

# Notes for Discrete Sampling Geometries Breakout

## Participants

Tim Whiteaker, Julien Chastang (Unidata), Bob Simons, Aleksandar Jelenak, Jessica Hausman, Kelsey Druken, Kevin O'Brien

## Use cases

Argo float - A ship drops a float which drifts with currents and measures parameters at various depths. Folks might request by buoy number, time slice, spatial coordinates

Glider - Like a float, but path is controlled remotely

Swath - NCEI has swath template, but there isn't a standard feature type for swaths.

Ajay - Time aggregate CTD casts, serve with THREDDS with netCDF subsetting service

Rich - Sensor measures at high frequency (burst), summarized/event data computed for users (see below)

## Solutions open to us via netCDF enhanced model

Groups, e.g., make a different group for each sample location with a uniquely sized temperature dimension in each group

VLEN (variable length arrays), e.g., make a vlen temperature dimension

Conclusions from the DAP discussion:

Should CF be involved in specifying a query and delivery system?

DSG can be accessed via cdmremote, but what about more generic simple flat, tables based (SQL-like) query API regardless of underlying data structure. Declarative language - specify what you want, not how to do it. It could greatly simplify requesting subsets of data from a wide variety of underlying data structures.

## Action Items

Sean willing to create a sample dataset, represent it using groups and again using VLEN, for >1,000,000 array axis size, and test performance. **Does anyone have a sample dataset they'd like him to use?**

Question was posed: how to store "burst" data (aka "events" -- high resolution time sampling embedded into a regularly-spaced low resolution time series)?

Could the bursts not be fully represented using the CF TimeSeries feature type with contiguous ragged array representation? The feature dimension would be used to dimension a time variable rather than the Lon and Lat variables of a typical TimeSeries DSG feature collection.

Request (Aleksandar Jelenak and Tim Whiteaker): Amend the current timeseries feature type to support time series that instead of a geo-location provide an identifier.

Use cases:

- satellite calibration coefficients time series (identifier: satellite (platform) and its instrument).
- A single streamflow value for all rivers in the conterminous U.S. at a given point in time.

- [Example file without geometry](#). Line geometry can be acquired from NHDPlus, or use script linked below. Probably want to subset for a small region, as there are 2.7 million river features.
- [Details about the data](#). Includes link to script to add lat/lon at upstream end of each stream segment to file, making it CF-friendly.

There is a need for tools and conventions that help the community of spreadsheet-using scientists to become part of our CF community. With a set of simple conventions describing how column headers (incl units) are specified, and how global attributes can be represented in a CSV header, tools that bi-directionally convert between CSV and netCDF-CF DSG (with no information loss) can easily be built.

Some candidate tools already exist: [Rosetta](#), and [Panoply](#)