JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

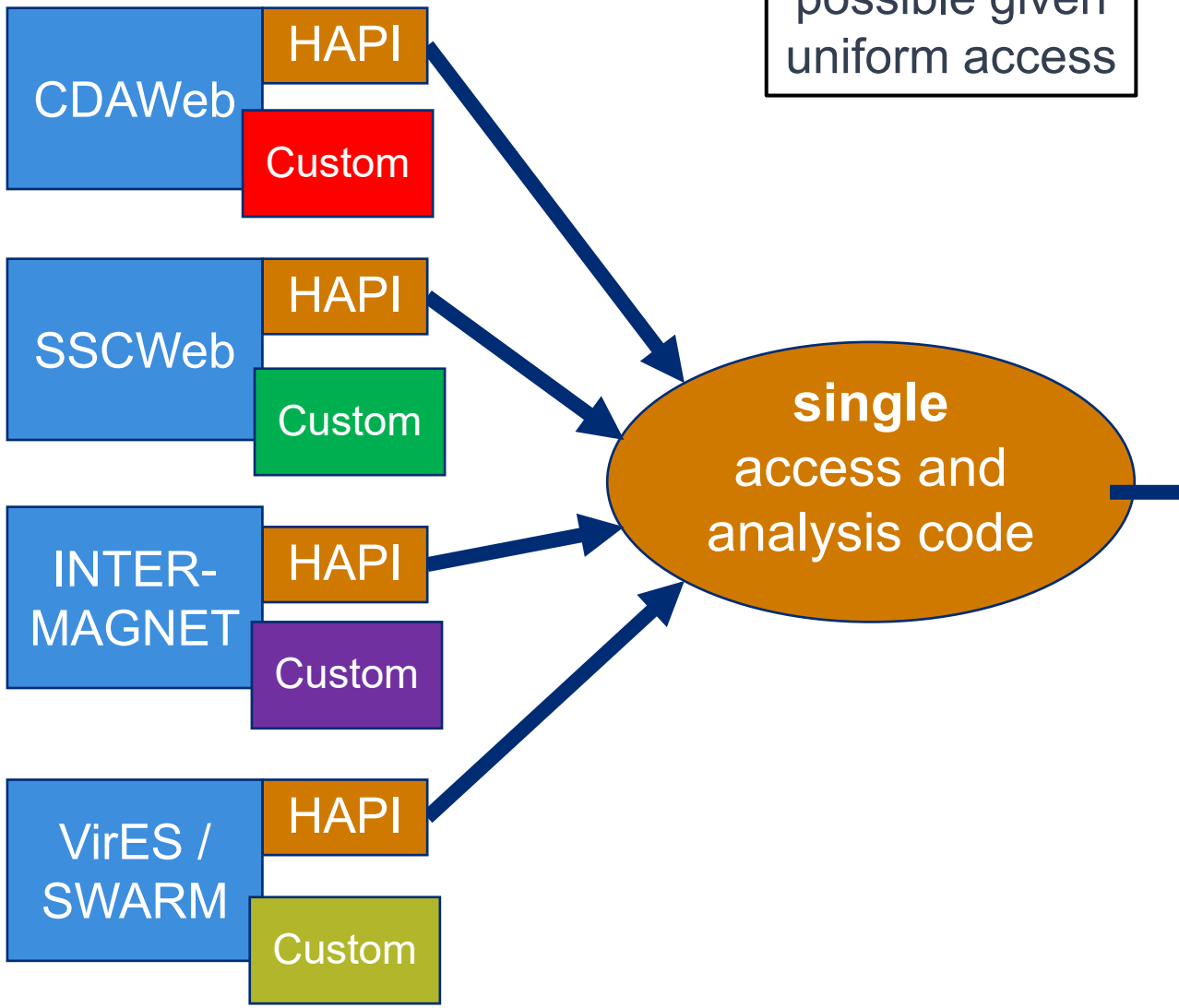
Pathways to Better Interoperability: Future Directions and Challenges for HAPI

