Open Geospatial Consortium

Approval Date:2012-09-25

Publication Date: 2013-01-17

External identifier of this OGC® document: http://www.opengis.net/doc/dp/netcdf-uncertainty

Reference number of this document: OGC 11-163

Category: Public Discussion Paper

Editors:Lorenzo Bigagli, StefanoNativi

NetCDF Uncertainty Conventions (NetCDF-U)

Warning

This document is not an OGC Standard. This document is an OGC Discussion Paper and is not an official position of the OGC membership. The document is distributed for review and comment. The document is subject to change without notice and may not be referred to as an OGC Standard.

Document type: OGC® Discussion Paper
Document subtype: NA
Document stage: Approved for public release
Document language: English

Contents

Page

Figures

Page

i. Abstract

This Discussion Paperproposes a set of conventions for managing uncertainty information within the netCDF3 data model and format: the NetCDF Uncertainty Conventions (NetCDF-U).

ii. Keywords

ogcdoc, netcdf, uncertainty, uncertml

iii. Preface

Suggested additions, changes, and comments on this document are welcome and encouraged. Such suggestions may be submitted by email message or by making suggested changes in an edited copy of this document.

The changes made in this document version, relative to the previous version, are tracked by Microsoft Word, and can be viewed if desired. If you choose to submit suggested changes by editing this document, please first accept all the current changes, and then make your suggested changes with change tracking on.

iv. Document contributor contact points

All questions regarding this document should be directed to the editor or the contributors:

Name	Organization
Lorenzo Bigagli	National Research Council of Italy
Stefano Nativi	National Research Council of Italy

v. Revision history

Date	Release	Editor	Primary clauses modified	Description
10/11/2011	1.0	Lorenzo Bigagli, Stefano Nativi	all	Initial draft for general review; based onaprevious draft presented and discussed at the September 2011 OGC TC Meeting and an UncertWeb Technical Report [2]
16/12/2011		Lorenzo Bigagli	6.4	Canceled _CoordinateAxisType attributes from examples to improve clarity
16/03/2012		Lorenzo Bigagli	6.6	Fixed example on distribution encoding

vi. Future work

The netCDF standard is evolving from version 3 (netCDF3) to version 4 (netCDF4) that adopts an abstract Common Data Model (CDM) to be implemented in different encodings (e.g. netCDF and ncML). This transition affects the possible ways to accommodate uncertainty, including what proposed in this document.

In thefuture, a netCDF4approach for uncertainty encodingwill be investigated. Since netCDF4 accepts data structures as values,it is feasible to extend the netCDF model to include uncertain data types. A data structure can be used to represent an uncertain data type (such as an uncertain integer or an uncertain float) according to the UncertML model. This approach is the most flexible, since it will allow the definition of an algebra for uncertainty and to implement it in a netCDF API extension.

Foreword

The intended use of the conventions proposed in this document is to qualify the uncertainty related to geospatial data encoded in the netCDF3 format. Hence, this document may relate to other OGC standards that make use of the netCDF3 data format, in particular the netCDFEncoding Format Extension of theWeb Coverage Service Interface standard (under discussion).

The work presented in this document does not cancel or replace other OGC documents.

This work was partially supported by the UncertWeb project[1],that has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° [248488].

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this draft document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

[1] http://www.uncertweb.org/

Introduction

As discussed in [4], from a theoretical perspective, it can be said that no dataset is a perfect representation of the reality it purports to represent.

Inevitably errors, arising from the observation process, including the sensor system and subsequent processing, differences in scales of phenomena and the spatial support of the observation mechanism as well as a lack of knowledge about the detailed conversion between the measured quantity and the target variable means that in principle all data should be treated as uncertain.

The most natural representation of an uncertain quantity is in terms of random variables (or fields / functions for spatially and temporally distributed variables), with a probabilistic approach.

However, it must be acknowledged that almost all existing data resources are not treated in this way. Most datasets come simply as a series of values, often without any uncertainty information. If there is uncertainty information, then this is typically contained within the metadata, in a data quality element. This is typically a global (dataset wide) representation of uncertainty, often derived through some form of validation process. Typically, it is a statistical measure of spread, for example the standard deviation of the residuals (data set measured minus ‘true’ value).

The introduction of a mechanism by which such descriptions of uncertainty can be integrated into existing geospatialapplications is considered a practical step towards a more accurate modeling of our uncertain understanding of any natural process.

