Active working group members:
James Biard (NCEI/NOAA)
Jonathan Yu (CSIRO, Australia)
Mark Hedley (UK Met Office)
Adam Leadbetter (Marine Institute, Ireland)

Acknowledgements
Nick Car & Alex Ip (Geoscience Australia)
Kelsey Druken (NCI Australia)
Sean Arms (UCAR)
[ Early contributors ]
Great uptake & tools.
Many conventions.
Lots of data in netCDF!

Way to interlink, discover, and integrate data on the web.
Rapidly growing galaxy of information spanning many disciplines.
Linked Data profiles – JSON-LD, CSVW.

Recipe for constructing Linked Data from netCDF files.
Define conventions and validate data.
Enable linking to other resources.
Recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge *on the web.*

Standard format …
Reachable …
Relationships between data …

Collection of interrelated data → Linked Data

**Key concept:** Give each *thing* in the data an individual identity or URI

https://www.w3.org/standards/semanticweb/data
Linked Open Data Cloud  http://lod-cloud.net/

32 billion triples in 2014

... 192 billion+ triples in 2017
See http://stats.lod2.eu/stats

Can we plug netCDF/HDF data in?
Motivations

1. Interoperability
Handle multiple metadata standards/conventions
  e.g. combining CF with project conventions

2. Reuse and discovery
Leverage web resources and Linked Data tools to *enhance discovery*
across large collections of files (e.g. represent separate files as graphs)
Journey so far

2014-2015
Conceived during eReefs in discussions with Ethan Davis and Adam Leadbetter

Collaboration with Mark Hedley (UK Met.)

DOI: 10.1007/978-3-319-15994-2_9

EarthCube Advancing CF for geosciences workshop (Boulder, CO), May 2016

WG proposal endorsed by CF community

EU: UK Met office, Marine Inst. Ireland
AU: GA, NCI
US: NOAA, NCICS (notably Jim Biard)

Drafting OGC Standard netCDF-Classic-LD

OGC TC NZ

EC CF meeting Sept 2017

Fall meeting Poster Dec 2016

Piloting encoding rules, tools and tests on GitHub

netCDF-LD WG and telecons (Aug 2016)
Design principles

1. Work with current netCDF files

   Design a simple mechanism to that works with existing netCDF files as-is to encode in a Linked Data friendly format.

2. Allow consistent & precise naming of each *thing* in netCDF/HDF metadata

   Implies introducing new syntax (compatible with netCDF / HDF) to build URIs for each attribute name and property value

   Enable consistent way to link to references, e.g. model, instrument, etc.

3. Provide useful Linked Data enabled outputs (RDF, schema.org)
Overview

Tools

Syntax (aliasing, prefixes)

Supporting registries

Next steps
Tools

Python libraries (Github bald repo) - (bald = binary array linked data)  
https://github.com/binary-array-ld/bald

Command line tools (in development):

nclddump 

nc2rdf

Demos
nclddump

$ python nclddump.py example.cdl

CDL or netCDF/HDF file (reads metadata) → Hotlinked HTML styled ncdump output → Information on web pages and registries

Example on https://binary-array-ld.github.io/netcdf-ld/#examples
nc2rdf

$ python nc2rdf.py example.cdl

CDL or netCDF/HDF file (reads metadata)

(visualisation representation)

Load into triple store DB for semantic queries

Other data
Demo visualisations of graphs from CDL examples in bald repo

ncld example 1 :: BALD Containers and Arrays

http://waterinformatics-ext1-cdc.it.csiro.au/ncld-demo/
Binary Array LD Syntax (for netCDF and HDF)

Methods to encode or process nc/hdf for translating to RDF / Linked Data ready

**Aliasing**

Lookup table for ‘well-known’ or declared mappings

Can be explicit or implicit

**Pros:** Easy to convert current nc files

**Cons:** Resolving clashes

e.g. `title → acdd:title`

(netcdf) (RDF)

**Prefixing**

Kinda like namespacing

**Pros:** Easy to convert conformant files

**Cons:** Current files need tweaking

e.g. `acdd___title → acdd:title`

(netcdf) (RDF)
Binary Array Linked Data (BALD) model

http://binary-array-ld.net/_latest?classView=true
Aliasing example

variables:

```plaintext
int variable(pdim0, pdim1);
variable: `SDN_ParameterDiscoveryCode` = "BactTaxaAbundSed";

int cfvariable(pdim0, pdim1);
cfvariable: `standard_name` = "air_temperature";

// global attributes:
:isAliasedBy = "alias_list";
```
Aliasing example – adding context (explicit aliases)

variables:

    int alias_list;
    alias_list: SDN_ParameterDiscoveryCode = "http://vocab.nerc.ac.uk/isoCodelists/sdnCodelists/cdisrCodeList.xml#SDN_ParameterDiscoveryCode";
    alias_list: BactTaxaAbundSed = "http://vocab.nerc.ac.uk/collection/P02/current/BAUC/";
    alias_list: standard_name = "https://def.scitools.org.uk/CFTerms/standard_name";
    alias_list: air_temperature = "http://vocab.nerc.ac.uk/collection/P07/current/CFSN0023/";

int variable(pdim0, pdim1);
    variable: SDN_ParameterDiscoveryCode = "BactTaxaAbundSed";

int cfvariable(pdim0, pdim1);
    cfvariable: standard_name = "air_temperature";

// global attributes:
    :isAliasedBy = "alias_list";
Aliasing example – RDF representation

<example> a bald:Container ;
    bald:contains <variable>, <cfvariable> .
    ...

<variable> a bald:Array ;
    ns1:SDN_ParameterDiscoveryCode
        <http://vocab.nerc.ac.uk/collection/P02/current/BAUC/> ;

<cfvariable> a bald:Array ;
    ns2:standard_name
        <http://vocab.nerc.ac.uk/collection/P07/current/CFSN0023/> .
float eta(time, j, i) ;
    eta:units = "metre" ;
    eta:long_name = "Surface elevation" ;
    eta:standard_name = "sea_surface_height_above_sea_level" ;
    eta:medium_id = "ocean"
    eta:scaledQuantityKind_id = "sea_surface_elevation"
    eta:substanceOrTaxon_id = "ocean_near_surface"
Prefix example – ereefs with prefixes added

variables:

```plaintext
float eta(time, j, i) ;
  eta:units = "metre" ;
  eta:cf__long_name = "Surface elevation" ;
  eta:cf__standard_name = "cfsn__sea_surface_height_above_sea_level" ;
  eta:ereefs_medium_id = "feature__ocean"
  eta:ereefs_scaledQuantityKind_id = "property__sea_surface_elevation"
  eta:ereefs_substanceOrTaxon_id = "feature__ocean_near_surface"
```
Prefix example – added prefix mappings

variables:
    int prefix_list;
        prefix_list:cf__ = https://def.scitools.org.uk/CFTerms/
        prefix_list:cfsn__ = http://mmisw.org/ont/cf/parameter/

    float eta(time, j, i) ;
        eta:units = "metre" ;
        eta:cf__ long_name = "Surface elevation" ;
        eta:cf__ standard_name = "cfsn__sea_surface_height_above_sea_level" ;
        eta:ereefs__medium_id = "feature__ocean" 
        eta:ereefs__scaledQuantityKind_id = “property__sea_surface_elevation”
        eta:ereefs__substanceOrTaxon_id = “feature__ocean_near_surface”

Prefix Mappings
Variable metadata
Demo visualisations of graphs from CDL examples in bald repo

http://waterinformatics-ext1-cdc.it.csiro.au/nclc-demo/
## Supporting registries

### List all registers

<table>
<thead>
<tr>
<th>Name</th>
<th>Notation</th>
<th>Description</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACDD</td>
<td>ACDD</td>
<td>Vocabulary of terms used in the Attribute Conventions Dataset...</td>
<td>experimental</td>
</tr>
<tr>
<td>CFTerms</td>
<td>CFTerms</td>
<td>Vocabulary of terms used in the CF conventions for netCDF files...</td>
<td>experimental</td>
</tr>
<tr>
<td>NetCDF</td>
<td>NetCDF</td>
<td>Vocabulary of terms used in the netCDF User Guide.</td>
<td>experimental</td>
</tr>
</tbody>
</table>

[https://def.scitools.org.uk/](https://def.scitools.org.uk/)
Next steps

Push netCDF standardisation via OGC forward

Looking for collaborators to work on demonstrators

- Interoperability experiments support enhanced discovery of netCDF data
- Link controlled vocabularies with netCDF data
- Want to contribute?
Thanks

Active working group members:
James Biard (NCEI/NOAA)
Jonathan Yu (CSIRO, Australia)
Mark Hedley (UK Met Office)
Adam Leadbetter (Marine Institute, Ireland)

Python libraries (bald = binary array linked data)

https://github.com/binary-array-ld/bald

https://github.com/opengeospatial/netCDF-Classic-LD

http://tinyurl.com/netcdf-ld

Demo http://waterinformatics-ext1-cdc.it.csiro.au/ncld-demo/