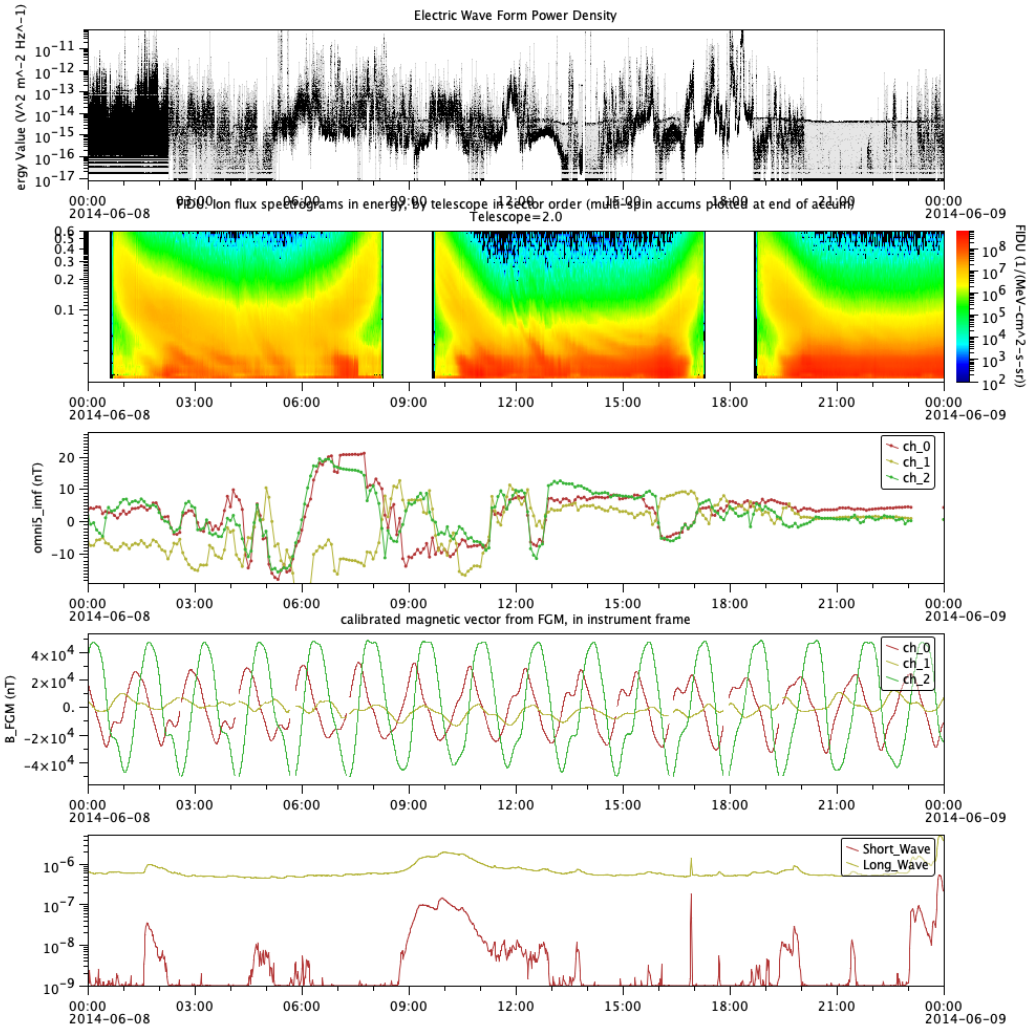
Jon Vandegriff, JHU / APL
Robert S. Weigel, GMU
Sandy Antunes, JHU / APL
Jeremy Faden, Cottage Systems

Servers ***add*** HAPI as a standard access mechanism.

Interoperability becomes possible given uniform access

Powerful analysis capabilities become possible





Where does HAPI fit in the ecosystem?

- meant to solve the **ACCESS** problem:
  - “I know what data I want, I just need to get this data into my workflow”
- focused on numerical data **indexed by time**
- a lowest common denominator offering **simple but universal access**
- targets “fully cooked” data -- higher level products in **well arranged “datasets”**)
- is meant to **become invisible plumbing** once widely adopted
  - this is happening to astronomy VO standards