This document proposes a linked data approach forthe widely used netCDF format. To encode uncertainty information in a netCDF dataset, it is possible to consider at least the following approaches:

1. To define a netCDF convention to specify uncertainty-related concepts in a netCDF file, enabling the representation of uncertain quantities, as well as values known with a given uncertainty;
2. **Linked netCDF**: to adopt the Semantic Web linked data approach, expressing the uncertainty in an XML document according to the UncertML model, which is then linked to netCDF-CF attributes and variables using inbound or outbound XLink references. This approach is valuable for re-use of data sets since uncertainty information can be associated to known values without affecting the existing encodings, applying Semantic Web principles and technologies;

This document takes the first approach and introduces the NetCDF Uncertainty Conventions (NetCDF-U)for capturing uncertainty in netCDF-encoded data.

OGC® NetCDF-U Discussion Paper

1 Scope

This OGC® document introduces a set of conventions and mechanisms that extend and qualify the netCDF3 data model and format to model uncertain information: the NetCDF Uncertainty Conventions (NetCDF-U).

Given the generality and flexibility of the netCDF data model, conventions on naming, semantics and data structure have been adopted by several community of practice, as a means of improving data interoperability. Some of the existing conventions may include specific provisions on uncertain elements and concepts, but, to our knowledge, no general convention on the encoding of uncertainty has been proposed, to date, what is the scope of this document.

This OGC™ document is applicable to data encoded in the netCDF3 format, with possible limitations, depending on conflicting conventions to which a given netCDF dataset may be compliant, in addition to NetCDF-U (in general, a netCDF may be compliant with more than one convention).

The NetCDF-U Conventions are designed to be fully compatible with the netCDF Climate and Forecast Conventions, the de-facto standard for a large amount of data in the Fluid Earth Science community.

2 Normative references

The following normative documents contain provisions that, through reference in this text, constitute provisions of this document. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the latest edition of the normative document referred to applies.

B. Eaton, et al., *NetCDF Climate and Forecast (CF) Metadata Conventions*, version 1.5, 25 October 2010.

IETF RFC 3986, *Uniform Resource Identifier (URI): Generic Syntax.*

OGC 10-090r3,*Network Common Data Form (NetCDF) Core Encoding Standard version 1.0*.

OGC 10-092r3,*NetCDF Binary Encoding Extension Standard: NetCDF Classic and 64-bit Offset Format*.

3 Terms and definitions

For the purposes of this standard, the definitions specified in Clause 4 of the OWS Common Implementation Specification [OGC 05-008] shall apply.

4 Conventions

4.1 Abbreviated terms

Most of the abbreviated terms listed in Subclause 5.1 of the OWS Common Implementation Specification [OGC 05-008] apply to this document.

4.2 UML notation

The diagrams that appear in this standard are presented using the Unified Modeling Language (UML) static structure diagram, as described in Subclause 5.2 of [OGC 05-008].

5 Compatibility with NetCDF-CF

An important constraint for NetCDF-U is to preserve compatibility with the NetCDF Climate and Forecast Conventions (NetCDF-CF) [Eaton 2010], whose data model is represented below.

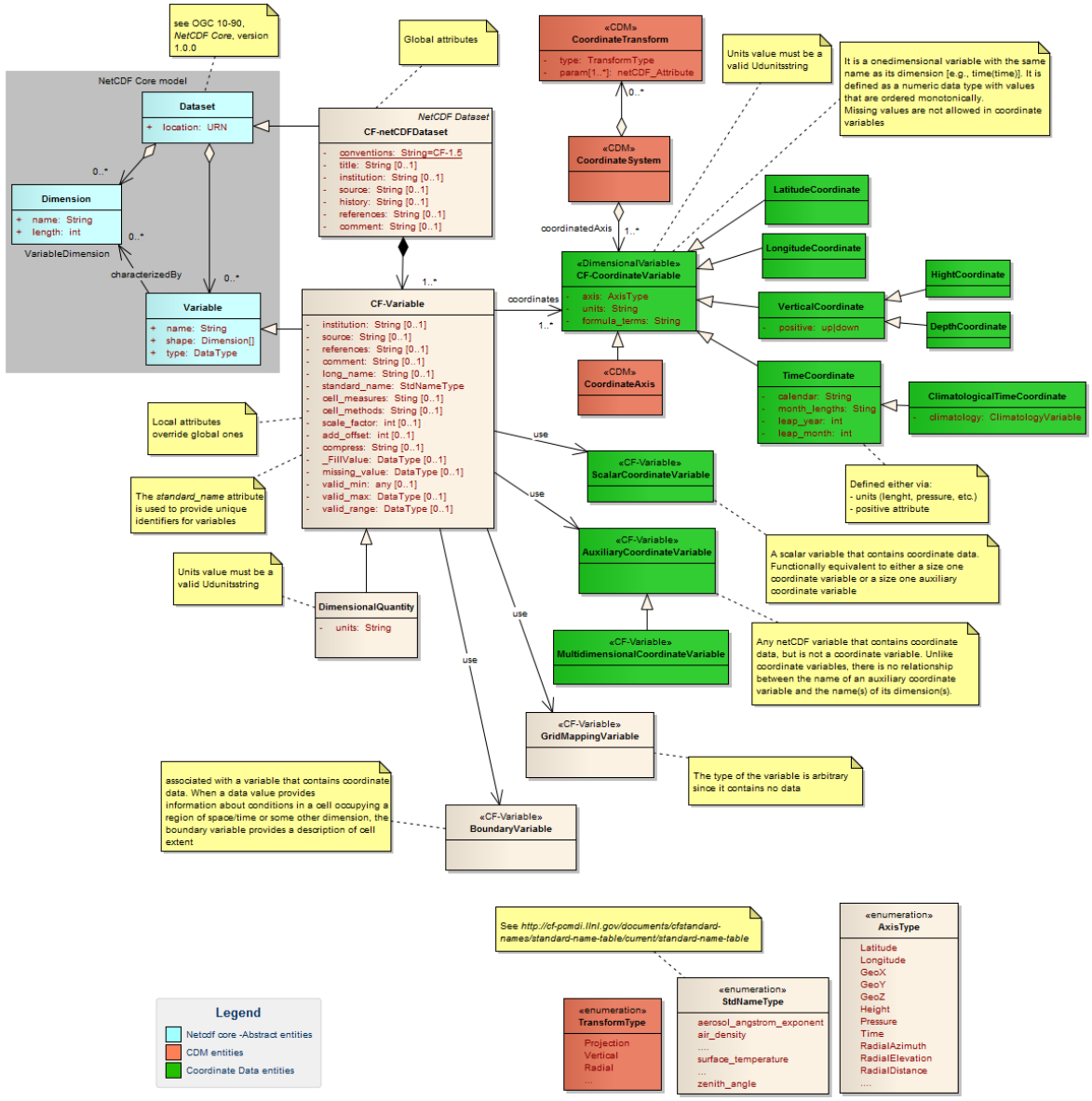

Figure 1 – NetCDF-CF 1.5 data model (source: [1])

NetCDF-CF mentions the issue of uncertainty representation and provides mechanisms for its expression, including the following:

- Data variables may be associated to variables representing their uncertainty via the *ancillary_variables* attribute;
- Data variables representing uncertainty may be expressed via modifiers of the quantity standard name attribute (actually, only the *standard_error* modifier is relevant to the present context);
- Common statistical operators are expressed via the *cell_methods* attribute, e.g. *mean*, *standard deviation*, *variance*;

Where possible, the above mechanisms are leveraged by NetCDF-U, e.g. using the same attributes name with compatible semantics.

6 NetCDF Uncertainty Conventions

This section describes the proposed NetCDF Uncertainty Conventions (NetCDF-U) for capturing uncertainty in netCDF-encoded data.

NetCDF files compliant with these conventions may be indicated as "netCDF-U files" in the following.

6.1 Rationale

These conventions have the following rationale:

- Compatibility with netCDF-CF Conventions 1.5;
- Human-readability of conforming datasets structure;
- Minimal difference between certain/agnostic and uncertain representations of data (e.g. with respect to dataset structure).

6.2 Global provisions

6.2.1 Identification of Conventions

NetCDF-U files shall declare the global attribute *Conventions*[2] with the string value "UW-1.0" (see the example in the next section).

6.2.2 Primary and ancillary variables

A netCDF file typically contains several variables. Some of these may depend on others, i.e. contain ancillary information referred to or qualifying concepts encoded by other variables (e.g., netCDF-CF introduces the concepts of coordinate variable, boundary variable, flag variable.)

[2]Lower-case "*conventions*" attribute name should be acceptable. However, netCDF-CF specify the upper-case form.

Similarly to netCDF-CF, netCDF-U variables may declare an *ancillary_variables* attribute for expressing dependency relationships among variables.

Dually, netCDF-U files may declare a *primary_variables* global attribute, whose value is a white-separated list of variable identifiers.

The intended use of the *primary_variables* attribute is to support applications in directly accessing their presumable data of interest, particularly since netCDF-U files are likely to contain a large number of ancillary variables, as in the example below.

```
netcdfbiotemperature {
// global attributes:
:Conventions = "CF-1.5 UW-1.0";
:primary_variables = "biotemperature";
dimensions:
lon = 240 ;
lat = 163 ;
variables:
doublelon(lon) ;
lon:long_name = "longitude";
lon:units = "degrees_east";
lon:standard_name = "longitude";
doublelat(lat);
lat:long_name = "latitude";
lat:units = "degrees_north";
lat:standard_name = "latitude";
doublebiotemperature_mean(lat, lon);
biotemperature_mean:missing_value = -999.;
biotemperature_mean:ref =
"http://www.uncertml.org/distributions/normal#mean";
doublebiotemperature_variance(lat, lon);
biotemperature_variance:missing_value = -999.;
biotemperature_variance:ref =
"http://www.uncertml.org/distributions/normal#variance";
doublebiotemperature(lat, lon);
biotemperature:units = "degC";
biotemperature:ancillary_variables =
"biotemperature_meanbiotemperature_variance";
biotemperature:ref = "http://www.uncertml.org/distributions/normal";
}
```

6.3 Uncertain concepts

The main mechanism for capturing uncertain concepts in netCDF-U files consists of encoding them (and their possible values) in netCDF data variables annotated with uncertainty-related semantics.

To this end, a netCDF variable referring to uncertain concepts shall declare a *ref* attribute, whose value is a list of space-separated URIs of UncertML 2.0 dictionary entries (see the conceptual model in Figure 2).

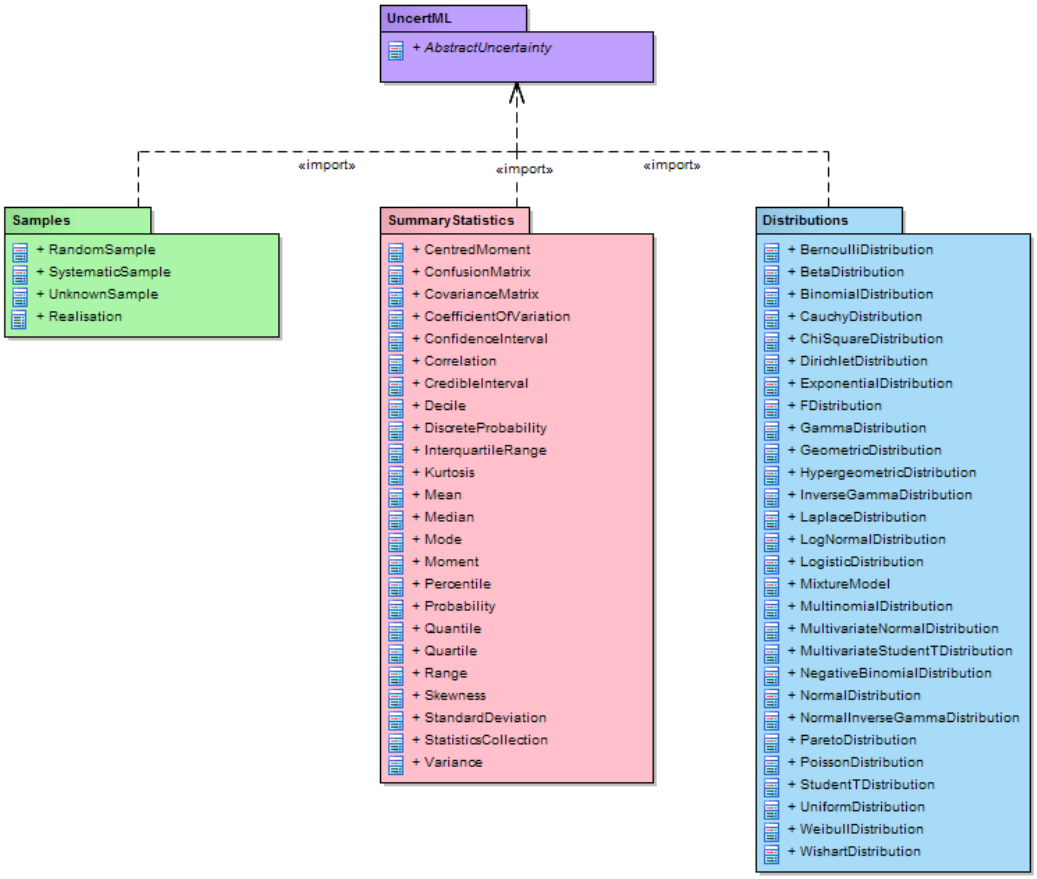

Figure 2 – UncertML 2.0 conceptual model

For the sake of generality, we allow for multiple URIs: this could support the expression of compound UncertML concepts (although this use-case is probably not so relevant, within the current version of the UncertMLdictionary).

Besides, anticipating a general support to semantic annotations in netCDF, an uncertain variable may also declare a *rel* attribute, whose value is a list of space-separated identifiers of the relationships between the variable and the corresponding URI in the *ref* attribute. In particular, the identifier for the above relationship shall be the string value "uncertainty". If the *rel* attribute is missing, all the URIs are considered uncertainty-related annotations.

An uncertainty-related URI shall be resolvable to appropriate human-readable documentation of the respective dictionary concept. The syntax and semantics of the URI should be compliant to [IETF 3986].

6.3.1 **Concept parameters**

The typical concept in the UncertML 2.0 dictionary is parameterized, with parameters of disparate types, from single scalar values to complex structures.

A complex parameter shall be encoded in a separate variable, to be declared in the *ancillary_variables* attribute of the referring concept variable, and annotated with the concept URI and the parameter name as a fragment part[3], as in the following example:

```
variables:
    float prob_greater_than_limit(Lat=163, Lon=240);
prob_greater_than_limit:ref =
"http://www.uncertml.org/statistics/probability";
prob_greater_than_limit:ancillary_variables = "limits";
float limits(Lat=163, Lon=240);
limits:ref = "http://www.uncertml.org/statistics/probability#gt";
```

For space optimization, a scalar parameter may be encoded as a variable attribute with the corresponding name, as in the following example:

```
variables:
    float second_order_moment(Lat=163, Lon=240);
second_order_moment:ref = "http://www.uncertml.org/statistics/moment";
second_order_moment:order = "2";
```

6.3.2 **Concepts without values**

Since the current implementation of the netCDF library seems to allocate space for variables as soon as they are defined (even when no data have been written), concepts whose values are irrelevant or undefined (e.g. distributions characterized by parameters only) may be encoded as scalar variables, for the sake of space optimization.

---

[3]Note that the semantics of the fragment part should be compliant to [IETF 3986]. In particular, if the referenced resource is an HTML document, the semantics of the fragment should be that of HTML anchors.

In this case, the information about the intended shape of the variable shall be preserved in the attribute *shape*, as in the example below:

```
doublebiotemperature;
biotemperature:ref = "http://www.uncertml.org/distributions/normal";
biotemperature:shape = “latlon”;
biotemperature:ancillary_variables =
"biotemperature_meanbiotemperature_variance";
```

6.4 Encoding of Samples

A sample is defined as a set of realizations. Hence, it shall be encoded as a concept without values (§6.3.2) annotated with the related concept URI, linked to data variables corresponding to its related realizations by means of the *ancillary_variables* attribute, as in the example below.

```
netcdf file:/Users/bigagli/Downloads/biotemperature_randomSamples.nc {
dimensions:
lon = 240;
lat = 163;
variables:
doublebiotemperature();
biotemperature:missing_value = -999.0; // double
biotemperature:shape = “latlon”;
biotemperature:units = "degC";
biotemperature:ref = "http://www.uncertml.org/samples/random";
biotemperature:ancillary_variables = "realisation1 realisation2";
doublelon(lon=240);
lon:units = "degrees_east";
doublelat(lat=163);
lat:units = "degrees_north";
double realisation1(lat=163, lon=240);
realisation1:missing_value = -999.0; // double
realisation1:units = "degC";
realisation1:ref = "http://www.uncertml.org/samples/realisation";
realisation1:ID = "#12345";
double realisation2(lat=163, lon=240);
realisation2:missing_value = -999.0; // double
realisation2:units = "degC";
realisation2:ref = "http://www.uncertml.org/samples/realisation";
realisation2:ID = "#12346";

:Conventions = "CF-1.5 UW-1.0";
:primary_variables = "biotemperature";
}
```


For reducing the number of variables, especially when the single realizations have the same attributes, the sample variable may group all the realizations, indexing them along an extra additional dimension. In this case, the dimension variable corresponding to the additional dimension shall be annotated as a realization, as in the example below:

```
netcdf file:/Users/bigagli/Downloads/biotemperature_randomSamples.nc {
dimensions:
lon = 240;
lat = 163;
realisation = 10;
variables:
doublebiotemperature(realization=10, lat=163, lon=240);
biotemperature:missing_value = -999.0; // double
biotemperature:shape = “latlon”;
biotemperature:units = "degC";
biotemperature:ref = "http://www.uncertml.org/samples/random";
doublelon(lon=240);
lon:units = "degrees_east";
doublelat(lat=163);
lat:units = "degrees_north";
doublerealisation(realisation=10);
realisation:ref = "http://www.uncertml.org/samples/realisation";

:Conventions = "CF-1.5 UW-1.0";
:primary_variables = "biotemperature";
}
```

6.5 Encoding of Summary Statistics

A summary statistics shall be encoded as a data variable annotated with the related concept URI. Summary statistics referring to the same stochastic variable shall be declared in the *ancillary_variables* attribute of a data variable annotated with the concept URI of a statistics collection, which may be declared in the *primary_variables* global attribute, as in the example below:

```
netcdfbiotemperature {
// global attributes:
:Conventions = "CF-1.5 UW-1.0";
:primary_variables = "biotemperature";
dimensions:
lon = 240 ;
lat = 163 ;
variables:
doublelon(lon) ;
lon:long_name = "longitude";
lon:units = "degrees_east";
```

```
lon:standard_name = "longitude";
doublelat(lat);
lat:long_name = "latitude";
lat:units = "degrees_north";
lat:standard_name = "latitude";
doublebiotemperature_mean(lat, lon);
biotemperature_mean:missing_value = -999.;
biotemperature_mean:ref = "http://www.uncertml.org/statistics/mean";
doublebiotemperature_variance(lat, lon);
biotemperature_variance:missing_value = -999.;
biotemperature_variance:ref =
"http://www.uncertml.org/statistics/variance";
doublebiotemperature(lat, lon);
biotemperature:units = "degC";
biotemperature:ancillary_variables =
"biotemperature_meanbiotemperature_variance";
biotemperature:ref = "http://www.uncertml.org/statistics/statistics-
collection";
}
```

6.6 Encoding of Distributions

A distribution shall be encoded as a concept without values (§6.3.2) annotated with the related concept URI, linked to data variables corresponding to its parameters, which shall be declared in its *ancillary_variables* attribute. The distribution data variable may be declared in the *primary_variables* global attribute, as in the example below:

```
netcdfbiotemperature {
// global attributes:
:Conventions = "CF-1.5 UW-1.0";
:primary_variables = "biotemperature";
dimensions:
lon = 240 ;
lat = 163 ;
variables:
doublelon(lon) ;
lon:long_name = "longitude";
lon:units = "degrees_east";
lon:standard_name = "longitude";
doublelat(lat);
lat:long_name = "latitude";
lat:units = "degrees_north";
lat:standard_name = "latitude";
doublebiotemperature_mean(lat, lon);
biotemperature_mean:missing_value = -999.;
biotemperature_mean:ref =
"http://www.uncertml.org/distributions/normal#mean";
doublebiotemperature_variance(lat, lon);
biotemperature_variance:missing_value = -999.;
biotemperature_variance:ref =
"http://www.uncertml.org/distributions/normal#variance";
```



```
doublebiotemperature();
biotemperature:units = "degC";
biotemperature:ancillary_variables =
"biotemperature_meanbiotemperature_variance";
biotemperature:shape = “latlon”;
biotemperature:ref = "http://www.uncertml.org/distributions/normal";
}
```

Bibliography

[1] B. Domenico and S. Nativi, OGC CF-netCDF Data Model extension specification v. 2.0 draft specification, June 2011.

[2] UncertWeb Consortium. Technical Report – NetCDF Uncertainty Conventions (NetCDF-U), version 1.0, 21 September 2011.

[3] UncertWeb Consortium. UncertML best practice proposal. Deliverable D1.2, 2011.

[4] UncertWeb Consortium. User Analysis. Deliverable D2.1, 2011.