

# Compliance Checker

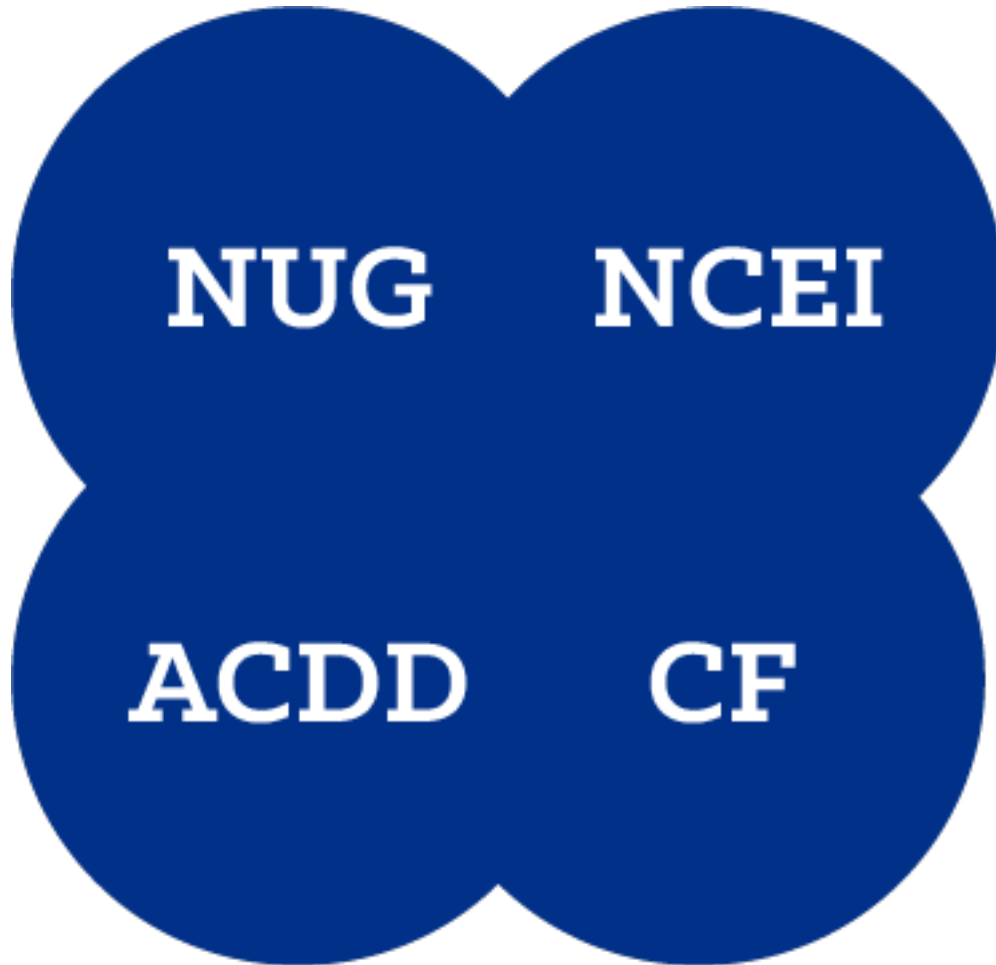
Presented by Luke Campbell



# Background

The standards in practice within IOOS and Communities

- [The netCDF User's Guide \(NUG\)](#)
- [Cooperative Ocean/Atmosphere Research Data Service \(COARDS\)](#)
- [netCDF Attribute Convention for Dataset Discover \(ACDD\)](#)
- [NetCDF Climate and Forecast \(CF\) Metadata Conventions](#)
- [NCEI NetCDF Templates](#)



- Many Standards
- Each overlap in some degree
- They don't always intersect nicely

# Background

The goal of the IOOS Compliance Checker is to provide data providers with confidence that their data is compliant and consistent with the community.

**GDAC**

**UGrid**

**CF**

**ACDD**

**NCEI**

# Compliance Checker: Core Standards

# Compliance Checker: Core Standards

Original Goal was for ACDD attributes.

- CF 1.6
- ACDD 1.1
- ACDD 1.3

From ACDD we expanded into CF and IOOS Vocabulary Metadata

Our latest release has ACDD 1.3 and support for plugins

# Compliance Checker: Core Standards

## Corrective Actions

Name	Priority	Corrective action
Conventions	3	Attr Conventions does not contain 'ACDD-1.3'
acknowledgment/acknowledgement	3	Neither 'acknowledgment' nor 'acknowledgement' attributes present
creator_email	2	Attr creator_email is empty or completely whitespace
creator_institution	1	Attr creator_institution not present
creator_name	2	Attr creator_name is empty or completely whitespace
creator_type	1	Attr creator_type not present
date_metadata_modified	1	Attr date_metadata_modified not present



# Plugin-based Framework

# Plugin-based Framework

Introducing a plugin-based framework



- Provides support for more checkers without overwhelming the main project
- Allows groups and developers to build their own compliance checkers based on the plugin framework

# Plugin-based Framework

- Our first plugin that we published online was the GliderDAC checker
- Why we needed a checker just for GliderDAC
- Integrating it with the web-checker

# Plugin-based Framework

- More plugins on the way:
  - A plugin for the NCEI Templates. Version 1.1 checker is almost ready
  - NCEI Templates checker 2.0 will follow shortly
  - A plugin for s-grid started in February

# Plugin-based Framework

- The frameworks approach:
  - Given a dataset, return a result
  - Each result has a severity, LOW, MED, HIGH
  - Each result has a score and an "out of" value
  - Each result without a perfect score has help messages for how to fix
- This approach enables virtually any kind of metadata checking for netCDF datasets.

More than a command line tool

# More than a command line tool

```
1. tmux (tmux)
-----
Reasoning for the failed tests given below:
-----
Name                Priority:  Score:Reasoning
-----
Coordinate Variables  :3:  15/16 :
  z has attribute ancillary_variables:2:  0/ 1 : z is missing attribute
                                           ancillary_variables
QA/QC Variables      :3:  14/16 :
  temperature_qc      :3:  7/ 8 :
  flag_attributes_lengths :3:  0/ 1 : the length of
                                           flag_mask/values does not
                                           match flag_meanings for
                                           temperature_qc
  temperature_spike_qc :3:  7/ 8 :
  flag_attributes_lengths :3:  0/ 1 : the length of
                                           flag_mask/values does not
                                           match flag_meanings for
                                           temperature_spike_qc
Required Variables    :1:  0/ 1 :
  timeSeries          :1:  0/ 1 :
  exists              :1:  0/ 1 : timeSeries does not exist.
                                           This is okay if there is
                                           only one Time Series in
                                           the dataset.
-----
(ccweb) → data git:(changes) |
git log --oneline -- 15. 1: .. el/tests/data* 6 days Tue 1:35:18 PM 2016-06-14
```

After a year of development we saw the immediate limitations of having something that only works as a command line tool:

- Challenging to install
- Output was hard to use in other programs
- Corrective actions messages weren't intuitive
  - "What do you mean "cf\_role exists!?"

# More than a command line tool

## Improvements as a library

- Switching to Anaconda
- Windows support and UDUnitspy
- Unicode Troubles
- Improving Messages
- Adding new output capabilities:
  - JSON
  - HTML
  - Plaintext



# More than a command line tool

- Creating a web interface
- Online web checker



Compliance Checker

## IOOS Compliance Checker

Select the test you want to run

GliderDAC Compliance Check ▾

Upload your dataset here: (Max Size: 16.0MB)

Drop files here to upload

Browse

Or provide a remote OPeNDAP Data URL here

Enter URL for dataset

Submit

# Online Web Checker

- No python required
- Just an internet connection and a dataset
- Supports uploads and links to OPeNDAP
- Reports are in a printable format

## Compliance Checker

Your dataset scored 123 out of 132 points

During the ncei-profile-orthogonal check

For dataset multi\_station\_timeseries.nc

### Scoring Breakdown:

#### High Priority

Name	Score
Coordinate Variables	15 / 16
Dataset contains NCEI profile required and highly recommended attributes	2 / 4
Dataset contains required time dimensions	0 / 2
QA/QC Variables	14 / 16
Required global attributes	6 / 6
Science Variables	0 / 1
latitude required attributes	3 / 3
longitude required attributes	3 / 3
temperature required attributes	3 / 3

#### Medium Priority

Name	Score
Container variable for storing grid_mapping information	4 / 4
Recommended Global Attributes	46 / 46
ctd_id is a proper instrument variable	2 / 2
latitude recommended attributes	5 / 5
longitude recommended attributes	5 / 5
platform is a proper platform variable	3 / 3
temperature recommended attributes	12 / 12

# Demonstration