Production HAPI Servers

Institution	Server	Version	Type of Data	Num Datasets
NASA / GSFC	CDAWeb	2.0	Heliophysics	3425
	SSCWeb	2.0	Ephemeris	298
	Helioviewer	3.1	Solar Images (URLs)	79
	CCMC	2.0	Space Weather Indices	275
IRAP Plasma Data Ctr	AMDA	2.0	Helio. & Planetary Data and Ephemeris	1029
University of Iowa	Das2 Server	1.1	Helio. & Planetary	19
CU Boulder / LASP	LISIRD	2.0	Solar Irradiance	29
SWARM Mission	ViRES Data Server	3.0	Space Mag Data	123
INTERMAGNET	INTERMAGNET	3.1	Ground-based Mag	2915
Royal Netherlands Meteorological Institute	KNMI	2.0	Space Weather	665
ESA / ESAC	ESAC / Cluster Mission Data	3.2	Helio. (magnetosphere)	1989
			Total	**10846**

Institution	Server	Type of Data	Num. Datasets
ESA / ESAC	Solar Orbiter and others	Heliophysics	lots
JHU / APL	SuperMAG	global ground mag	~500
JHU / APL	TIMED / GUVI	ionospheric images	~10
UCLA	PDS PPI Node	Planetary Plasma, Particle, and Fields	~1000
NOAA / SWPC	SWPC	Space Weather	~50

On the horizon			
Univ. of Calgary	Space Environment Canada	Ground-based ionospheric data (riometers)	~dozens?
CEDAR / Madrigal	Open Madrigal	Space Weather	1000+
NCAR ? (in discussion)	*Multiple*	*Atmospheric*	*TBD*

highlight indicates server is written and scheduled for deployment

HAPI Clients

Plain reader libraries in:

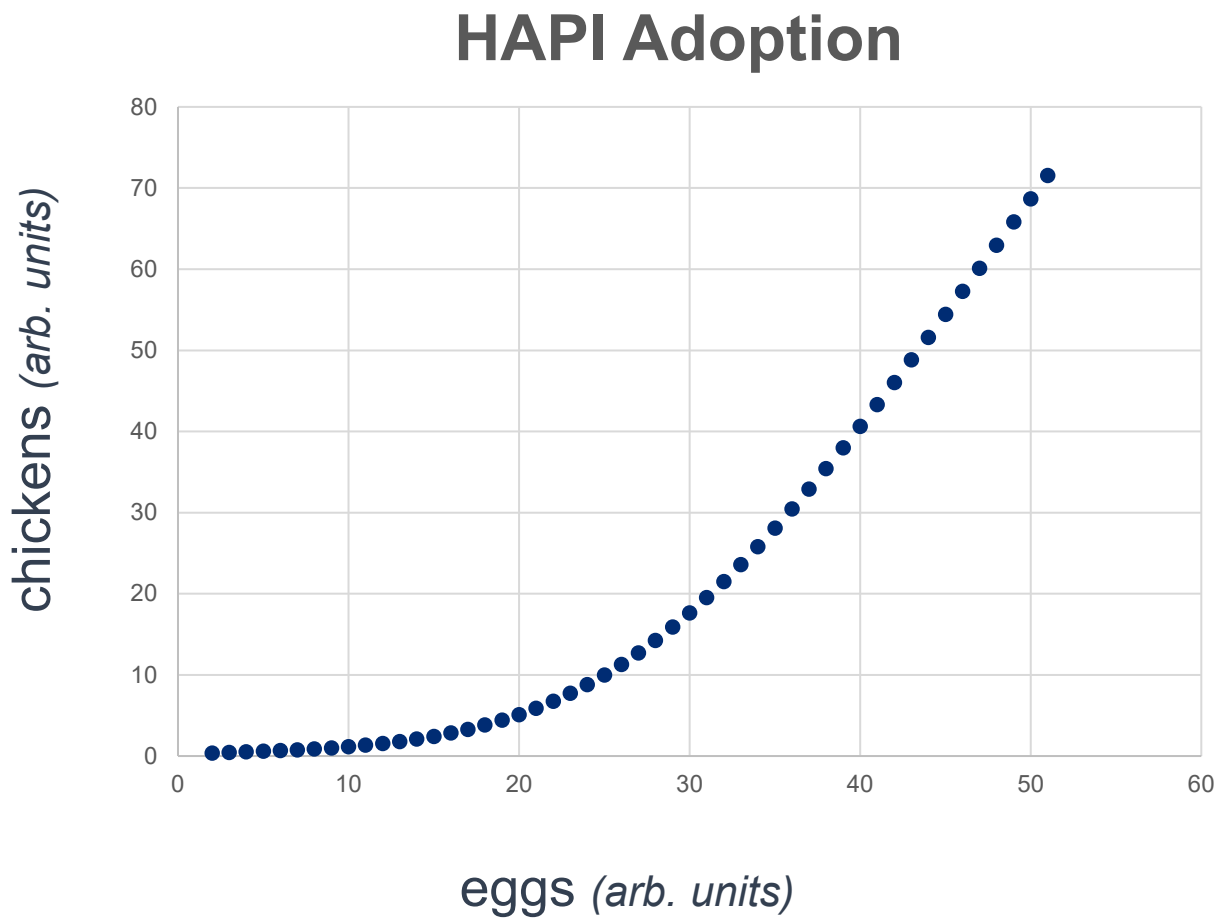
- Python
- Matlab
- IDL
- Java
- R
- JavaScript

