

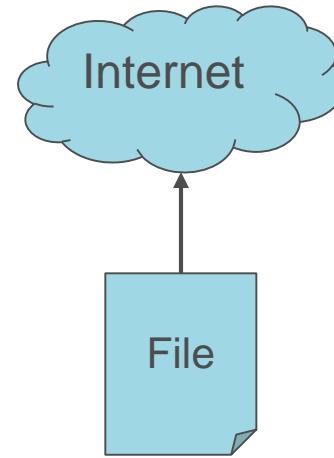


# netCDF-LD SKOS: demonstrating Linked Data vocabulary use within netCDF-compliant files

Nicholas Car  
*Data Architect*  
Geoscience Australia  
[nicholas.car@ga.gov.au](mailto:nicholas.car@ga.gov.au)

Prepared for ISESS2017 conference  
(<http://www.isess2017.org/>)

Presented on 2017-05-11



# Outline

- What we are doing and why
- Intro to main concepts
- ncskos
- ncskosdump tool
- Deployment scenarios
- Future (current) work
- Conclusions

# Outline

- **What we are doing and why**
- Intro to main concepts
- ncskos
- ncskosdump tool
- Deployment scenario
- Future (current) work
- Conclusions

# What we are doing and why

- **What**
  - Linking data files to reference metadata
- **Why**
  - Better metadata means better data use
- Not all metadata is equal
  - Curated vocab better than un-curated range values
  - Curated vocab better than static code lists
- New tools mean new metadata possibilities
  - Faster networks, better UIs, many more

# Outline

- What we are doing and why
- **Intro to main concepts**
  - netCDF
  - netCDF metadata communities
  - Linked Data
  - SKOS
- ncskos
- ncskosdump tool
- Deployment scenarios
- Future (current) work
- Conclusions

# Intro to main concepts

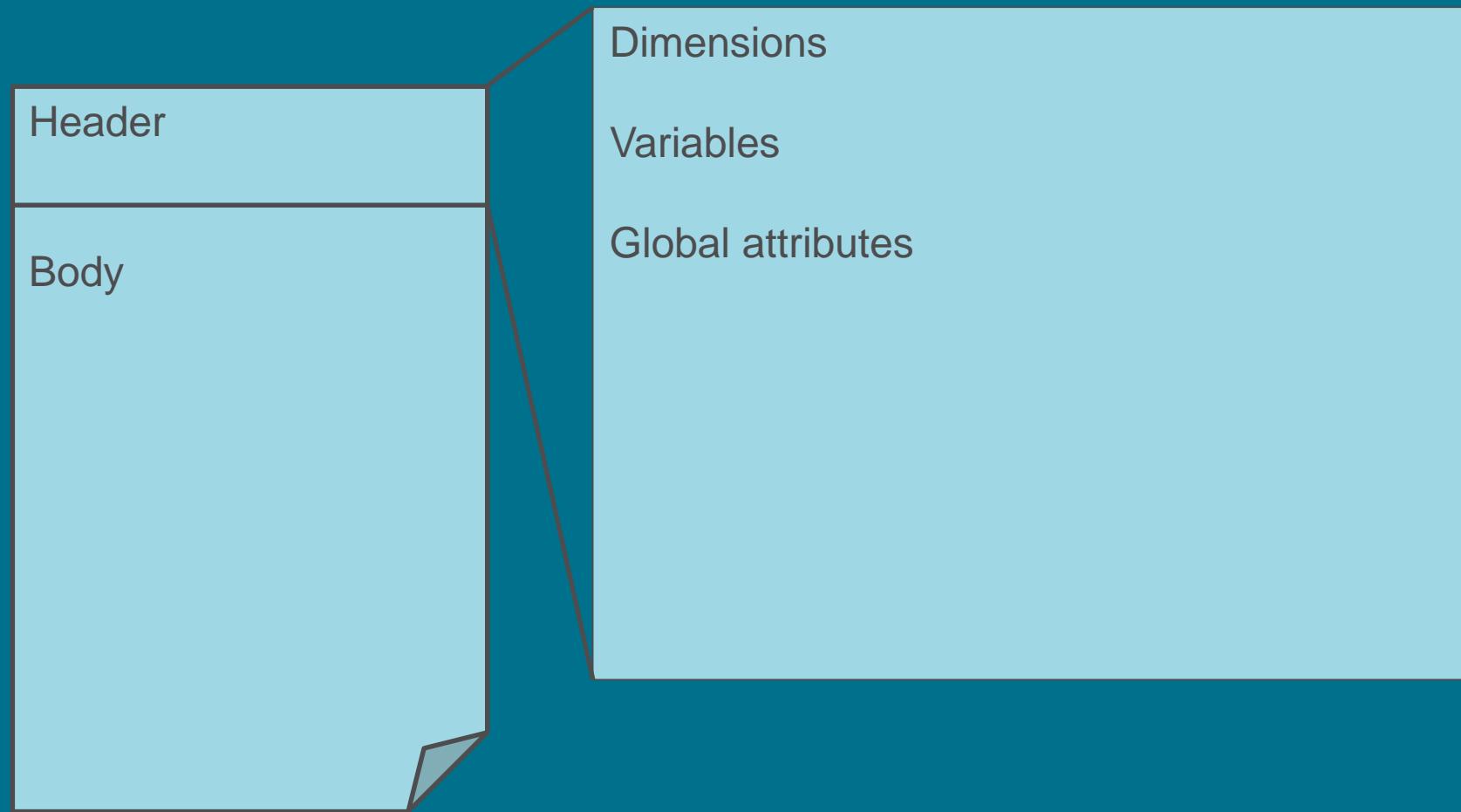
- **netCDF – Network Common Data Form**
  - Containers that include both data – usually array-oriented scientific data – and metadata
  - Intended to be ‘self-describing’
  - Very widely used
  - Example:  
[https://www.unidata.ucar.edu/software/netcdf/examples/sresa1b\\_ncar\\_ccsm3-example.cdl](https://www.unidata.ucar.edu/software/netcdf/examples/sresa1b_ncar_ccsm3-example.cdl)

