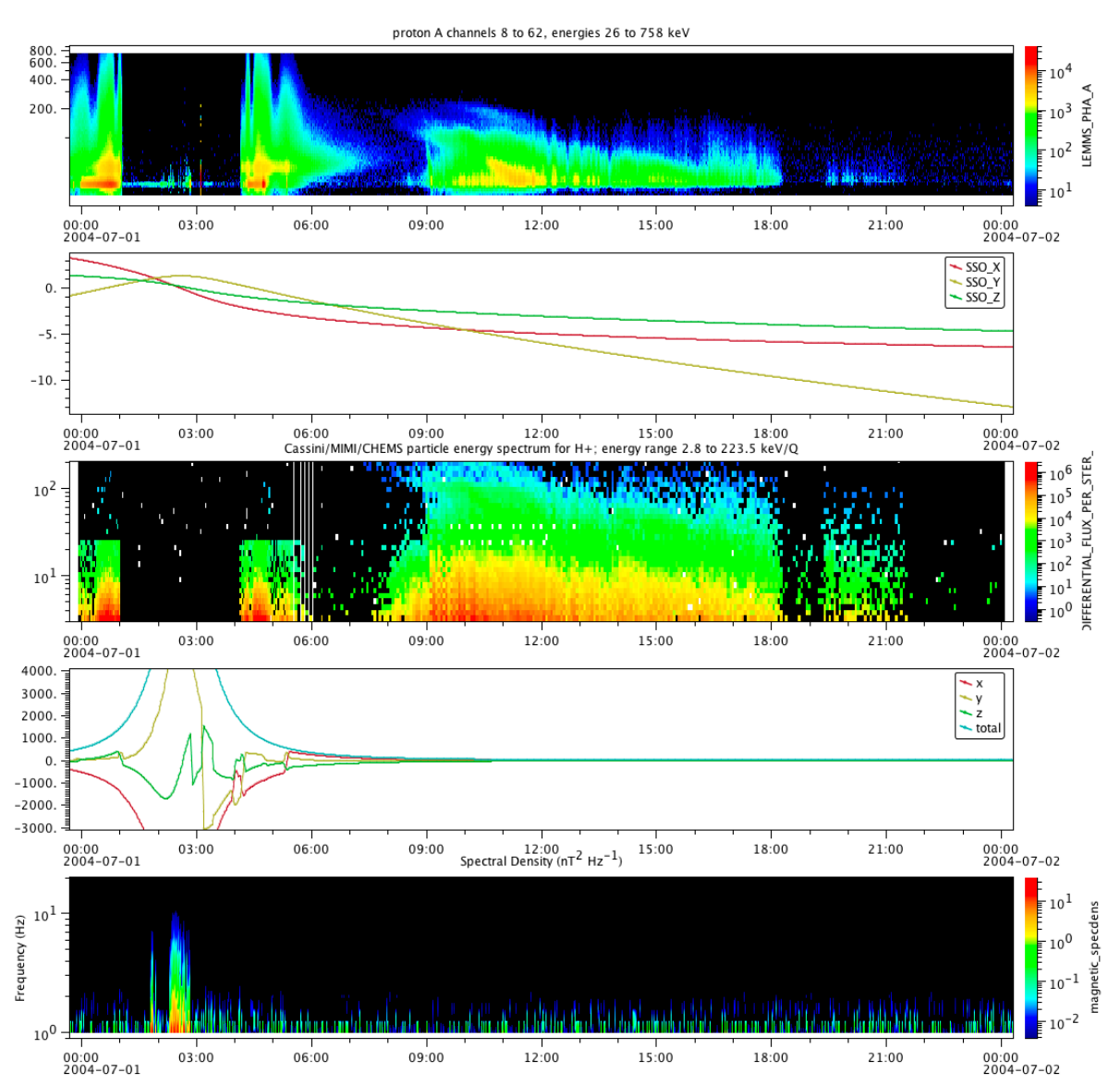
Soon:
LASP
UCLA PDS PPI Node

Clients

Autoplot
SPEDAS
MIDL4
Python client
IDL client

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

Autoplot with panels from several HAPI servers



energetic particles (LEMMS proton energy spectra)