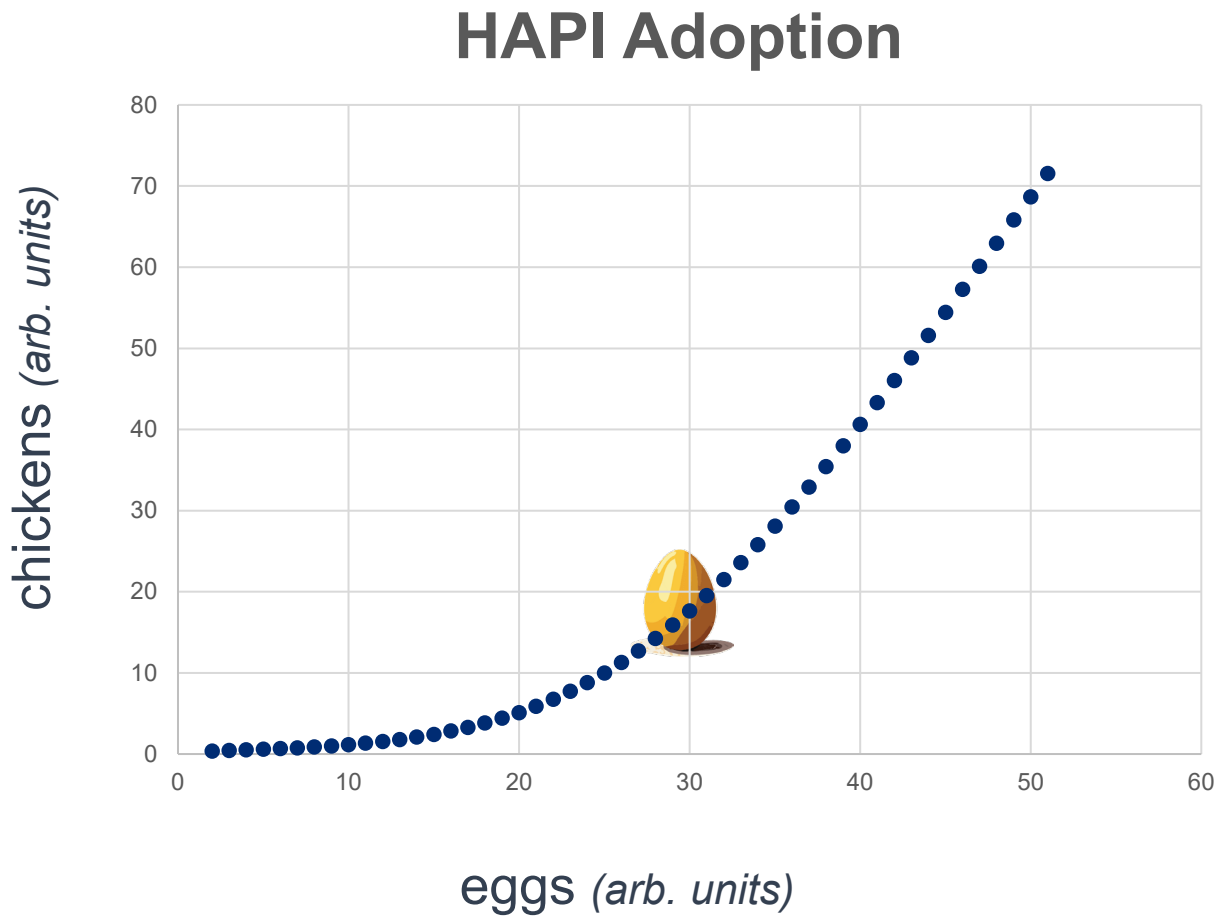
Client Applications:

- **TopCat**
- SPEDAS
- PySPEDAS
- Autoplot
- KNMI Viewer
- LASP Space Weather Viewer

Coming Soon:
SpEasy / SciQLop

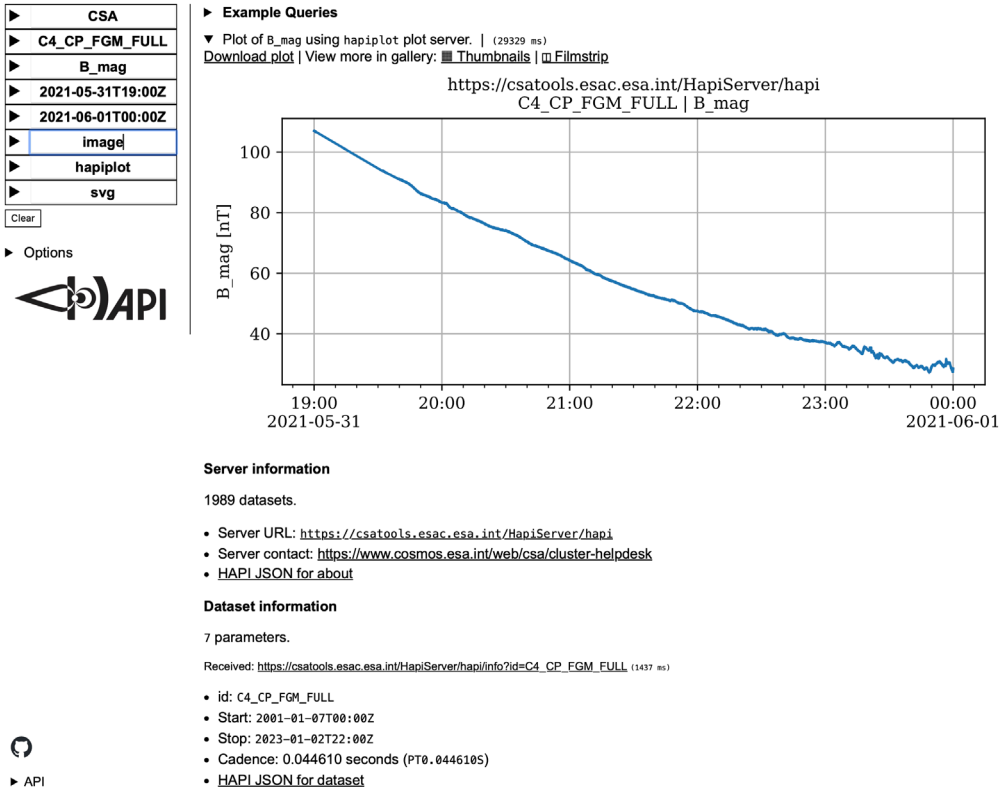
HAPI Adoption


HAPI Adoption
chickens (arb. units)
0
10
20
30
40
50
60
70
80
0
10
20
30
40
50
60
eggs (arb. units)


HAPI Adoption
chickens (arb. units)
eggs (arb. units)
0
10
20
30
40
50
60
70
80

Best way to explore and use HAPI data

https://hapi-server.org/servers/



- curated list of production HAPI servers
- simple web page illustrates how to use the various parts of the API
- view metadata from each HAPI server
- simple plots for quick exploration
- code generation in many languages

How to check the validity of a newly implemented server?

Use the HAPI verifier!

- Runs a battery of tests to see if the server meets the spec
  - hits all the endpoints
  - requests data
  - tries bogus inputs
  - lists each test done and gives **GREEN (pass)** **YELLOW (warning)** **RED (fail)**
- Run locally (recommended)
  - code and instructions are here: https://github.com/hapi-server/verifier-nodejs
- Web version (just as comprehensive, but slow – lots of requests!)
  - http://hapi-server.org/verify/
  - enter the top URL to your server, and another server will run all the tests

For more info on HAPI

Main HAPI web site: https://hapi-server.org

Weigel et al, 2021, **HAPI: An API Standard for Accessing Heliophysics Time Series Data,** Journal of Geophysical Research, Vol 126, Issue 12, https://doi.org/10.1029/2021JA029534

HAPI Specification at Github: https://github.com/hapi-server/data-specification

HAPI Organization at Github (many projects: clients in Python, IDL, Java, R, HAPI servers, a verifier for serv er developers): https://github.com/hapi-server

Tutorial Jupyter Notebooks: https://github.com/hapi-server/tutorial-python

(These tutorials were used at two PyHC summer schools.)

