

HDF5 Community Building

June 1, 2022



Dana Robinson
The HDF Group

Whither HDF5 ?



How do we make HDF5 a sustainable piece of software?

1. More maintainable code

- Reduce
- Refactor
- Simplify
- Document
- Automate

2. More community participation

- **Tune in tomorrow!**

Focus of this talk



The HDF5 Library

- Also applies to other core HDF Group software, like HDFView, hsd, etc.
- Not about "community" in the sense of all the supporting software that we've seen this week.
- As far as I can tell, that part of "community" is going great!

Agenda



- Escaping the walled garden
- Transparency
- GitHub
- Documentation
- External Contributors
- Communication with key downstream software
- Challenges

Escaping the walled garden



- HDF5 has always been open source
- Also historically been a walled garden run by benign dictators
- But this is not how we want to run things in the future!
- Going to spend effort ensuring we are more open to outside participation
- Why?
 - Limited funds at the HDF Group
 - External input will help HDF5 meet community needs
 - Will help ensure HDF5 lives on even if the HDF Group closes up shop

Transparency



- We want HDF5 development to be much more transparent
- For everybody, not just people who are in frequent contact with us
- Release schedule on GitHub and kept up-to-date
- More frequent blogging
- Public performance metrics
- Website improvements
- Documentation

GitHub



- Good move for the company – huge success!
- Reduce number of branches
 - Too many maintenance branches
 - Too many branches, in general
- Overhaul PR checks
- Pay more attention to GitHub issues
- Quicker responses to pull requests

Documentation



- THREE sources of reference manual entries (old support site, portal, Doxygen)
- Other documentation scattered across the web, hdf5doc GitHub, etc.
- Goal: All user documentation in Doxygen (RM, UG, etc.)
- Architecture documentation
- Better guides for getting new people up to speed

External Contributors



- We want more people to contribute to HDF5 and in bigger ways
- Unfortunately, the barrier to entry for this is high
- We are going to work to reduce this barrier:
 - Library internals documentation
 - Better contributing guides
 - Reduce and simplify HDF5 quirks
 - Refactor complex code

Communication with key downstream software



- We have often had good communication with key HDF5-using infrastructure like h5py and netCDF
- Want to strengthen these relationships
- Help with bug fixes
- Ensure new development doesn't cause problems
- "Cross-pollination" – e.g., more HDF Group developers should know about h5py internals, contribute PRs for h5py, etc. (and v.v.)

Challenges (minor)

- Lots of legacy practices that will take some time and effort to change
- Finding effective, inexpensive, and open tools (e.g., code review)
- Unifying disparate data / resources
 - Scattered documentation (THREE sources of reference manual entries!)
 - Issues / bug reports scattered across JIRA and GitHub

Challenges (major)



- I don't think enthusiasm and cooperation are a big problem.
- The HDF5 ecosystem is an incredible community full of friendly, talented people who want to solve problems.
- The biggest problem, as always, is **resources**
- All the things described here – writing, meeting with people, coordinating, restructuring HDF5 supporting infrastructure to be more open – they all take time, money, and people.
- Without resources, all of this is just aspirational

Funding "Sustainable Engineering"

- No pile of money for this
- Most funding pays for features, not maintenance
- We basically get enough money to do releases, testing, and staff the help desk
- Addressing maintainability issues has largely been a labor of love by a small number of developers
- This has been a challenge for The HDF Group for many years



How can you help?



- Pitch in! Become an external contributor!
- Talk to us! – Feedback, ideas
- Funding: Support contracts, donations
- Alert us to funding opportunities that support open-source software
- Be patient! We really are trying to do our best to support you. We just have very limited resources.

THANK YOU!

Questions & Comments?