

Interoperable Heliophysics Data Access via HAPI

Jon Vandegriff (JHU/APL)
Robert Weigel (George Mason University)
Jeremy Faden (Cottage Systems)
Sandy Antunes (JHU/APL)
Robert Candey (GSFC)
Eelco Doornbos (KNMI)

Outline

- What is HAPI (1 min version)
- Latest updates for version the HAPI 3.3 Specification
- Usefulness of HAPI in the Science Data System landscape
- HAPI and Heliophysics data standards in general
- What new capabilities can / should be built on top of HAPI?

What is HAPI? *In 1 minute.*

HAPI is:

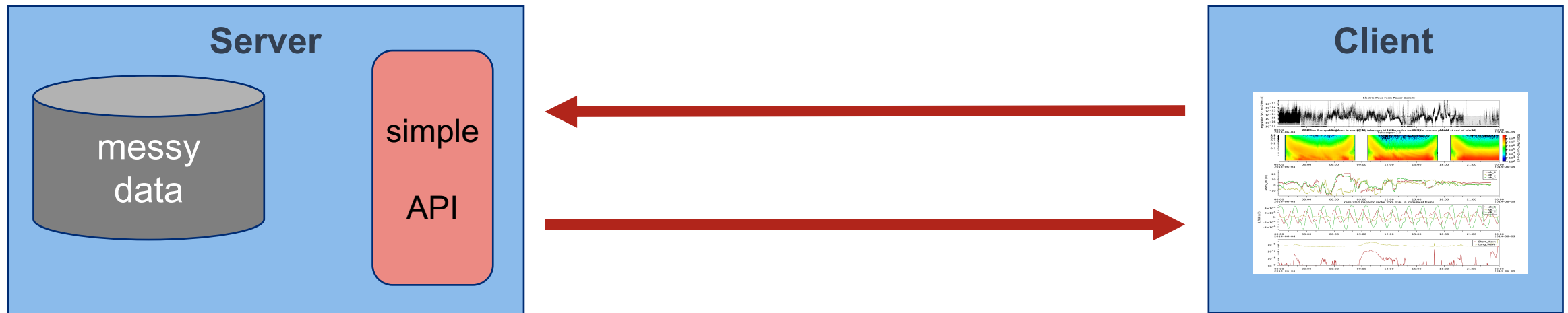
a data access specification for time-series data



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

HAPI is:

a way to hide data storage details behind a simple access API



HAPI is:

a widely adopted standard in Heliophysics *

* Heliophysics in the expansive sense used by DASH:

- ground-based measurements (mag, radars, aurora)
- upper atmosphere
- ionosphere
- magnetosphere
- the sun
- heliosphere
- planetary plasma environments
- space weather
- modeling of any of the above

HAPI is:

a widely adopted standard in **Heliophysics**

HAPI is:

a widely adopted standard in Heliophysics

- official COSPAR standard for space weather data
- in use at multiple Heliophysics institutions
- now being adopted at Madrigal

HAPI is:

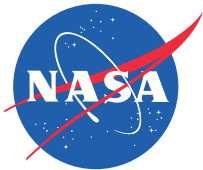
a widely adopted standard in Heliophysics

- official COSPAR standard for space weather data
- in use at multiple Heliophysics institutions
- now being adopted at Madrigal

HAPI is:



a widely adopted standard in Heliophysics



- official COSPAR standard for space weather data
- in use at multiple Heliophysics institutions
- now being adopted at Madrigal



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat



INTERMAGNET



British
Geological
Survey



HAPI is:

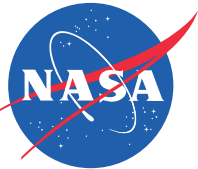


UCLA



JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

a widely adopted standard in Heliophysics



- official COSPAR standard for space weather data access
- in use at multiple Heliophysics institutions
- now being adopted at Madrigal



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat



INTERMAGNET



British
Geological
Survey

| Data Center | # HAPI Datasets |
|--------------------|------------------------|
| CDAWeb | 3532 |
| SSCWeb | 301 |
| AMDA | 1044 |
| CSA | 1989 |
| CCMC | 300 |
| VirES for SWARM | 162 |
| Intermagnet | 3025 |
| SuperMAG | 3270 |
| WDC | 4889 |
| LISIRD | 29 |
| KNMI | 934 |
| Helioviewer | 79 |
| Das2 | 19 |
| UCLA | 426 |
| Total | 19999 |

| Data Center | # HAPI Datasets |
|-----------------|-----------------|
| CDAWeb | 3532 |
| SSCWeb | 301 |
| AMDA | 1044 |
| CSA | 1989 |
| CCMC | 300 |
| VirES for SWARM | 162 |
| Intermagnet | 3025 |
| SuperMAG | 3270 |
| WDC | 4889 |
| LISIRD | 29 |
| KNMI | 934 |
| Helioviewer | 79 |
| Das2 | 19 |
| UCLA | 426 |
| Total | 19999 |

At AGU 2024,
the count was
10,846

HAPI 3.3 Features

HAPI 3.3 – fully backwards compatible with 3.2

- added 2 optional attributes to /info response about where measurements were made
 - **location** and **geoLocation**
- **altitude** quantity now allowed as vector component types (common for geo-location)

To Support Open Science:

- now have distinct **serverCitation** in about endpoint and **datasetCitation** in info/endpoint, and plain **citation** is deprecated
- added **resourceID** in /about and **provenance** and **licenseURL** in /info so that HAPI can describe FAIR data
- added a section and appendix on how FAIR principles map to HAPI

Units

- added new units schema for VOUnits to enumerated list of allowed schemas

HAPI 3.3 – fully backwards compatible with 3.2

- added 2 optional attributes to /info response about where measurements were made
 - **location** and **geoLocation**
- **altitude** quantity now allowed as vector component types (common for geo-location)

To Support Open Science:

- now have distinct **serverCitation** in about endpoint and **datasetCitation** in info/endpoint, and plain **citation** is deprecated
- added **resourceID** in /about and **provenance** and **licenseURL** in /info so that HAPI can describe FAIR data
- added a section and appendix on how FAIR principles map to HAPI

Units

- added new units schema for VOUnits to enumerated list of allowed schemas

Minor fixes / clarifications in the 3.2 HAPI spec

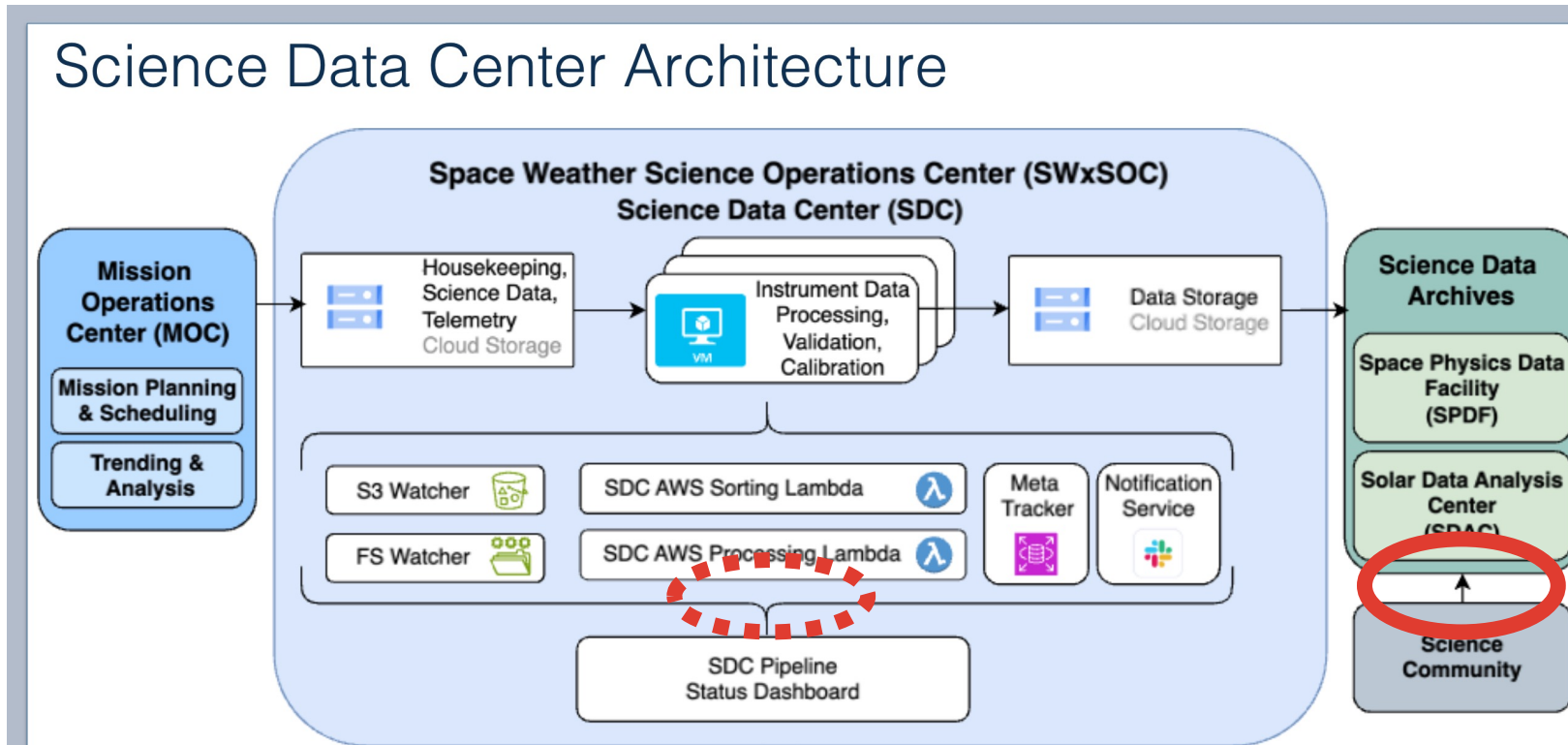
- clarified how extra trailing slash should be handled ([#248]#248)
- clarified how to name a custom (i.e., non-standard) endpoint (#245)
- clarified how scalar parameters can also contain vector components (#244)
- clarified requirements for which of bin centers and ranges need to be present (#237)
- clarified that HAPI is RESTful rather than strictly based on the original REST concept (#236)
- clarified that any non-standard data format in the **outputFormats** of the /capabilities endpoint needs to begin with **x_** (#222)
- clarified the difference between **title** (a short label) in /catalog endpoint versus the **description** (few lines of details) in /info endpoint (#221)
- clarified expectations for **id** and **title** in /about endpoint (acronyms ok in **id**, but expand in **title**; don't include word HAPI) (#219)
- fixed some typos and inconsistencies (#241)
- rearranged info section for clarity (#247)

