

[Community Conventions for Sharing and Understanding Data](#) [1]



Submitted by Ted.Habermann on Wed, 2014-05-28 20:52 Tuesday, July 8, 2014 - 09:00 to 10:30

Event: [Summer Meeting 2014](#) [2]

Session Type: [Workshop](#) [3]

Collaboration Area: [Information Technology and Interoperability](#) [4]

Abstract/Agenda:

HDF gives communities the opportunity to define conventions for organizing and describing their data in order to facilitate sharing and understanding. Earth Science communities use several conventions for facilitating discovery (ACDD) and use (HDF-EOS, NUG, CF) of data in HDF. In addition, many groups are in the process of adopting ISO standards for metadata and data quality. Multiple conventions are needed to support different use cases and phases of the data life-cycle. We will begin with discussions of several approaches to multi-conventional datasets and finish with a discussion of external metadata augmentation techniques as an approach to supporting various conventions.

Notes:

The HDF Workshop Part 1 (of 4)

- HDF presentations have been put up on Slideshare and have had over 27,000 views.
- Presentations for this year's workshop will also be on Slideshare

Community Conventions for Sharing and Understanding Data - Habermann

- Groups take HDF and put things on top & then call it something else - like NetCDF - this is the theme of the first session

Ajay Krishnan (NODC - NOAA) - The National Oceanographic Data Center's NetCDF Templates

- Conventions - ACDD (Attribute Conventions for Dataset Discovery) and CF (Climate and Forecast)
- Use a flow chart (decision tree) to use the right template and use a guidance table as to how to populate template
- NODC added attributes includes their own list of standard names
- o Recommend make & model of instrument, ...
- o Make new uuid when creating products or making changes for analysis purposes
- Tables - people, products, institutions, and managers are maintained by NODC
- Unidata attribute conventions for data discovery reports
- o Simple changes can make a big difference in the scores

- Goals/benefits
 - o When have good metadata are able to good products from it
- Key principles
 - o Be explicit (ex. N/A can be not applicable or not available)
 - o For more information & non file specific - use ISO metadata
 - o Not enforcing a new convention
- Q - Aleksander - swath is not an official CF feature type
 - o Ken - no but we use it anyway
- Q - John - are you taking the existing NetCDF archive and build a new archive with the templates
 - o This is specifically for new data sets - will be applied backwards based on funding
- Q - Chris (NASA) - are you planning or exposing the authority tables - looking at machine accessible times
 - o Going to expose via scoft (sp?) & some is available via webservices
 - o Ken Casey - moving away from NODC vocabulary tables - moving towards international standards (ex. ISO country code)... if there is an international or larger table (ex. CF standard names)
- Ted - "these are the benefits" - without conventions, then may not be able to use in the future - expressing the benefits which is a big thing
 - o The rubric is a learning tool in addition to a compliance checker
 - o NODC templates combine many different templates - ACDD, CF, NetCDF and more

Mike Folk (HDF group)- Augment: Add to, amplify, extend, enhance (

- Augmentation was an original HDF design philosophy
- Why we augment
 - o Intended purpose better or even a new purpose
 - o Preservation - adding provenance or adding new data to support other things
- HDF augmentation
 - o Preserve the original use but adding extra information (ex. Rename - new version with a different name)
- 3 examples
 - 1) Augmenting HDF-EOS5 files for NetCDF-4 access
 - o HDF-EOS5 uses HDF5 as it storage layer

- o Adds EOS structural and content metadata
- o NetCDF4 uses HDF ALSO as its storage layer - adds structural and content metadata
- o Solution was to create NetCDF4 structural data and insert it in a Eos5 file without causing problems for EOS or NetCDF4
- 2) H5AUGJPSS
- o IDV - based on netCDF model (no groups)
- o Make JPSS HDF5 file look like netCDF-4 - hid HDF5 element not in netCDF, add coordinate variables and removes from view groups
- 3) HDF4 File Contents - User View
- o When a user opens a NetCDF file - expect people to look for specific objects and relationships (or metadata) & object data
- o Uses the HDF4 library to view their data - idea of opening a file without using the HDF4 library (difficult to do)
- o Question that Chris raised was would the library be around in the future - what to do if the library isn't there
- o Create a separate file - don't change data file, but create a file with all the data needed for the data - in a simple tool format
- o HDF4 mapping workflow - created a reader
- o Tested it with students - one student wrote application in python to read the data
- Q - has some tried to create an intermediate metadata format for other programs to read
- o It is on the to do list - but need funding
- o Written a tool that analyzes the HDF5 file
- o Map files can be used for a lot of other things than just writing a reader and learning
- o People who participated on early project jump on board the reader
- Ted - important uses of augmentation - such as a map -with mapping tool - able to extract data from 400-600 products and then use it to characterize the files
- o Hooked up the mapper to THREDDs and used the map ... this will continue in HDF and Services

Kent Yang (HDF Group) - Make CERES products user-friendly

- CERES upgrades their HDF4 to HDF5
- A good model to work with science teams
- o This will help solve a lot of problems - if fix the original data
- Ted - similar to work by NODC to work with data providers to improve the convention within

Community Conventions for Sharing and Understanding Data

Published on Commons (<http://commons.esipfed.org>)

their files – also for other uses – this is called the 1st mile which makes things easier in the 2nd mile

- Q – Robert – when first looked at CF convention – they are good for modeling data but not for observational data – how have speakers overcome this issue or has the CF root taken on the challenge to add key words (etc) for observational data

- o Ken – CF1.6 made huge step forward – mainly working to simplify the conventions

§ Getting lots of great feedback – gone a long way

- o Jeff Lee – to substitute engineering for observation – NASA missions have engineering – and CF doesn't have terms for this