Since 2021, HAPI is a COSPAR-endorsed standard for Space Weather data access

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.

Challenges facing HAPI

- Performance of HAPI servers ends up reflecting on the standard
  - if a server is down, or is slow, or has bad metadata, this translates to
  - *Solution: make centralized status dashboard for known HAPI servers*
  - *Solution: work with data centers more to get better support*
- People expecting HAPI to help with the data search process
  - there are enough HAPI servers now that people start looking through them to find stuff
  - HAPI metadata is focused on access and use, not discovery and sort-of not search
  - *Solution: HAPI team could build either a central or a federated search capability (SPASE is for this)*
  - *Solution: build SEARCH capability into the tools people are using for access, and then have search results that seamlessly bring in the data*
- Our generic servers have not been used as much as we thought
  - data centers have usually just added HAPI to their own code base
  - we had planned to be improving our test implementations over time, but this is only happening with the Java server
  - **this means server upgrades to a new spec will have to be done by each server project**
  - we need to support older HAPI versions for a long time
  - *Solution: write really robust generic server! (issue is funding)*
    Note: could be used for HAPI-fying small datasets (research data products)
- Some duplicate datasets are now appearing in more than one HAPI server
  - some ground magnetometer data has different arrangements and also different values!
  - *Solution: add robust Provenance to HAPI metadata*
- No built-in way to centrally capture HAPI usage at all servers
- Cloud services require a login – HAPI does not have this

Achieving greater interoperability in Heliophysics

What is needed for wider adoption of HAPI?

- **Should IHDEA recommend / promote HAPI as a standard?**
  - Used by: NASA and ESA Heliophysics data centers and missions, NOAA Space Weather Prediction Center, Ground Magnetometers (US National Science Foundation)
  - already is COSPAR standard
  - official standardization by IHDEA would encourage commitment by data centers and projects
- Continued development of client-side tools
  - **the analysis impact of uniform access is still largely untapped**
  - HAPI is underlying infrastructure – we don't really want scientists to think about it, just use it
  - this kind of invisible plumbing is hard to fund
- Easier ways needed for scientists to make their own data available using HAPI
  - serve out research products via "pop-up" HAPI severs
- Lots more user awareness and education
  - PyHC Summer School is helping with this
  - more outreach at other science meetings – follow the SPEDAS model for user education

IHDEA and Standards

- IHDEA is the entity to be making community recommendations for standards
- a lightweight process could be created or stolen / adapted from IVOA
- the HAPI team volunteers to help flesh this process out

- larger issue: role & activity level & expectations for IHDEA

Planned Enhancements to the HAPI Ecosystem

1. **adding ability to link HAPI datasets together**
   - allow automate4d connection of datasets differing by cadences: MAG_5_min **isAveragedVersionOf** MAG_Hi_res
   - could express other linkages too (isCalibratedVersion, isAvailabilityInfoFor)
   - would use JSON-LD terms, hopefully existing ones!
2. **making HAPI FAIR compliant**
   - need to add `license` and `provenance` attributes to each HAPI `dataset`
   - *these are in progress*
3. centrally tracking outages of HAPI servers
   - helps preserve the HAPI brand (it's not HAPI's fault that it's not working now)
   - it's a simple task to poll existing HAPI servers and report on status
   - we have a plan for this and will have something by early next year
4. collect all HAPI metadata centrally – could enable search
   - could enable a basic search capability
   - cycle through all servers once per day to harvest detailed metadata
   - can overlay any missing elements (FAIR ones, for example)
5. **develop a data caching client capability for speeding up client access**
   - this is well along – will have poster at AGU
6. HAPI to SPASE helper / converter
   - still in the planning phases
   - could use this at Madrigal if: a) we finish in time b) Madrigal implements HAPI in time

HAPI Enhancements (continued)

7. offer standard schemas to regularize HAPI content for common output types

HAPI now allows listing of files (URIs) in addition to data.