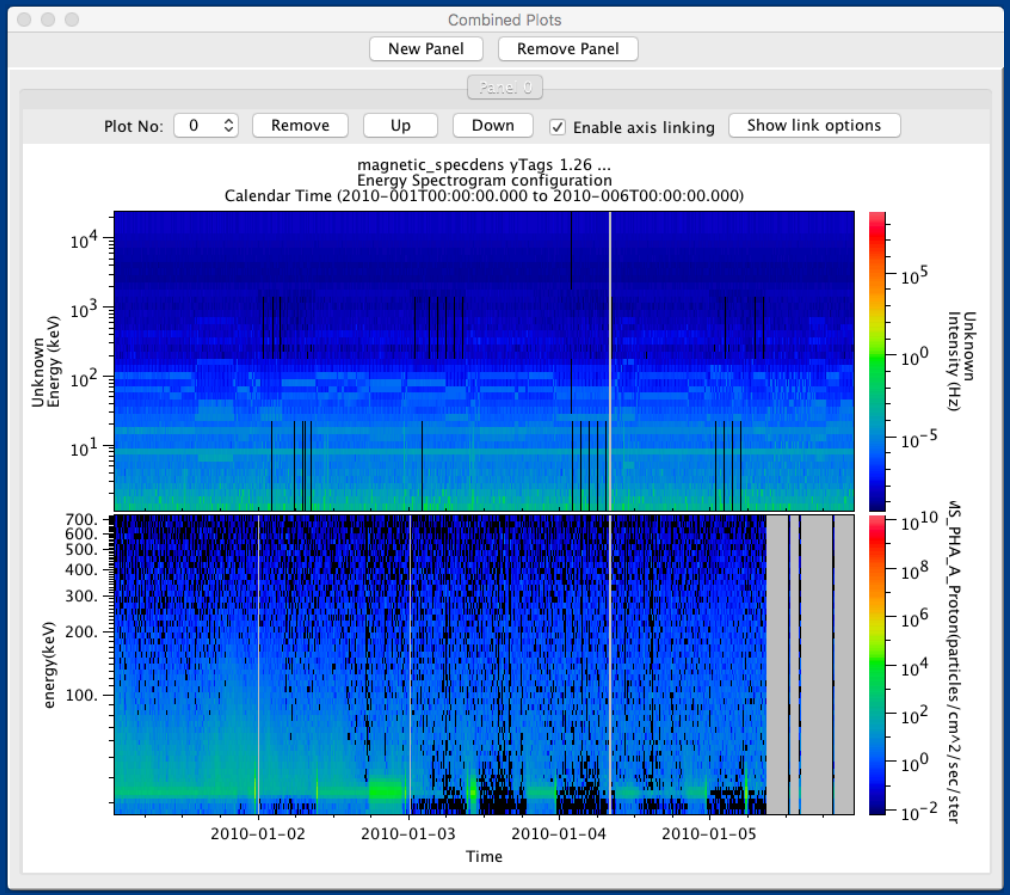
Cassini trajectory data

energetic particles (CHEMS proton energy spectra)

MAG data

plasma wave data RPWS

MIDL4 showing multiple panels



In development

- servers at PDS PPI node at UCLA
- server at LASP
- interest from European providers
- additional data from Cassini mission

- generic server capability

If you want to set up a HAPI server

- read the specification -- it is very complete
- see if HAPI protocols can be added within an existing access mechanism at your site
- test your server
  - use the HAPI verifier: **`http://tsds.org/verify-hapi`**
    - this checks all of the important parts of the spec, and many corner cases
  - try your server in an existing client (Autoplot, SPEDAS, MIDL4)
- sign up on the mailing list: **`hapi-news@hapi-server.org`**
- ask for help
  - email anyone involved in making the spec (authors on this paper)
  - contact other early adopters

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY