



THE HDF COMMUNITY

State of the Discussion

METAPHOR

Radar

Inference based on uncertain information

Senders

- HDF5 releases
- Community project releases
- Outreach activities
-

Receivers

- 20,600 citations on Google Scholar (all time)
 - 957 (year to date)
- HDF forum
- Pull requests, bug reports
- GitHub projects
- ...



THE HDF GROUP

What

We are here to make large complex data manageable and efficiently accessible, anywhere and indefinitely.

Why

We are doing that because we want to help communities of researchers advance science and engineering and enable companies and organizations to make better data-driven decisions faster.

How

We sustain ourselves by maintaining high-quality Open Source Software, adding value to a community software ecosystem, complemented by support, training, development, and consulting services in order to raise funds to invest in ideas, do research, and spark innovation.

MYTHOLOGY

- FREE Open-Source Software doesn't cost me anything
 - 'Free' at the point of access
- The HDF Group receives external funding for HDF5 library maintenance (sustaining engineering)
- HDF maintainers have access to an large pool of testing resources
- HDF Group employees are paid competitive salaries and tend to retire early (because of their compensation)
- HDF developers are secretive, don't like to share knowledge, sluggish to reply on the forum
- Everyone at The HDF Group is super-efficient at what they are doing, and sitting idle most of the time
- HDFG doesn't like pull requests
 - It's an easy fix, but they don't get it
- The HDF5 library is so large because of the complex file format specification...
- ...



SIGNALS

How can we help?

Language bindings & modules

- API-less HDF5?
- Can we eliminate I/O as a concern for most users?

Tools

- Web-based viewers (h5web, h5wasm)
- VSCode plugin
- The file format is not going away anytime soon
 - Forensics & recovery
- Time to architect our command line tools?

Standards

- FAIRness

Money



HDF ANYWHERE

Edge, Mobile, Server, HPC, Cloud

- Keeping HDF relevant
- Cohesive appearance
- HDF "Turing test"
- Scales
- Virtual Object Layer (VOL)
- The twilight of the general purpose processor(?)
 - Better support for accelerators, smart storage, in-memory processing, ...
- HDF5-X testbed (?)
 - Sorted key/value store
 - Concurrency
 - Asynchrony
 - Atomicity
 - Closure
 - Streaming

```
import h5pyd as h5py

# no changes below this line
```



everything
is
connected

HDF INDEFINITELY

FAIRness, Archivability

- HDF5 is only conditionally suitable for archival
- HDF/A
 - Bit-level reproducibility in reference environment
 - GNU Guix channel
 - Machine-digestible specification
 - Machine-learned specification
 - Auto-generated HDF5 SerDe
 - Compatibility/risk levels
- The power of linkage
- HDF5 automation
 - Package/plugin management
 - Flathub
- What is FAIR HDF5?



WHAT ARE WE MISSING?

- Missing bands
- Stealth objects / UFOs
- Jammers
- Information flow between radar sites (community projects)
 - Collaborative system