Minor fixes / clarifications in the 3.2 HAPI spec

- clarified how extra trailing slash should be handled ([#248]#248)
- clarified how to name a custom (i.e., non-standard) endpoint (#245)
- clarified how s (#244)
- clarified require (#237)
- clarified that H REST concept (#236)
- clarified that any non-standard data format in the **outputFormats** of the capabilities endpoint needs to begin with **x_** (#222)
- clarified the difference between **title** (a short label) in /catalog endpoint versus the **description** (few lines of details) in /info endpoint (#221)
- clarified expectations for **id** and **title** in /about endpoint (acronyms ok in **id**, but expand in **title**; don't include word HAPI) (#219)
- fixed some typos and inconsistencies (#241)
- rearranged info section for clarity (#247)

Yes, we are still actively maintaining the spec and responding to questions.

Usefulness of HAPI in the Science Data System Landscape



HAPI can be used
for science access
to the data

Robbertz, et al, Fall AGU 2024

<https://zenodo.org/records/14606063>

HAPI also encourages you to standardize your data

- don't do anything weird with your time values
- use good variable names
- include descriptions

- represent vector quantities (magnetic field) as arrays, not sets of scalars
- indicate what the FILL value is

HAPI and Heliophysics data standards in general

To make data fully machine interpretable, **HAPI needs other standards that don't exist yet!**

- Units
- Coordinate Systems / Frames
- *maybe Provenance (these exist but are complex)*

What new capabilities can / should be built on top of HAPI?

- HAPI to SPASE adapter (see Madrigal talk later today!)
 - convert HAPI metadata to a SPASE record for a dataset
 - This could give you guaranteed-correct SPASE that can be automatically kept in sync with the data!
- HAPI Data Amalgamator
 - combine, fuse, average, interpolate any HAPI data content from any server
 - this would be hugely useful
 - someone should propose to make this – it is a very fundable idea!!
- HAPI for Zenodo – regularized access to all the “stuff” people jam into Zenodo (or similar)
 - has been proposed once and did not win – talk to me if you want to try this!

Catch this talk
later today!!

Disciplinary Improvements: Accelerating Geoscience Discovery with Madrigal: HAPI+SPASE Integration for Transformative Open Science

**K. Cariglia, P. Erickson, L. Goncharenko, W. Rideout,
E. Rojas Villalba,**

MIT Haystack Observatory



DASH/IHDEA Workshop, October 20 - 24 2025

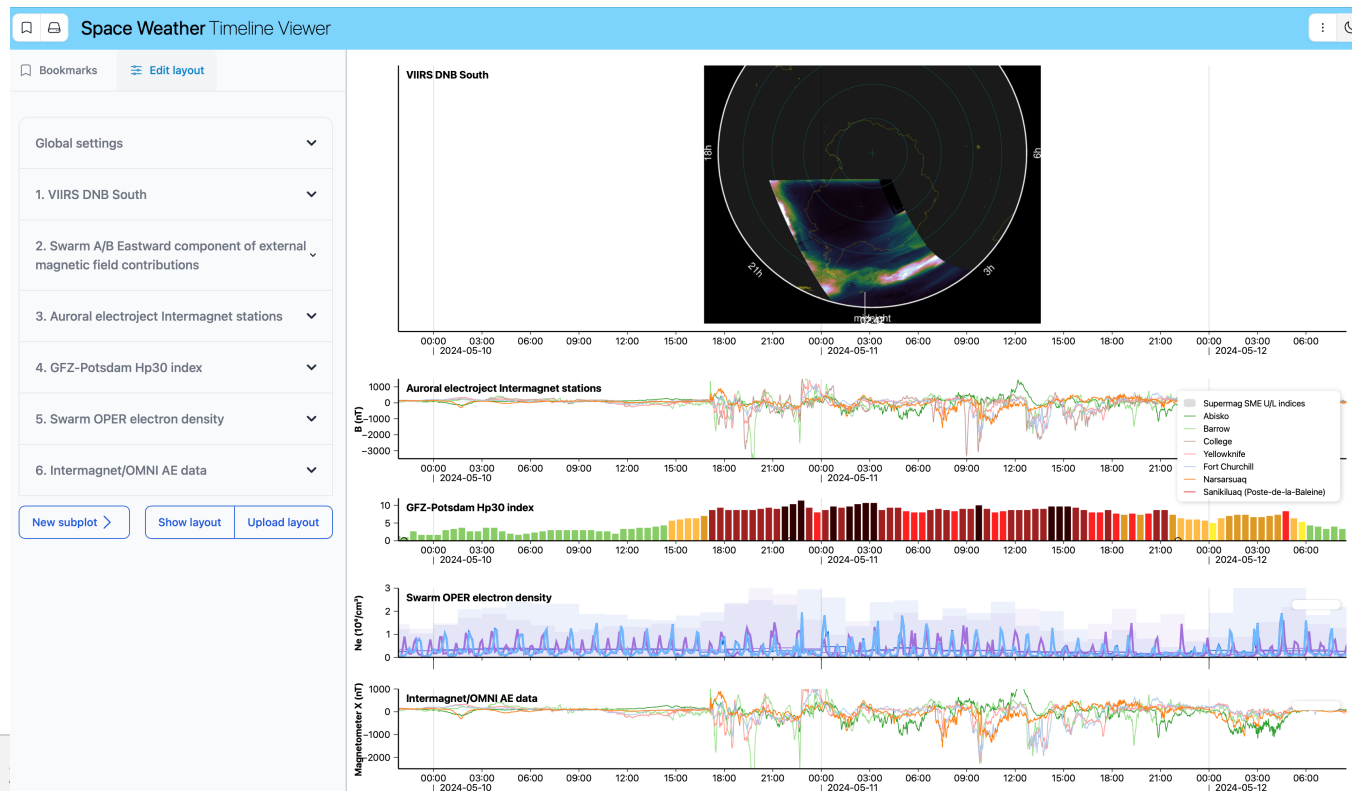


What new capabilities should be built on top of HAPI? (part 2)

- HAPI as a way to regularize other time series lists
 - file listings
 - good time intervals (a list of “good data is present” time windows)
 - event lists
- Model-data comparison engine
 - some elements exist – generalization would be hard
- Semantically knowable data structures for measurement types
 - what if every magnetometer dataset you accessed came into the same data structure in your code?

What new capabilities should be built on top of HAPI? (part 3)

- Machine-readable, scientifically useful connections between datasets, i.e., easily use / compare / combine datasets related by:
 - cadence (same measurements, but different averaging intervals)
 - coordinate frame (data transformed to other coord sys)
 - background subtracted version (or other calibration type)
 - mirror or copy of the data



Timeline Viewer at KNMI is client that auto-switches to correct cadence dataset based on time window size in current view.

We are generalizing this clever addition!

Connecting related datasets *(feature still in development!)*

- feature was added by a HAPI user (KNMI development team) using HAPI extension capability
- HAPI team working to make this a generic feature for relationships between data:
 - datasetA **isResampledTo** datasetB
 - C **isCoordTransformedTo** D
 - E **isCopyOf** F
 - G **hasFileListingAt** H
- The nouns and verbs in these triples will be derived from known vocabularies
 - probably DCAT or similar
 - example for the noun “dataset”: https://www.w3.org/TR/vocab-dcat-3/#Property:catalog_dataset

Takeaways

- HAPI adoption is growing and is helping standardize access
- capabilities built on HAPI will enhance multi-source analysis
- HAPI is being enhanced to communicate machine exploitable connections between data

Join us!

Subscribe: hapi-dev+subscribe@groups.io

Ask, learn or demonstrate at our weekly telecons, Monday, noon US Eastern.



JOHNS HOPKINS

APPLIED PHYSICS LABORATORY