- o Ted – CF came from level 3 or 4 (low resolution) data sets – usually global- but usually think of discrete samples – also have processing level 4 (models) to 3 (grids) to level 1 (from satellite) and 2 (some processing)... trying to extended processing levels down to level 1 and 2, which is very difficult

- o Documentation cluster tried to add names to CF – adding names related to radiances (and others) was a difficult task – many people have a lot of problems with CF

- o If look at it from CF statement – their focus is on the climate and forecast community (and it is working very well)

- o Adding complexity to CF is unlikely

- o Process of ESWG (Earth Science Working Group) – NASA working group – finding best elements of.... CF, ACDD, EOS And bring together into a new set of convention and supports the how data lifecycle

Hyo-Kyung Joe Lee (HDF Group) – How to Meet the CF Conventions with McML for NASA HDF/HDF-EOS

- Focus on few key conventions

- o Coordinate variable & attributes (blad)

- o Valid_range/_fillValue (fat)

- o Scale_factor/add_offset (short)

- o Units (\$ and cents)

- If your data product doesn't follow these conventions – then it is useless (like IDV) (i.e. it is not your type)

- But if you follow the conventions – then can see the data instantly and correctly

- OBPG group Sample data

- o IDV couldn't visualize it

- Why McML has more features and flexibility – also works with THREDDS

- What worked with IDV doesn't work with GoDiva2

- o Is CF/McML ready for “group”

Community Conventions for Sharing and Understanding Data

Published on Commons (<http://commons.esipfed.org>)

- Q - can't change dimension names in McML... dimension name/variable name can't change it
- o New CF convention doesn't
- o Still need to link the 2 - dimension name or coordinate attribute
- Ted - good example of need to see who things actually work
- o Mentioned the problem of getting groups into CF or getting CF to read files with groups
- o The non-invasive McML approach is the historic approach - make non-CF datasets look like CF
- o Use McML can create a flattened version without groups and tools that use groups can read the original file

Session Leads:

Name: [Ted Habermann](#) [5]
Organization(s): [The HDF Group](#) [6]
Email: thabermann@hdfgroup.org [7]

Presenters:

Name: [Ajay Krishnan](#) [8]
Organization(s): [NOAA National Oceanographic Data Center](#) [9]
Email: ajay.krishnan@noaa.gov [10]

Name: [Mike Folk](#) [11]
Organization(s): [The HDF Group](#) [6]
Email: mfolk@hdfgroup.org [12]

Name: [Kent Yang](#) [13]
Organization(s): [The HDF Group](#) [6]
Email: myang6@hdfgroup.org [14]

Name: [Joe Lee](#) [15]
Organization(s): [HDF](#) [16]

Notes takers:

Name: [Joel Plutchak](#) [17]
Organization(s): [The HDF Group](#) [6]
Email: plutchak@hdfgroup.org [18]

Name: [Kelly Monteleone](#) [19]
Organization(s): [University of New Mexico](#) [20]
Email: krbm@unm.edu [21]

Participants:

Jennifer Davis, Matt Cechini, Ellen Johnson, Nick Haddadk, Matt Tisdale, Ted Habermann, Ajay Krishnan, Gary Bers-cross, matt comerford, andrew collets, jeff lee, joel plutchak, john relph, james howison, ed johnson, minnie wong, lizzy mcmichael, joe lee, karjaan xei, Ted Hart, Dan Kowal, vicki ferrini, john morton, suresn vannan, amadan ross, ana pinhero preivette, nina jackson, doug fils, chris lynnes, charles thompson, zhenlogn Li, mike McGann, Karen Moe, Robert Wolfz, Navid Golpayegani, gao chen, john kusterer, lindsay powers, joe glassy, elmain Martinez, aleksandar jelenak, nawajish nanan, Anna Katz, Ross Bagwell, Steve oldign, stephan klene, gilberto vicente, mike folk, ken casey, thomas harris, Carl Recil, Yaxing Wei, Kelly Monteleone

Creative Common License: Creative Commons Attribution 3.0 License

Community Conventions for Sharing and Understanding Data

Published on Commons (<http://commons.esipfed.org>)

Teaser: There is no silver bullet in convention. HDF gives communities the opportunity to define conventions for organizing & describing their data.

Accepted:

Keywords: [HDF](#) [22]

Source URL: <http://commons.esipfed.org/node/2471>

Links:

- [1] <http://commons.esipfed.org/node/2471>
- [2] <http://commons.esipfed.org/2014SummerMeeting>
- [3] <http://commons.esipfed.org/session-type/workshop>
- [4] <http://commons.esipfed.org/collaboration-area/information-technology-and-interoperability>
- [5] <http://commons.esipfed.org/node/1847>
- [6] <http://commons.esipfed.org/taxonomy/term/822>
- [7] <mailto:thabermann@hdfgroup.org>
- [8] <http://commons.esipfed.org/node/2510>
- [9] <http://commons.esipfed.org/taxonomy/term/1160>
- [10] <mailto:ajay.krishnan@noaa.gov>
- [11] <http://commons.esipfed.org/node/1679>
- [12] <mailto:mfolk@hdfgroup.org>
- [13] <http://commons.esipfed.org/node/2511>
- [14] <mailto:myang6@hdfgroup.org>
- [15] <http://commons.esipfed.org/node/465>
- [16] <http://commons.esipfed.org/taxonomy/term/315>
- [17] <http://commons.esipfed.org/node/1524>
- [18] <mailto:plutchak@hdfgroup.org>
- [19] <http://commons.esipfed.org/node/430>
- [20] <http://commons.esipfed.org/taxonomy/term/297>
- [21] <mailto:krbm@unm.edu>
- [22] <http://commons.esipfed.org/taxonomy/term/316>