# Intro to main concepts: netCDF

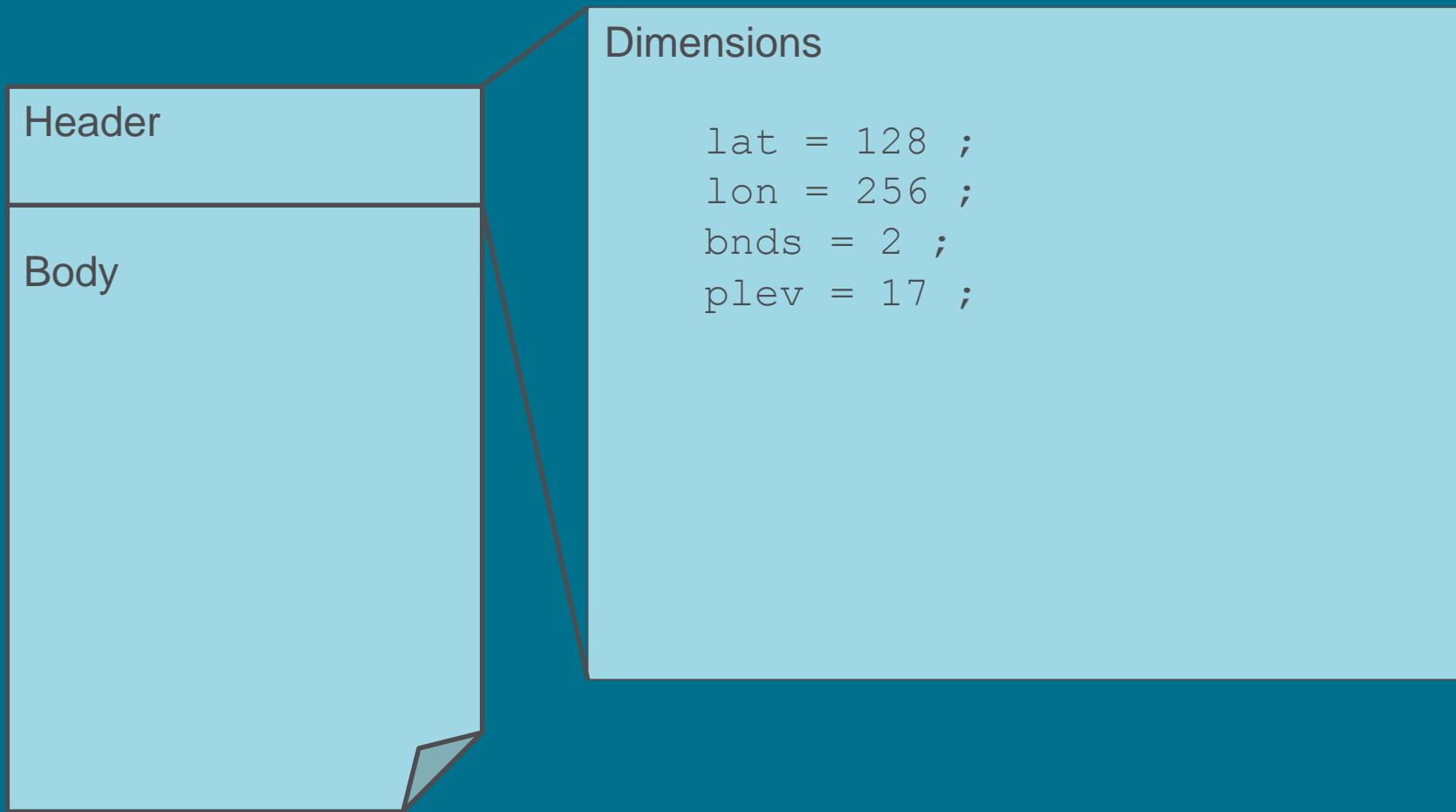
Header

Body

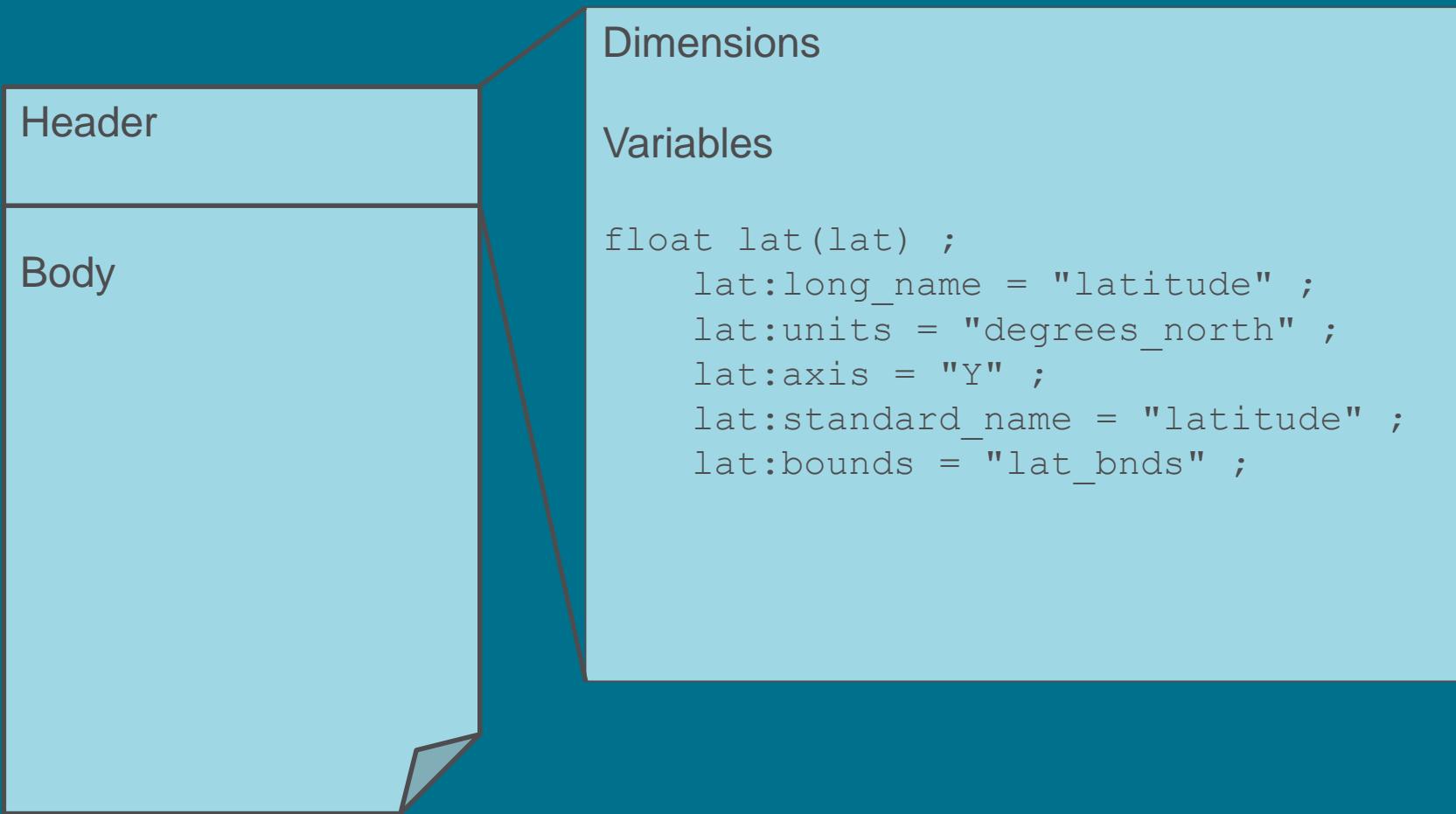
# Intro to main concepts: netCDF



# Intro to main concepts: netCDF



# Intro to main concepts: netCDF



# Intro to main concepts: netCDF

Header

Body

Dimensions

Variables

Global attributes

```
:title = "model output prepared for  
IPCC AR4" ;
```

```
:institution = "NCAR (National Center  
for Atmospheric"
```

```
:source = "CCSM3.0, version beta19" ;
```

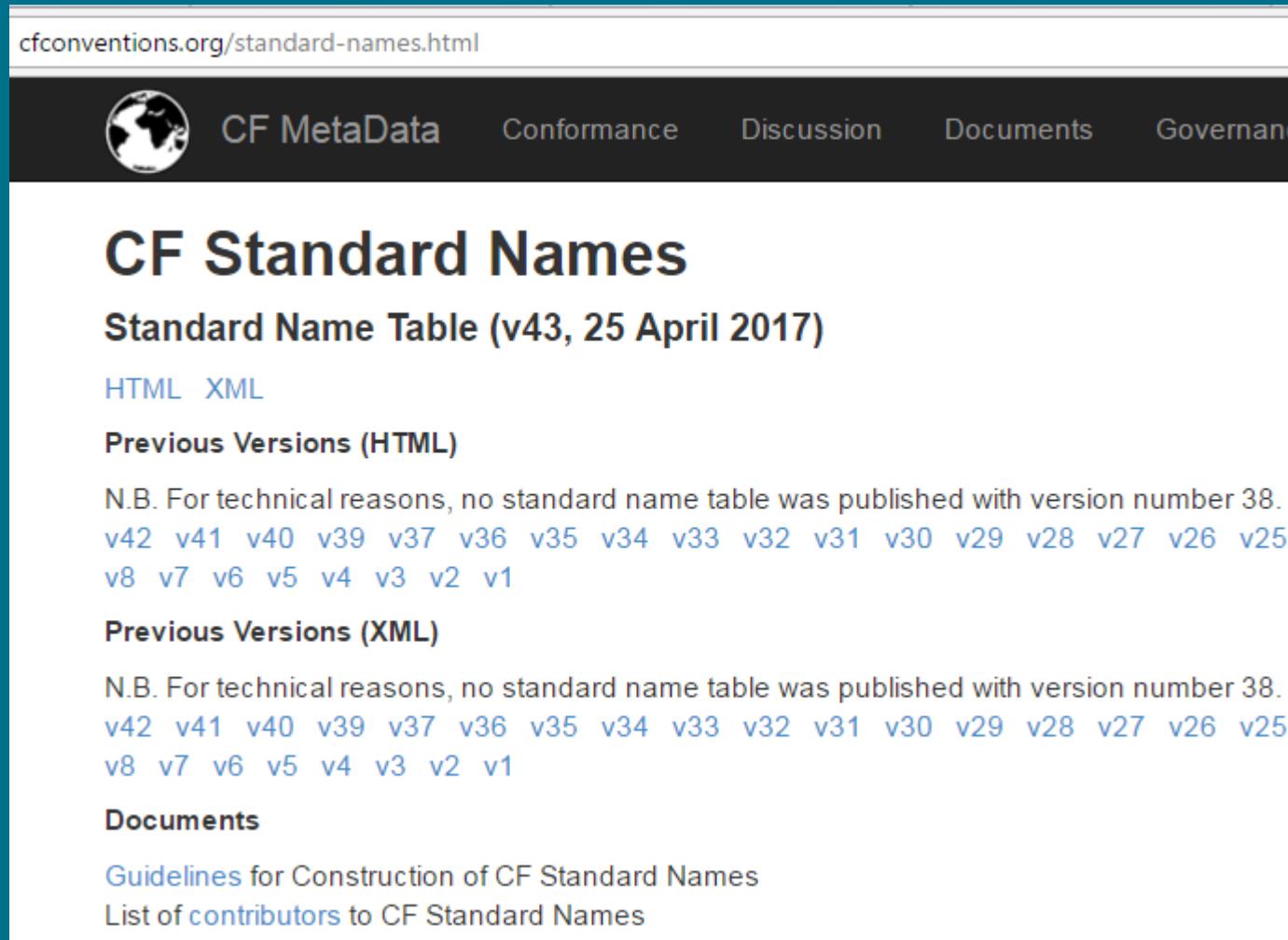
# Intro to main concepts

- **netCDF metadata communities**
  - Climate and Forecast (CF) conventions... for the description of Earth sciences data, intended to promote the processing and sharing of data files
  - provide a definitive description of what the data values found in each netCDF variable represent
  - Use code lists published at <http://cfconventions.org>

# Intro to main concepts

- netCDF metadata communities

[cfconventions.org/standard-names.html](http://cfconventions.org/standard-names.html)



The screenshot shows the homepage of the CF Standard Names website. The header includes a logo of a globe, navigation links for 'CF MetaData', 'Conformance', 'Discussion', 'Documents', and 'Governance', and the URL 'cfconventions.org/standard-names.html'. The main content area features a large title 'CF Standard Names' and a subtitle 'Standard Name Table (v43, 25 April 2017)'. Below this are links for 'HTML' and 'XML'. A section titled 'Previous Versions (HTML)' lists versions from v1 to v42. Another section titled 'Previous Versions (XML)' also lists versions from v1 to v42. A 'Documents' section includes links to 'Guidelines for Construction of CF Standard Names' and 'List of contributors to CF Standard Names'.