So what should a HAPI file listing look like? **A schema is needed!**

Following the HAPI mantra (keep it simple), then for file listings:

1. **must** have start time of file contents as first column, and file URI as second column
2. *may* have these optional fields after the required first two: end time, file size in bytes, mod time
3. *may* contain any other custom field that HAPI will just ignore

Time	**filename as URI string**	*end time*	*size*	*mod-time*	*other custom fields*
t0	http://data.org/image0.fits	*et0*	*s0*	*mt0*	
t1	http://data.org/image1.fits	*et1*	*s1*	*mt1*	
t1	http://data.org/image2.fits	*et2*	*s2*	*mt2*	
t2	http://data.org/image3.fits	*et3*	*s3*	*mt3*	
t4	http://data.org/image4.fits	*et4*	*s4*	*mt4*	

not final -- still under active discussion!!

Could use schemas for science data!

Schema defines essential aspects of a science measurement.
would be some requirements on both data and metadata (i.e., must specify coord. system)

Data providers will not want to change data to match this schema.
So this would require a mapping of existing data into the required elements.
=> Lots of metadata work, so this will be slow

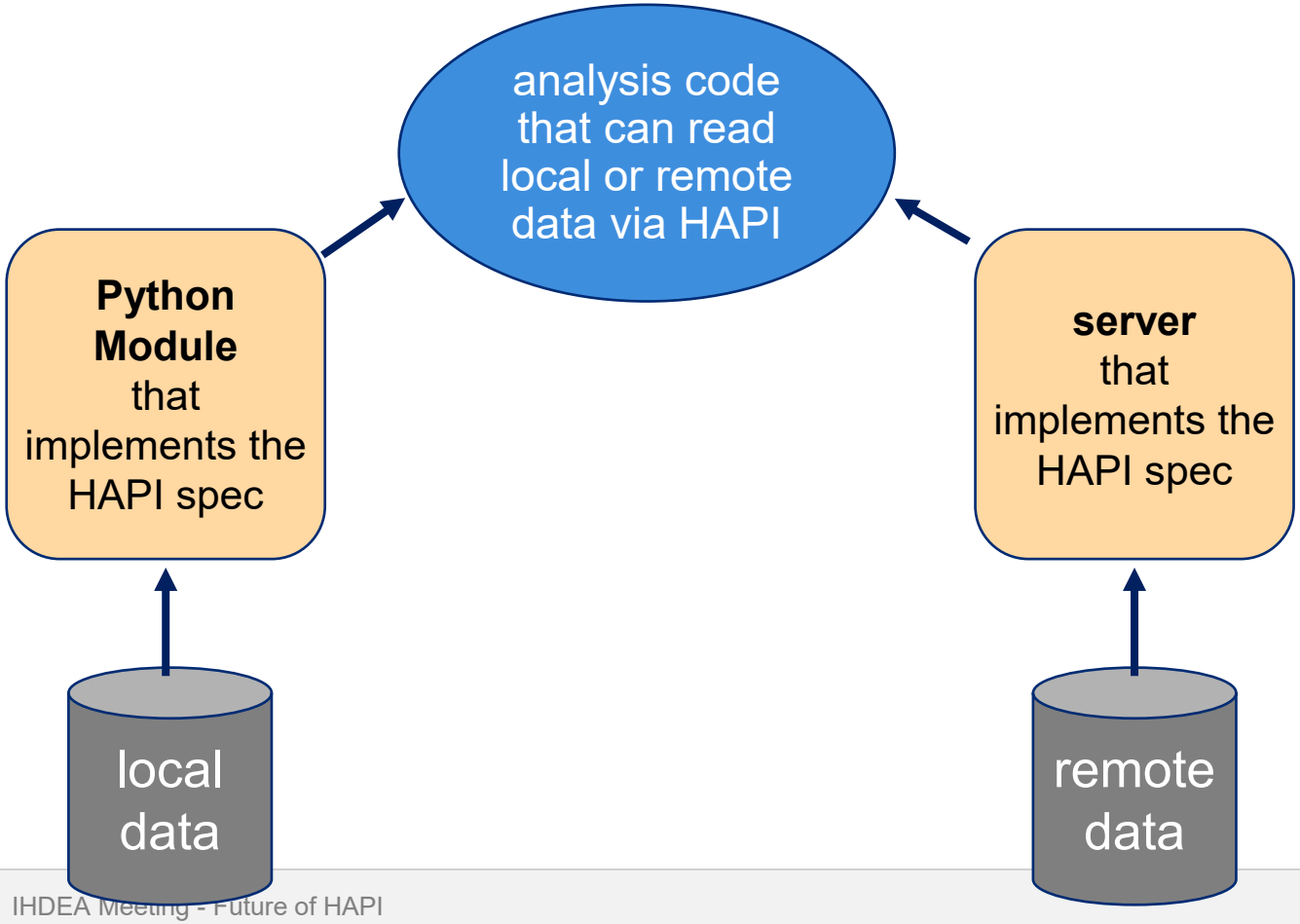
What if every magnetometer dataset looked like (or could be automatically mapped to) this:

Time	**Magnetic Field**	*Mag. Field Magnitude*	*Mag. Field Uncertainty*	*other custom fields*
t0	b0[3]	*m0[3]*	*u0[3]*	
t1	b1[3]	*m1[3]*	*u1[3]*	
t1	b2[3]	*m2[3]*	*u2[3]*	
t2	b3[3]	*m3[3]*	*u3[3]*	
t4	b4[3]	*m4[3]*	*u4[3]*	

Then:
define JSON-LD term for this content and you have a semantically recognized, computer-readable, standardized way to access Magnetometer data

Enhancements (continued)

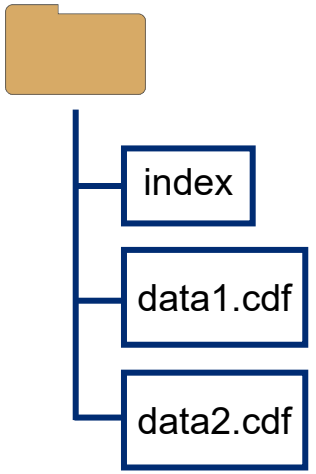
8. ways to analyze local HAPI data or (eventually) packaged HAPI data
   - a “pop-up” server is being developed (Sandy Antunes) for accessing your own data via HAPI
   - also will facilitate some data fusion – merging of multiple HAPI data sources



Using HAPI to capture research data outputs: seeking interoperability for "long tail" data

Consider:

- define a data packing arrangement for research data products (zip or tar) with specifically formatted contents – like "containerized data"
- put this data package on Zenodo
- have a HAPI server that can interpret these packages
- maybe cache the data packages, maybe run in HelioCloud



Also – use HAPI as a dynamic result of a more complex query

- HAPI only allows query by time range
- many data systems have more complex queries, but then the result is a time series table
- have the **result** be made available as a "pop-up" HAPI server with just the results of the query

use HAPI as a dynamic result of a more complex query

- HAPI servers only allow query by time range
- many data systems have more complex queries, but then the result is a time series table
- have the result be made available as a “pop-up” HAPI server with just the results of the query

Next Datasets for HAPI

Higher fruit:

large model output data
NCAR data (atmospheric data)

Medium Fruit:

smallish model output data
astronomy time-domain data
Earth science data - streamflow, earthquakes, gridded model output

Low Hanging Fruit:

HamSCI data – in discussion
Madrigal - interested

Thoughts / recommendations from the HAPI dev team

- when designing your data, keep it simple => seek advice / feedback outside your team
- when designing your visualization tools, use one layer of abstraction for representing data
  - Autoplot, TopCat, SpacePy and many others do this

More controversial:

- could we use HAPI metadata to generate SPASE records?
  (this has not been tried yet – there would be difficulties!!)
  - **might be the only way to reach non-NASA, non-Helio data (other fields, other countries, NSF, etc)**
  - universal: do for one HAPI server, then it works for all of them
  - robust: hard link to the data -- no human editing
  - automated: can regenerate new SPASE if data changes

Connect with the HAPI team

HAPI telecons are open; every Monday at noon, Eastern US time

`hapi-dev@groups.io`

Talk to us if you:

- prefer guidance on the initial part of the learning curve for using or implementing HAPI
- want to serve HAPI data
- need HAPI access in your own (client) code
- have input on any of our planned enhancements!

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY