JP1.5

CLIMATE SCIENCE MODELLING LANGUAGE: STANDARDS-BASED MARKUP FOR METOCEAN DATA

Andrew Woolf[1,*], Bryan Lawrence[2], Roy Lowry[3], Kerstin Kleese van Dam[1], Ray Cramer[3], Marta Gutierrez[2], Siva Kondapalli[3], Sue Latham[2], Kevin O'Neill[1], Ag Stephens[2]

[1] CCLRC e-Science Centre
[2] British Atmospheric Data Centre
[3] British Oceanographic Data Centre

1. INTRODUCTION

Data "interoperability" in the climate sciences is an increasingly important strategy. Whether for data assimilation, verification studies or model intercomparison projects, there is a growing need to access and integrate a range of data from multiple sources. Discovery, access and use of metocean data are all dimensions of the interoperability problem.

Achieving interoperability across these dimensions requires agreements on metadata formats, data access service interfaces, and dataset content models. Recent developments in standards for geographic information offer considerable potential for discovery and exchange of earth-related information (Woolf *et. al.*, 2004). The ISO Technical Committee 211[1] (ISO TC211) is developing a raft of specifications for digital geographic information, metadata and services. In addition, the Open Geospatial Consortium[2] (OGC) is developing a number of implementation specifications for web-enabled rendering and delivery of earth-related data.

We present here a new data model and XML markup language for the climate sciences, based on emerging ISO standards. We omit below consideration of metadata and data access services in the interoperability problem, except to note that a variety of discovery metadata may be used in conjunction with our model, and instantiation through various access services is possible. A design objective has been compatibility with the emerging (ISO and OGC) metadata and web service standards for geospatial data.

2. STANDARDS-BASED DATA MODELLING

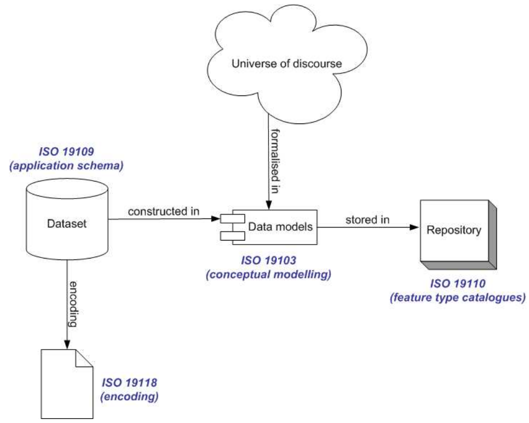

Figure 1: ISO framework for data modelling

An abstract framework for geographic data modelling has been proposed by ISO TC211 (ISO 19101). This programme is illustrated in Figure 1. Important data concepts from a universe of discourse are codified through formal data models (ISO 19103). Individual abstractions are known as "feature types" and may be catalogued for re-use in repositories ("feature type catalogues", ISO 19110). The logical structure and semantic content of datasets is then described through an *application schema* in terms of feature instances and other objects (ISO 19109). Finally, canonical encoding rules may be specified for serialisation of datasets (ISO 19118). (Alternatively, mappings may be defined from an application schema to the internal data models of third-party file formats or data delivery services.)

* *Corresponding author address*:
CCLRC Rutherford Appleton Lab
Chilton Didcot
Oxon, OX11 0QX, UK

[1] http://www.isotc211.org

[2] http://www.opengeospatial.org

A number of conceptual schemas for spatial and temporal referencing, geometry, topology, gridded data, etc., are being standardised by ISO TC211 for use in defining data models. An XML encoding for several of these is being developed as the Geography Markup Language (GML, ISO 19136). GML provides a generalised "toolkit" for defining feature types and building application schemas.

This programme offers a number of benefits for data interoperability. First, the focus is on the semantics of information, rather than its representation (content, not format). Second, governance of data models encourages a coordinated process and consensus within communities of practice. Third, the emergence of semantics repositories (feature-type catalogues) in service-oriented architectures will facilitate powerful new value-added processing chains. Finally, a standards-based framework anticipates harmonised geographic data sharing across domains and disciplines.

3. CLIMATE SCIENCE MODELLING LANGUAGE (CSML)

The Climate Science Modelling Language (CSML) has been developed as a first attempt to define structured semantic data models for the climate sciences within the standards framework described above. A prototype deployment across the curated data holdings of the British Atmospheric Data Centre and British Oceanographic Data Centre is being undertaken through the UK's NERC DataGrid project[3] (Lawrence *et. al.*, 2003). These holdings include a wide range of observational and model data. CSML is expected to evolve and develop with its application; the primary source for up-to-date information is the CSML website (http://ndg.nerc.ac.uk/csml).

As its name implies, CSML aims to provide semantically meaningful models of climate science data types. The emphasis on semantics and interoperability distinguish CSML from the XML markup representations of climate data possible with NcML[4], CDML[5] and ESML[6].

CSML has been designed explicitly with a dual purpose. In addition to modelling various climate science data types, it provides a wrapper mechanism to encapsulate the representation of those data objects in file-based storage artefacts.

CSML draws on ISO TC211 conceptual schemas for spatiotemporal referencing, geometry, etc., and employs GML for standardised encodings of these wherever possible.

3.1 CSML dataset structure

CSML represents an *application schema*, defining the semantic content and logical structure of climate science datasets. This structure is very simple and includes the following five components (all optional), as shown in Figure 2:

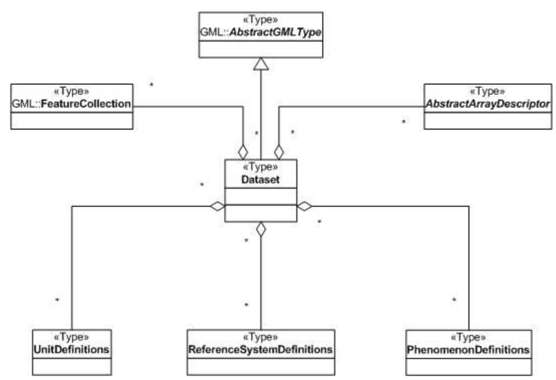

Figure 2: UML conceptual model for structure of CSML dataset

1. Local definitions of physical units,

---

[3] http://ndg.nerc.ac.uk

[4] NetCDF Markup Language, http://my.unidata.ucar.edu/content/software/netcdf/ncml/index.html

[5] Climate Data Markup Language, http://esg.llnl.gov/cdat/cdms_html/cdms-6.htm

[6] Earth Science Markup Language, http://esml.itsc.uah.edu

2. Local definitions of spatiotemporal coordinate reference systems,

3. Local definitions of 'phenomena' (the physical parameter being observed, measured, or modelled),

4. Numerical content descriptors, and

5. Collection of feature instances.

These components are described in following sections.

3.2 CSML feature types

Defined broadly as an "abstraction of real world phenomena" (ISO 19101), a feature type may represent any important aspect of the universe of discourse. As noted by ISO 19109, "(t)he classification of real world phenomena as features depends on their significance to a particular universe of discourse".

The definition of climate science feature types could take a multitude of forms. Major classes could be based on the measured parameter – for instance an abstract 'Temperature' feature type could be defined, and subclassed for potential temperature, wet-bulb temperature, etc. (This is the approach suggested in the OGC Observations and Measurements discussion paper.) Alternatively, classes could be defined on the basis of geometric and topologic structure of data types (the approach adopted here). An effective strategy for feature types is likely to be a set of comprehensive models for the full range of climate science data types. Thus, feature types would be associated with current meters, tide gauges, radiosondes, wind profilers, satellite remote sensors, numerical models, etc. It is expected that transformations would be possible from specialised to more general feature types.

The question of typing granularity for feature types is an open issue. Atkinson has suggested (Atkinson, 2004) that feature types should be defined up to the level of governance structures available to support those definitions. Thus, a full spectrum of strongly typed feature definitions as mentioned above would be defined most appropriately by a body with a significant international remit in the climate sciences. Alternatively, more specialised authorities might maintain feature type definitions for a narrower section of the climate domain (say, measurements from marine instruments).

The intent with CSML is to define a small number of feature types that have general utility across a wide range of climate science data types, as a means both of exploring the new standards-based data interoperability framework and of providing a basis for more specialised feature types later. To that end, the following principles have been adopted in defining CSML feature types:

1. ***Offloading semantics onto CRS.*** If two features would 'look the same', apart from the underlying coordinate reference systems, then they are modelled as the same feature. Thus, a vertical sounding radar and a scanning radar are modelled as the same feature type, distinguished through their respective coordinate reference systems.

2. ***Offloading semantics onto parameter type.*** If two features would 'look the same', apart from the parameter types, then they're modelled as the same feature. For instance, a vertical wind-profiler is plotted with wind barbs, while a temperature sounding is plotted as a line – these are regarded as the same feature type, distinguished by the parameter type.

3. ***Sensible plotting as discriminant.*** A principle to suggest a workable minimum granularity of feature type definition is to use the requirement for 'sensible plotting' as a discriminant. That is, there should be sufficient detail in, and sufficient difference between, feature types to enable unsupervised 'sensible plotting' by an appropriate piece of software.

As a result of the above principles, CSML feature types are defined primarily on the basis of geometric and topologic structure, and not the semantics of the observable or measurand. Seven feature types are defined, listed here in Table 1.

CSML feature type	*Description*	*Examples*
TrajectoryFeature	Discrete path in time and space of a platform or instrument.	ship's cruise track, aircraft's flight path
PointFeature	Single point measurement.	raingauge measurement
ProfileFeature	Single 'profile' of some parameter along a directed line in space.	wind sounding, XBT, CTD, radiosonde
GridFeature	Single time-snapshot of a gridded field.	gridded analysis field
PointSeriesFeature	Series of single datum measurements.	tidegauge, rainfall timeseries
ProfileSeriesFeature	Series of profile-type measurements.	vertical or scanning radar, shipborne ADCP, thermistor chain timeseries
GridSeriesFeature	Timeseries of gridded parameter fields.	numerical weather prediction model, ocean general circulation model

Table 1: CSML feature types

These feature types do not carry any explicit topologic description currently. For instance, both a scanning radar and a vertical sounding radar are modelled with the same feature type (ProfileSeries). The spatial geometry of the first may be modelled topologically as a series of straight curves connected at an origin node, while the second is a single straight curve (repeated in time). Similarly, the topological relationship between a series of marine CTD casts and the associated ship track is left implicit in the ProfileSeries feature type. GML provides rich schemas for describing topology – a requirement may arise in the future to add this information to CSML.

The CSML feature types fall into two broad categories – 'geometry' and 'coverage'. Only one geometry feature type is defined currently – the TrajectoryFeature. The remainder all are coverage feature types. A coverage is a "feature that acts as a function to return values from its range for any direct position within its spatiotemporal domain" (ISO 19123), Figure 3.

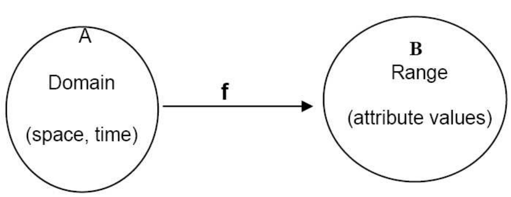

Figure 3: Conceptual model of a coverage ***(from Geography Markup Language (ISO/CD 19136, 2004))***

All CSML coverage feature types have the following properties:

- **parameter**: the physical parameter (temperature, wind-direction, salinity etc.) associated with the feature instance
- **domainSet**: locations of data points in the CSML feature
- **rangeSet**: physical parameter values for the CSML feature
- **coverageFunction**: the mapping sequence between the domain locations and range values

3.3 CSML units, phenomena, and reference system definitions

The physical units, parameter types, and coordinate reference systems associated with a feature instance are all namespace-qualified. Thus, multiple controlled vocabularies may be used. Local definitions within a dataset may be used when a physical unit, parameter type or reference system is not defined elsewhere in a reference-able manner.

GML provides XML markup for defining coordinate reference systems (ISO/CD 19136, 2004 §12,14.4) and dictionaries (§15.2). For instance, Figure 4 shows an XML snippet for a

GML encoding of the 'udunits' dictionary[7] widely used in atmospheric science.

The encoding of parameter types is based on the conceptual model for 'phenomena' from the OGC Observations and Measurements discussion paper (OGC, 2003). This supports the following classes of observable:

- **Phenomenon**: a basic definition of a phenomenon, characterised by an identifier, description and namespace-qualified name
- **CompositePhenomenon**: for example vector winds. Each component is itself a Phenomenon.
- **ParametrisedPhenomenon**: a phenomenon with multiple components derived by applying constraints to some base phenomenon; for example, multi-spectral radiance bands apply wavelength constraints to a base 'radiance' phenomenon; particle size distributions are classified by bands of particle size range

An example of parameter type (phenomena) encodings is shown in Figure 5, which represents a GML translation of a small part of the 'Standard Names Table' from the NetCDF Climate and Forecast Metadata Convention[8] which is widely used in the climate modelling community.

Examples of reference system encodings are available in GML (ISO/CD 19136, 2004) and at the CSML website.

3.4 CSML numeric content descriptors

CSML is intended to provide a data storage 'wrapper' in addition to modelling the semantic content and logical structure of datasets. This is done by allowing any numerical content for feature instances to be replaced with descriptors for file-based storage. The current mechanism supports only logical numerical arrays, but this is sufficient for the CSML feature types defined above.

The following four array descriptor subclasses are defined:

- **InlineArray**: an inline character encoding of array values
- **ArrayGenerator**: a formulaic expression implicitly generating array values
- **AbstractFileExtract**: values extracted from file-based storage
- **AggregatedArray**: an aggregation of two or more arrays along an existing or new dimension

In addition a numeric or regular expression transformation can be applied to any of the above as appropriate.

Examples are provided in Figure 6, and the UML conceptual model is shown in Figure 7.

---

[7] http://my.unidata.ucar.edu/content/software/udunits/index.html

[8] http://www.cgd.ucar.edu/cms/eaton/cf-metadata/

```xml
<Dataset xmlns="http://ndg.nerc.ac.uk/csml" xmlns:gml="http://www.opengis.net/gml"
xmlns:om="http://www.opengis.net/om" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink">
    <UnitDefinitions gml:id="udunits">
        <gml:name>udunits</gml:name>
        <gml:dictionaryEntry>
            <gml:BaseUnit gml:id="ampere">
                <gml:name>ampere</gml:name>
                <gml:quantityType>electric current</gml:quantityType>
            <gml:catalogSymbol
codeSpace="http://www.unidata.ucar.edu/packages/udunits">ampere</gml:catalogSymbol>
                <gml:unitsSystem xlink:href="http://www.unidata.ucar.edu/packages/udunits"/>
            </gml:BaseUnit>
        </gml:dictionaryEntry>
        <gml:dictionaryEntry>
            <gml:BaseUnit gml:id="kelvin">
                <gml:name>kelvin</gml:name>
                <gml:name>degree_Kelvin</gml:name>
                <gml:name>degree_Kelvin</gml:name>
                <gml:name>degK</gml:name>
                <gml:name>degreeK</gml:name>
                <gml:name>degree_K</gml:name>
                <gml:name>degree_k</gml:name>
                <gml:name>deg_K</gml:name>
                <gml:name>deg_k</gml:name>
                <gml:name>K</gml:name>
                <gml:name>Kelvin</gml:name>
                <gml:quantityType>thermodynamic temperature</gml:quantityType>
            <gml:catalogSymbol
codeSpace="http://www.unidata.ucar.edu/packages/udunits">kelvin</gml:catalogSymbol>
                <gml:unitsSystem xlink:href="http://www.unidata.ucar.edu/packages/udunits"/>
            </gml:BaseUnit>
        </gml:dictionaryEntry>
        <gml:dictionaryEntry>
            <gml:ConventionalUnit gml:id="degree_Celsius">
                <gml:name>degree_Celsius</gml:name>
                <gml:name>Celsius</gml:name>
                <gml:name>celsius</gml:name>
                <gml:name>degree_centigrade</gml:name>
                <gml:name>degC</gml:name>
                <gml:name>degreeC</gml:name>
                <gml:name>degree_C</gml:name>
                <gml:name>degree_c</gml:name>
                <gml:name>deg_C</gml:name>
                <gml:name>deg_c</gml:name>
                <gml:quantityType>THERMODYNAMIC TEMPERATURE</gml:quantityType>
                <gml:conversionToPreferredUnit uom="degree_Kelvin">
                    <gml:formula>
                        <gml:a>273.15</gml:a>
                        <gml:b>1</gml:b>
                        <gml:c>1</gml:c>
                    </gml:formula>
                </gml:conversionToPreferredUnit>
            </gml:ConventionalUnit>
        </gml:dictionaryEntry>
        <!-- ... -->
    </UnitDefinitions>
</Dataset>
```


Figure 4: Example GML encoding of 'udunits' dictionary

```xml
<?xml version="1.0" encoding="UTF-8"?>
<Dataset xmlns="http://ndg.nerc.ac.uk/csml" xmlns:gml="http://www.opengis.net/gml"
xmlns:om="http://www.opengis.net/om" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink">
	<PhenomenonDefinitions gml:id="CFStandardNames">
		<gml:name>CFStandardNames</gml:name>
		<gml:dictionaryEntry>
			<om:Phenomenon gml:id="air_potential_temperature">
				<gml:description>Potential temperature is the temperature a parcel of air or sea water would have if
moved adiabatically to sea level pressure.</gml:description>
				<gml:name codeSpace="http://www.cgd.ucar.edu/cms/eaton/cf-
metadata/">air_potential_temperature</gml:name>
				<gml:name codeSpace="GRIB">13</gml:name>
				<gml:name codeSpace="PCMDI">theta</gml:name>
			</om:Phenomenon>
		</gml:dictionaryEntry>
		<gml:dictionaryEntry>
			<om:Phenomenon gml:id="atmosphere_cloud_condensed_water_content">
				<gml:description>"condensed_water" means liquid and ice. "Content" indicates a quantity per unit area.
The "atmosphere content" of a quantity refers to the vertical integral from the surface to the top of the atmosphere. For
the content between specified levels in the atmosphere, standard names including content_of_atmosphere_layer are
used.</gml:description>
				<gml:name codeSpace="http://www.cgd.ucar.edu/cms/eaton/cf-
metadata/">atmosphere_cloud_condensed_water_content</gml:name>
				<gml:name codeSpace="GRIB">76</gml:name>
				<gml:name codeSpace="PCMDI">clwvi</gml:name>
			</om:Phenomenon>
		</gml:dictionaryEntry>
		<gml:dictionaryEntry>
			<om:Phenomenon gml:id="baroclinic_eastward_sea_water_velocity">
				<gml:description>A velocity is a vector quantity. "Eastward" indicates a vector component which is
positive increasing eastward (negative westward).</gml:description>
				<gml:name codeSpace="http://www.cgd.ucar.edu/cms/eaton/cf-
metadata/">baroclinic_eastward_sea_water_velocity</gml:name>
			</om:Phenomenon>
		</gml:dictionaryEntry>
		<gml:dictionaryEntry>
			<om:Phenomenon gml:id="brightness_temperature">
				<gml:description>The brightness temperature of a body is the temperature of a black body which
radiates the same power per unit solid angle per unit area.</gml:description>
				<gml:name codeSpace="http://www.cgd.ucar.edu/cms/eaton/cf-
metadata/">brightness_temperature</gml:name>
				<gml:name codeSpace="GRIB">118</gml:name>
			</om:Phenomenon>
		</gml:dictionaryEntry>
	</PhenomenonDefinitions>
</Dataset>
```

Figure 5: Example GML translation of part of 'CF Standard Names Table'

```xml
<InlineArray>
      <arraySize>5 2</arraySize>
      <uom>udunits.xml#degreeC</uom>
      <numericType>float</numericType>
      <regExpTransform>s/10/9/ge</regExpTransform>
      <numericTransform>+5</numericTransform>
      <values>1 2 3 4 5 6 7 8 9 10</values>
</InlineArray>

<ArrayGenerator>
      <arraySize>10001</arraySize>
      <uom>udunits.xml#minute</uom>
      <numericType>float</numericType>
      <expression>0:5:50000</expression>
</ArrayGenerator>

<NASAAmesExtract>
      <arraySize>526</arraySize>
      <numericType>double</numericType>
      <fileName>/data/BADC/macehead/mh960607.asv</fileName>
      <variableName>coefficient a1</variableName>
</NASAAmesExtract>

<GRIBExtract>
      <arraySize>320 160</arraySize>
      <numericType>double</numericType>
      <fileName>/e40/ggas1992010100rsn.grb</fileName>
      <parameterCode>203</parameterCode>
      <recordNumber>5</ recordNumber>
      <fileOffset>289412</fileOffset>
</GRIBExtract>

<AggregatedArray gml:id="feat05cruisetrack">
      <arraySize>2 50</arraySize>
      <aggType>new</aggType>
      <aggIndex>1</aggIndex>
      <component>
            <NetCDFExtract>
                  <arraySize>50</arraySize>
                  <fileName>cruisetrack.nc</fileName>
                  <variableName>alat</variableName>
            </NetCDFExtract>
      </component>
      <component>
            <NetCDFExtract>
                  <arraySize>50</arraySize>
                  <fileName>cruisetrack.nc</fileName>
                  <variableName>alon</variableName>
            </NetCDFExtract>
      </component>
</AggregatedArray>
```

Figure 6: Example CSML encoding of numeric content descriptors

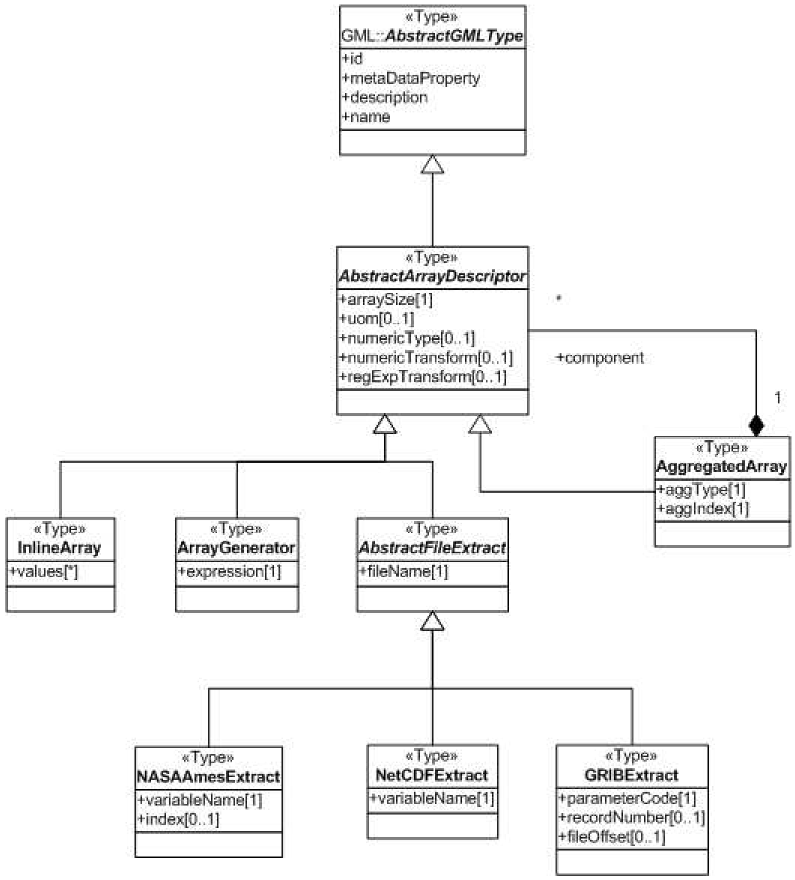

Figure 7: UML conceptual model for CSML numeric content descriptors

4. APPLICATION

The initial prototyping of CSML is being applied to a variety of data types across the curated holdings of the British Atmospheric Data Centre (BADC) and British Oceanographic Data Centre (BODC). A parsing and processing software suite is being written concurrently. Software releases will be available through the CSML website.

This work is being undertaken as part of the UK project "NERC DataGrid" (NDG), which is building a data-grid infrastructure to provide seamless discovery and access to a range of earth-science data in the UK. CSML will be used as an abstraction layer for climate-science data, with NDG data delivery services layered over it. These will include file instantiation (e.g. CF-compliant netCDF files), and delivery via OPeNDAP and the mapping and data web services of the Open Geospatial Consortium (Web {Map, Feature, Coverage} Server). A file instantiation service has been implemented for an earlier version of the NDG data model (Woolf *et. al.*, 2003).

5. CONCLUSIONS AND FUTURE

A number of benefits to a standards-based data infrastructure have been suggested. These include a focus on the semantics of data, and improved scope for interoperability.

We have presented the Climate Science Modelling Language (CSML) as an attempt to apply the emerging ISO standards for geographic information to climate science data. A number of feature types have been identified and implemented as an application schema of the Geography Markup Language (GML).

Several issues have been identified in the process. For instance, only regularly-spaced grids are supported, and work is needed to characterise the range of vertical coordinate systems used with climate data. We shall seek and recommend improvements to the emerging standards to try to ensure they meet the requirements of the climate community. (CSML, for instance, includes an extension to GML grids to provide for arbitrary node locations.)

In addition to describing the logical structure of datasets, CSML incorporates a wrapper mechanism to 'connect' with file-based storage. Aggregation and limited transformations are supported. CSML is a work-in-progress and this, as with other CSML elements, will be evaluated and improved with deployment experience.

Immediate plans for CSML include stabilising the feature type definitions and schemas through application across BADC and BODC data stores, implementing software tooling (including NDG delivery services), normalisation against GML and other ISO standards as they emerge and stabilise, and publishing associated standards-compliant, reference-able dictionaries. (The latter may include, for instance, a 'udunits'-compatible GML units dictionary and a GML climate sciences coordinate-reference system repository.) CSML schemas will be lodged in appropriate ISO-compliant feature type catalogues where possible. (ISO standards issues are being considered by both the WMO and IOC.)

CSML has already been used by the EU MarineXML project to provide feature types for a testbed exercise on mappings against the new GML encoding of IHO's S-57 navigational charts.

6. ACKNOWLEDGEMENTS

This work was undertaken by the NERC DataGrid project funded under the UK e-Science program through grant NER/T/S/2002/00091 from the Natural Environment Research Council.

7. REFERENCES

Atkinson, R., *et. al.*, 2004: Next steps to interoperability – Mechanisms for semantic interoperability, *EOGEO Workshop*, University College London, 23-25 June 2004 [Online: http://y.eogeo.org/files/eogeo2004/eogeo-2004-full-atkinson.ppt, 8 November, 2004].

ISO 19101, 2002: Geographic information – Reference model.

ISO/DTS 19103, 2003: Geographic information – Conceptual schema language.

ISO/FDIS 19109, 2003: Geographic information – Rules for application schema.

ISO/DIS 19110, 2004: Geographic information – Methodology for feature cataloguing.

ISO/DIS 19118, 2003: Geographic information – Encoding.

ISO/DIS 19123, 2004: Geographic information – Schema for coverage geometry and functions.

ISO/CD 19136, 2004: Geographic information – Geography Markup Language [Online: http://portal.opengis.org/files/?artifact_id=4700, 8 November 2004]

Lawrence, B.N., *et. al.*, 2003: The NERC DataGrid Prototype, *Proceedings of the U.K. e-science All Hands Meeting*, S.J.Cox(Ed) ISBN 1-904425-11-9 [Online: http://epubs.cclrc.ac.uk/work-details?w=30086, 8 November 2004]

OGC, 2003: Observations and Measurements, *OGC 03-022r3, Open Geospatial Consortium* [Online: http://www.opengeospatial.org/docs/03-022r3.pdf, 8 November 2004]

Woolf, A., *et. al.*, 2003: Data Virtualisation in the NERC DataGrid, *Proceedings of the U.K. e-science All Hands Meeting*, S.J.Cox(Ed) ISBN 1-904425-11-9 [Online: http://epubs.cclrc.ac.uk/work-details?w=30104, 8 November 2004]

Woolf, A., *et. al.*, 2004: Standards-based data interoperability for the climate sciences, *European Geosciences Union, 1st General Assembly*, Nice (J. Met. Apps. submitted).

APPENDIX: EXAMPLE CSML DATASET ENCODING

Listed below is an example hypothetical CSML dataset containing some unit, reference system and phenomenon definitions, a number of numeric array descriptors, and five feature instances: a thermometer temperature reading (PointFeature), raingauge timeseries (PointSeries), wind profile (Profile), scanning radar timeseries (ProfileSeries), and cruise CTD salinity (ProfileSeries).