## CF Standard Names

### Standard Name Table (v43, 25 April 2017)

[HTML](#) [XML](#)

**Previous Versions (HTML)**

N.B. For technical reasons, no standard name table was published with version number 38.  
[v42](#) [v41](#) [v40](#) [v39](#) [v37](#) [v36](#) [v35](#) [v34](#) [v33](#) [v32](#) [v31](#) [v30](#) [v29](#) [v28](#) [v27](#) [v26](#) [v25](#)  
[v8](#) [v7](#) [v6](#) [v5](#) [v4](#) [v3](#) [v2](#) [v1](#)

**Previous Versions (XML)**

N.B. For technical reasons, no standard name table was published with version number 38.  
[v42](#) [v41](#) [v40](#) [v39](#) [v37](#) [v36](#) [v35](#) [v34](#) [v33](#) [v32](#) [v31](#) [v30](#) [v29](#) [v28](#) [v27](#) [v26](#) [v25](#)  
[v8](#) [v7](#) [v6](#) [v5](#) [v4](#) [v3](#) [v2](#) [v1](#)

**Documents**

[Guidelines for Construction of CF Standard Names](#)  
[List of contributors to CF Standard Names](#)

# Intro to ma

- netCDF m

← → × ⓘ cfconventions.org/Data/cf-standard-names/42/build/cf-standard-name-table.html

units which are physically equivalent (not necessarily identical) to the canonical units, possibly modified. [Section 1.3 of the CF conventions](#) states: "The values of the units attributes are character strings that are "Units".". For example, a variable with the standard name of "air\_temperature" may have a units attribute refer to the [Udunits documentation](#). Refer to the [CF conventions](#) for full details of the units attribute.

The screenshot shows a web browser window displaying the CF Standard Name Table. The URL in the address bar is cfconventions.org/Data/cf-standard-names/42/build/cf-standard-name-table.html. The page has a header with the CF logo and navigation links for HTML and XML versions. A search bar at the top right contains the placeholder text "Search". Below it are two radio buttons: "AND" (selected) and "OR (separate search terms with spaces)", and a checkbox for "Also search help text". To the right are buttons for "Search Standard Names" and "Show All Standard Names". The main content area is titled "View by Category" and features a grid of links: Atmospheric Chemistry, Atmosphere Dynamics, Carbon Cycle, Cloud, Hydrology, Ocean Dynamics, Radiation, Sea Ice, and Surface. Below this is a table with a single column labeled "Standard Name" containing a list of standard names preceded by a blue triangle icon.

Standard Name
► <a href="#">acoustic_signal_roundtrip_travel_time_in_sea_water</a>
► <a href="#">aerodynamic_particle_diameter</a>
► <a href="#">aerodynamic_resistance</a>
► <a href="#">age_of_sea_ice</a>
► <a href="#">age_of_stratospheric_air</a>
► <a href="#">age_of_surface_snow</a>

# Intro to map

- netCDF m

units which are physically equivalent (not necessarily identical) to the canonical units, possibly modified. [Section 1.3 of the CF conventions](#) states: "The values of the units attributes are character strings that are "Units".". For example, a variable with the standard name of "air\_temperature" may have a units attribute refer to the [Udunits documentation](#). Refer to the [CF conventions](#) for full details of the units attribute.

The screenshot shows a search interface for the CF conventions. At the top, there is a search bar with the placeholder text "Search Standard Names" and a "Show All Standard Names" button. Below the search bar, there is a list of standard names. The first item in the list is "air potential temperature", which is expanded to show its definition: "Potential temperature is the temperature a parcel of air or sea water would have if moved adiabatically to sea level pressure." To the right of the definition, there is a table with four columns: "Standard Name", "K", "theta", and "13".

Standard Name	K	theta	13
---------------	---	-------	----

The screenshot shows a search interface for the CF conventions. At the top, there is a search bar with the placeholder text "Search Standard Names" and a "Show All Standard Names" button. Below the search bar, there is a list of standard names. The first item in the list is "air potential temperature", which is expanded to show its definition: "Potential temperature is the temperature a parcel of air or sea water would have if moved adiabatically to sea level pressure." To the right of the definition, there is a table with four columns: "Standard Name", "K", "theta", and "13".

Standard Name	K	theta	13
---------------	---	-------	----

## Previous Versions

N.B. For technical details see the v42 v41 v40 v39 v38 v37 v36 v35 v34 v33 v32 v31 v30 v29 v28 v27 v26 v25 v24 v23 v22 v21 v20 v19 v18 v17 v16 v15 v14 v13 v12 v11 v10 v9 v8 v7 v6 v5 v4 v3 v2 v1 v0

## Documents

Guidelines for Authors

List of control variables

## Standard Name

► [acoustic\\_signal\\_roundtrip\\_travel\\_time\\_in\\_sea\\_water](#)

► [aerodynamic\\_particle\\_diameter](#)

► [aerodynamic\\_resistance](#)

► [age\\_of\\_sea\\_ice](#)

► [age\\_of\\_stratospheric\\_air](#)

► [age\\_of\\_surface\\_snow](#)

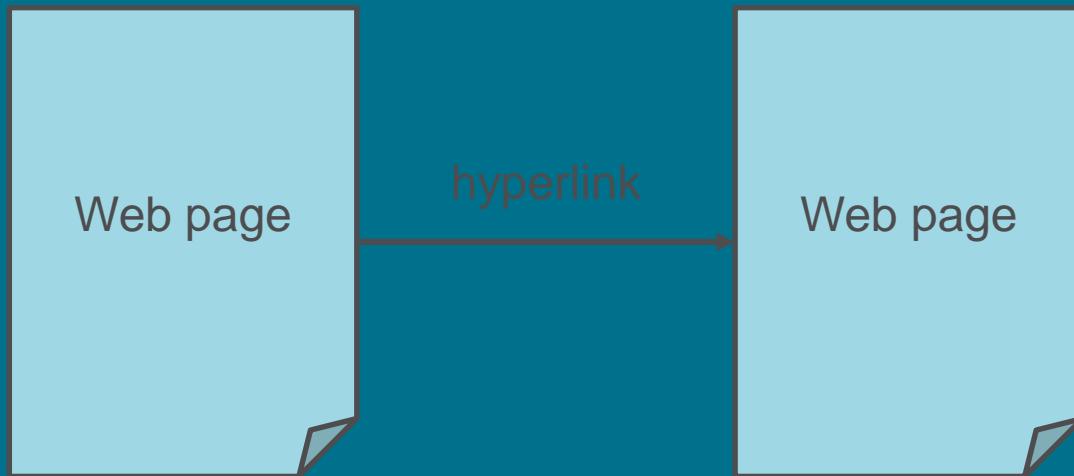
# Intro to main concepts

- **Linked Data**

- “Linked Data provides a publishing paradigm in which not only documents [web pages] but also data can be first class citizens of the Web”

– Heath & Bizer, 2011

- Normal web pages:



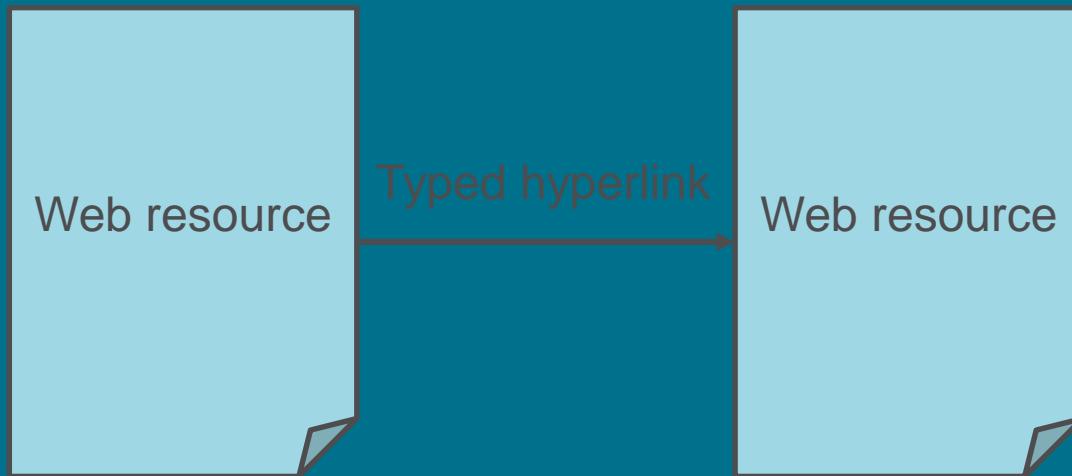
# Intro to main concepts

- **Linked Data**

- “Linked Data provides a publishing paradigm in which not only documents [web pages] but also data can be first class citizens of the Web”

– Heath & Bizer, 2011

- Linked Data:



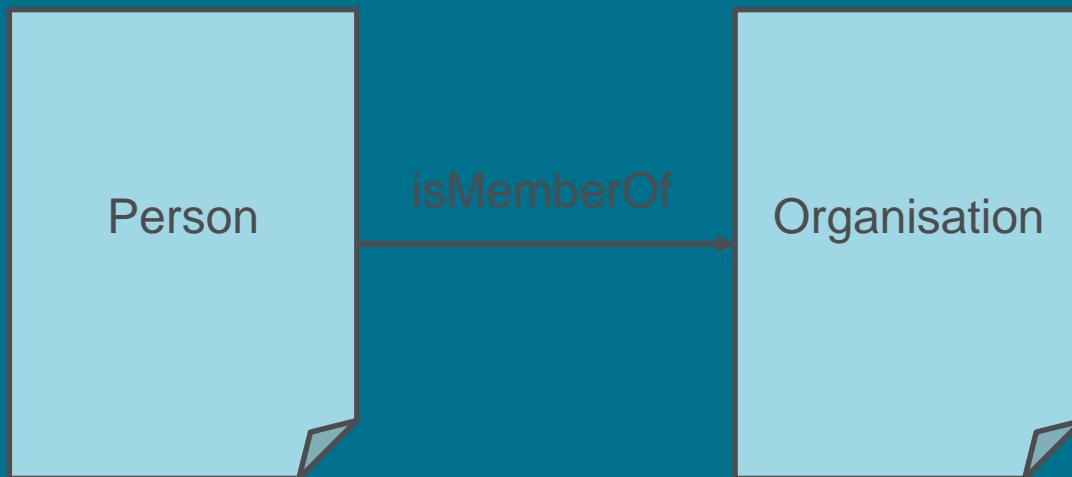
# Intro to main concepts

- **Linked Data**

- “Linked Data provides a publishing paradigm in which not only documents [web pages] but also data can be first class citizens of the Web”

– Heath & Bizer, 2011

- Linked Data:

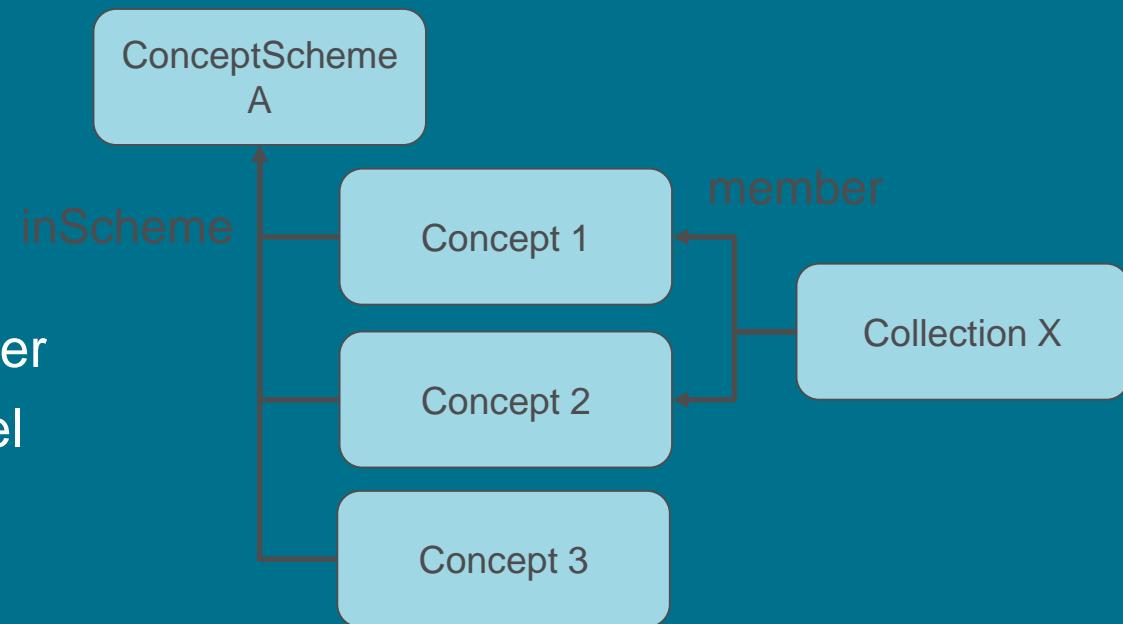


Person  
Organisation  
isMemberOf

...are all defined

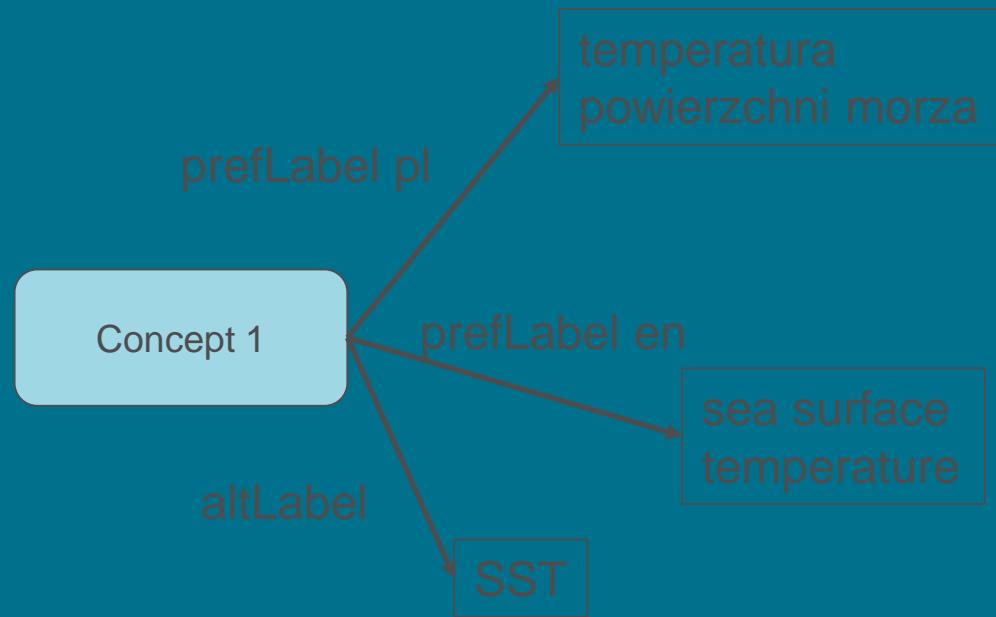
# Intro to main concepts

- **SKOS – Simple Knowledge Organization System**
  - A vocabulary data model
  - Main classes:
    - Concept
    - ConceptScheme
    - Collection
  - Main relations:
    - hasTopConcept
    - inScheme, member
    - prefLabel, altLabel



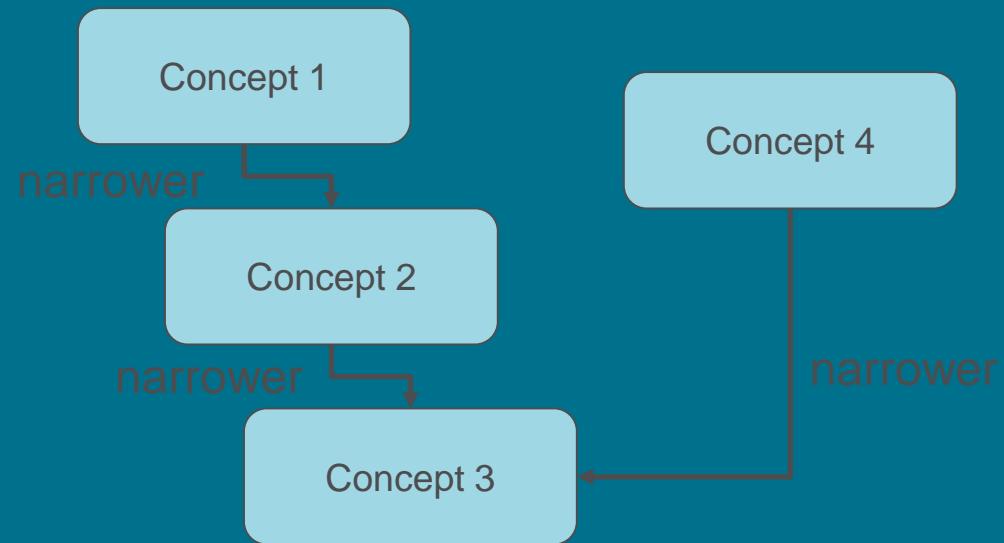
# Intro to main concepts

- **SKOS – Simple Knowledge Organization System**
  - A vocabulary data model
  - Main classes:
    - Concept
    - ConceptScheme
    - Collection
  - Main relations:
    - hasTopConcept
    - inScheme, member
    - prefLabel, altLabel



# Intro to main concepts

- **SKOS – Simple Knowledge Organization System**
  - A vocabulary data model
  - Main classes:
    - Concept
    - ConceptScheme
    - Collection
  - Main relations:
    - hasTopConcept
    - inScheme, member
    - prefLabel, altLabel
  - broader, narrower



# Outline

- What we are doing and why
- Intro to main concepts
- **ncskos**
- ncskosdump tool
- Deployment scenarios
- Future (current) work
- Conclusions

# ncskos

- We proposed a semi-sophisticated LD encoding for netCDF files in ISESS 2015
  - “Towards linked data conventions for delivery of environmental data using netCDF”
  - Used JSON-LD style “context” variables to specify namespaces
  - Proposed but did not define a netCDF structural ontology

```
var:ref= "http://vocab.nerc.ac.uk/collection/P07/current/CFSN0600/";  
var:unit = "Meters";  
var:unit_ref = " http://qudt.org/vocab/unit#Meter";
```

# ncskos

- Now, a simpler method of including only SKOS definition links in netCDF files
- Aiming for
  - Scale deployment to make case for utility
  - Tool generation
  - Establishment of useful reference datasets

```
float tos(time, lat, lon) ;  
  tos:standard_name = "sea_surface_temperature" ;  
  tos:long_name = "Sea Surface Temperature" ;  
  tos:units = "K" ;  
  tos:original_units = "degC" ;  
  ...
```

# ncskos

- Now, a simpler method of including only SKOS definition links in netCDF files
- Aiming for
  - Scale deployment to make case for utility
  - Tool generation
  - Establishment of useful reference datasets

```
float tos(time, lat, lon) ;  
    tos:skos_concept_uri =  
"http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-  
ToS/sea_surface_temperature";  
    tos:standard_name = "sea_surface_temperature" ;  
    tos:long_name = "Sea Surface Temperature" ;  
    tos:units = "K" ;  
    tos:original_units = "degC" ;  
    ...
```

## ncskos

- Now, a simpler method of including only SKOS definition links in netCDF files
- Aiming for
  - Scale deployment to make case for utility
  - Tool generation
  - Establishment of useful reference datasets

```
float tos(time, lat, lon) ;  
    tos:skos_concept_uri =  
"http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-  
ToS/sea_surface_temperature";
```

# ncskos

Concept of “sea surface temperature”, RESTful link:

[http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS/sea\\_surface\\_temperature](http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS/sea_surface_temperature)

Whole vocab:

<http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS>

Also in RDF for machines



## SISSVoc Default

**sea surface temperature / temperatura powierzchni morza**

alt label	SST	
broader	<a href="http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS/surface_temperature">http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS/surface_temperature</a> [original]	
narrower	<a href="#">sea_surface_temperature</a> [original]	
	<a href="http://skos.um.es/unesco6/251007">http://skos.um.es/unesco6/251007</a> [original]	
narrower	<a href="#">sea_surface_temperature</a> [original]	
definition	Sea surface temperature is usually abbreviated as "SST". It is the temperature of sea water near the surface (including the part under sea-ice, if any). More specific terms, namely sea_surface_skin_temperature, sea_surface_subskin_temperature, and surface_temperature are available for the skin, subskin, and interface temperature, respectively. For the temperature of sea water at a particular depth or layer, a data variable of sea_water_temperature with a vertical coordinate axis should be used.	
narrower	<a href="http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS/sea_surface_skin_temperature">http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS/sea_surface_skin_temperature</a> [original]	
broader	<a href="#">sea_surface_temperature</a> [original]	
	<a href="http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS/sea_surface_subskin_temperature">http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS/sea_surface_subskin_temperature</a> [original]	
broader	<a href="#">sea_surface_temperature</a> [original]	

# Outline

- What we are doing and why
- Intro to main concepts
- ncskos
- **ncskosdump tool**
- Deployment scenarios
- Future (current) work
- Conclusions

# ncskosdump tool

- Many well used netCDF tools:
  - ncdump
  - ncgen
  - ncview
  - Operators Toolkit

# ncskosdump tool

- Many well used netCDF tools:

- ncdump
- ncgen
- ncview
- Operators Toolkit

```
# ncdump -h some_file.nc  
  
sst:skos_concept_uri =  
"http://pid.geoscience.gov.au/def/voc/netCDF-LD-  
eg-ToS/sea_surface_temperature" ;
```

# ncskosdump tool

- Many well used netCDF tools:
  - ncdump
  - ncgen
  - ncview
  - Operators Toolkit
- ncskosdump
  - Wraps ncdump, adds commands for SKOS info retrieval

```
# ncskosdump -h some_file.nc --lang=pl altlabels=True  
  
sst:skos__prefLabel_pl = "temperatura powierzchni morza" ;  
sst:skos__altLabels = "SST" ;
```

# ncskosdump tool

- Many well used netCDF tools:
  - ncdump
  - ncgen
  - ncview
  - Operators Toolkit
- ncskosdump
  - Wraps ncdump, adds commands for SKOS info retrieval

```
# ncskosdump -h sst.nc --narrower=True  
  
sst:skos__narrower = "http://...01, http://...02, http://...03"
```

# ncskosdump tool

- Many well used netCDF tools:
  - ncdump
  - ncgen
  - ncview
  - Operators Toolkit
- ncskosdump
  - Wraps ncdump, adds commands for SKOS info retrieval
  - Session (memory) and location (file) metadata retrieval caching enabled

# ncskosdump tool

- Commands:

- verbose** to enable verbose output

- lang=<lang\_code>** where <lang\_code> is a two-character ISO 639-1:200 code for the language in which the results are sought

- narrower** to recursively create complete tree of narrower concepts, not just ones resolved directly from URIs

- altLabels=<altLabel\_list>** where <altLabel\_list> is a comma-separated list of altLabels to match in order to list their associated datasets

- retries=<max\_retries>** where <max\_retries> is the maximum number of retries to attempt for unresolved URIs. Default retries = 0

- delay=<retry\_delay\_seconds>** where <retry\_delay\_seconds> is the number of seconds to wait before each retry. Default delay = 2s

- refresh** to discard current file cache and repopulate the cache from scratch

# Outline

- What we are doing and why
- Intro to main concepts
- ncskos
- ncskosdump tool
- **Deployment scenarios**
- Future (current) work
- Conclusions

# Deployment scenarios

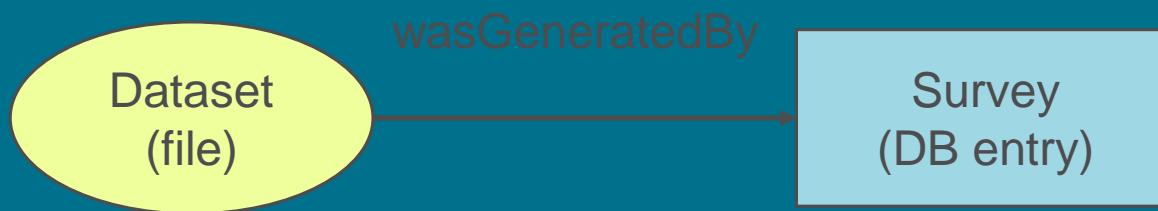
- Example in the paper & GitHub
  - Series of empty netCDF files
  - Test SKOS vocab:  
<http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS>
  - Demo program to sort files, based on hierarchical relations
- Real-world scenario
  - 10,000+ netCDF files at the Aust. Nat. Comp. Center
  - Geophysics data, multiple variables per file
  - Variables being described in a Geophysics SKOS vocab
  - Automated selection using ncskosdump

# Outline

- What we are doing and why
- Intro to main concepts
- ncskos
- ncskosdump tool
- Deployment scenario
- **Future (current) work**
  - PROV
  - netCDF-LD

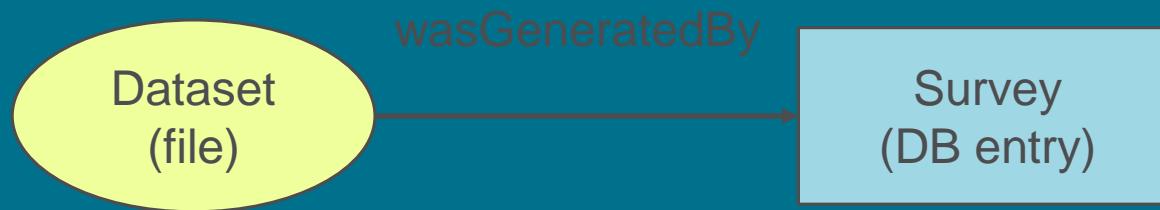
# Future (current) work

- PROV
- We have extended the use of a variable name prefix for the PROV ontology
  - file:prov\_{var\_name}
  - file:prov\_wasGeneratedBy



# Future (current) work

- PROV
- We have extended the use of a variable name prefix for the PROV ontology
  - file:prov\_{var\_name}
  - file:prov\_wasGeneratedBy



:prov\_wasGeneratedBy = "http://pid.geoscience.gov.au/survey/801"



## Applying geoscience to Australia's most important challenges

Australian Government  
Geoscience Australia

GA Home Surveys API Home Surveys Register

# Survey

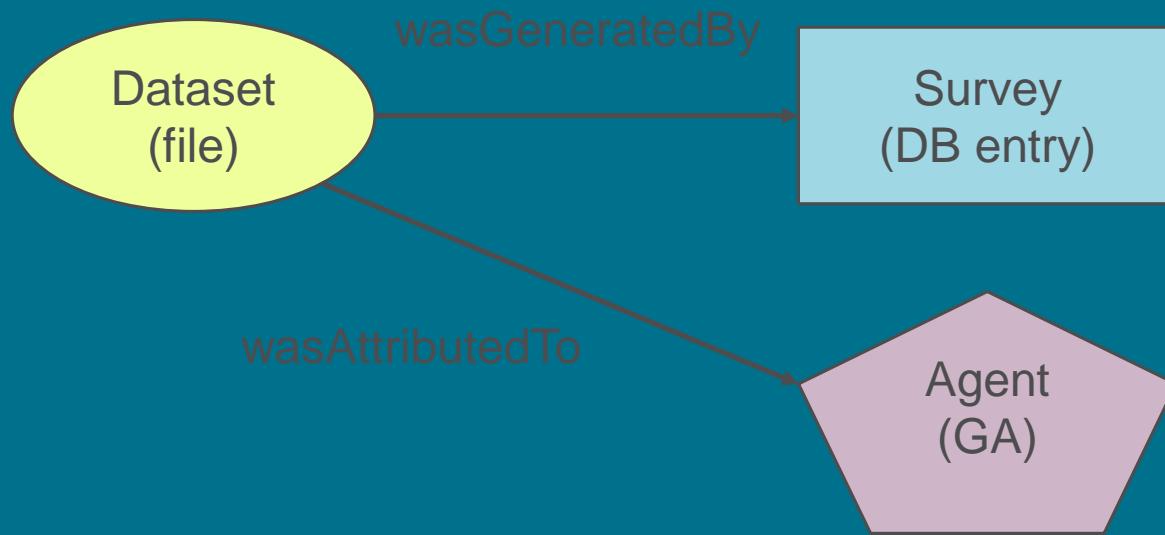
## GAPD view

These properties are those listed in the [GAPD Model](#) of a Survey.

Property	Value
ID	801
Name	Basin/Quondong, Vale, SA Company, 1988 (88SA14) (74pt)
State	SA
Agents:	
Contractor	Kevron Geophysics Pty Ltd
Operator	MESA & Dominion Mining
Processor	Aerodata Holdings Pty Ltd
survey_type	Semi-detailed
data_types	MAG
Vessel:	
Type	Plane
Name	None
release_date	1989-01-01 00:00:00
onshore_offshore	Onshore

# Future (current) work

- PROV
- We have extended the use of a variable name prefix for the PROV ontology
  - file:prov\_{var\_name}
  - file:prov\_wasGeneratedBy



# Future (current) work

- **netCDF-LD**
- We are working on a full netCDF LD implementation
  - WG with netCDF developers and users
- Making an encoding format for any LD data in netCDF files
  - It would be a new netCDF specification
  - Interest looks high (AGU, EGU, other confs)
- Using Spatial Data on the Web WG conventions
- Re-using JSON-LD encoding where possible

# Outline

- What we are doing and why
- Intro to main concepts
- ncskos
- ncskosdump tool
- Deployment scenario
- Future (current) work
- **Conclusions**

# Conclusions

- Just using SKOS is:
  - Easy
  - Useful – standardised vocab, hierarchy
  - Better than CF – open world, curation still possible
- Using limited PROV looks to be:
  - Easy
  - Useful – people really want provenance, reuse important
- Use of SKOS & PROV:
  - Paving the way to full LD

# Conclusions

- Final thoughts:
  - This work brings well managed yet Open World to scientific data
  - Overcomes several major problems with Linked Data:
    - Relevance to non ontologists – domain application
    - Computational efficiency – appropriateness
    - Tooling – wrapping existing tools



## netCDF-LD SKOS:

demonstrating Linked Data vocabulary use within netCDF-compliant files

**Nicholas Car**

*Data Architect*

Geoscience Australia

[nicholas.car@ga.gov.au](mailto:nicholas.car@ga.gov.au)

Code: <http://pid.geoscience.gov.au/dataset/103620>

Vocab: <http://pid.geoscience.gov.au/def/voc/netCDF-LD-eg-ToS>