This is for illustration purposes only, as schema details are likely to change – latest versions are available from the CSML website (http://ndg.nerc.ac.uk/csml).

```xml
<?xml version="1.0" encoding="UTF-8"?>
<Dataset xmlns="http://ndg.nerc.ac.uk/csml" xmlns:gml="http://www.opengis.net/gml"
xmlns:om="http://www.opengis.net/om" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink" gml:id="TestDataset">
    <gml:metaDataProperty xlink:href="http://ndg.nerc.ac.uk/Metadata/TestDataset"/>
    <gml:description>Test CSML Dataset.</gml:description>
    <gml:name codeSpace="http://ndg.nerc.ac.uk/NDGData">DataEntity0000001</gml:name>
    <!--===================================================================-->
    <!--===== Local unit definitions. =====-->
    <!--===================================================================-->
    <UnitDefinitions gml:id="UnitDefs">
        <gml:name codeSpace="http://ndg.nerc.ac.uk">TestDatasetUnitDefs</gml:name>
        <gml:definitionMember>
            <gml:UnitDefinition gml:id="psu">
                <gml:description>Conventional practical salinity units.</gml:description>
                <gml:name>practical salinity units</gml:name>
                <gml:quantityType>sea water salinity</gml:quantityType>
                <gml:catalogSymbol codeSpace="http://ndg.nerc.ac.uk/units">psu</gml:catalogSymbol>
            </gml:UnitDefinition>
        </gml:definitionMember>
    </UnitDefinitions>
    <!--===================================================================-->
    <!--===== Local reference system definitions. =====-->
    <!--===================================================================-->
    <ReferenceSystemDefinitions gml:id="RefSysDefs">
        <gml:name codeSpace="http://ndg.nerc.ac.uk">TestDatasetRefSysDefs</gml:name>
        <gml:definitionMember>
            <TimeCoordinateSystem gml:id="RefSys01">
                <gml:description>Temporal reference system representing hours since midnight January 1,
2000.</gml:description>
                <gml:name>RaingaugeTimeRefSys</gml:name>
                <gml:originPosition>2001-01-01T00:00:00</gml:originPosition>
                <gml:interval unit="hour">1</gml:interval>
            </TimeCoordinateSystem>
        </gml:definitionMember>
        <gml:definitionMember>
            <gml:EngineeringCRS gml:id="RefSys02">
                <gml:remarks>A reference system representing linear distance in metres along some
path.</gml:remarks>
                <gml:srsName>RangeRefSys</gml:srsName>
                <gml:usesCS>
                    <gml:LinearCS gml:id="RefSys02Coords">
                        <gml:remarks>The coordinate system is linear with a single axis.</gml:remarks>
                        <gml:csName>RangeCoords</gml:csName>
                        <gml:usesAxis>
                            <gml:CoordinateSystemAxis gml:id="RefSys02CoordAxis1" gml:uom="udunits.xml#m">
                                <gml:axisName>range</gml:axisName>
                                <gml:axisAbbrev>r</gml:axisAbbrev>
                                <gml:axisDirection>Positive away from origin.</gml:axisDirection>
                            </gml:CoordinateSystemAxis>
                        </gml:usesAxis>
                    </gml:LinearCS>
                </gml:usesCS>
                <gml:usesEngineeringDatum>
```

```xml
					<gml:EngineeringDatum gml:id="RefSys02Datum">
						<gml:remarks>Defines positions along profile relative to origin of profiling instrument (radar,
CTD, etc).</gml:remarks>
						<gml:datumName>Profiler origin</gml:datumName>
					</gml:EngineeringDatum>
				</gml:usesEngineeringDatum>
			</gml:EngineeringCRS>
		</gml:definitionMember>
		<gml:definitionMember>
			<TimeCoordinateSystem gml:id="RefSys03">
				<gml:description>Temporal reference system representing five minute intervals since 08:30, 2 June,
2001.</gml:description>
				<gml:name>RadarTimeRefSys</gml:name>
				<gml:originPosition>2001-06-02T08:30:00</gml:originPosition>
				<gml:interval unit="minute">5</gml:interval>
			</TimeCoordinateSystem>
		</gml:definitionMember>
		<gml:definitionMember>
			<TimeCoordinateSystem gml:id="RefSys04">
				<gml:description>Temporal reference system representing day fractions since 12 May,
1999.</gml:description>
				<gml:name>CruiseTimeRefSys</gml:name>
				<gml:originPosition>1999-05-12</gml:originPosition>
				<gml:interval unit="day">1</gml:interval>
			</TimeCoordinateSystem>
		</gml:definitionMember>
		<gml:definitionMember>
			<gml:EngineeringCRS gml:id="RefSys05">
				<gml:remarks>A reference system representing depth pressures in decibar</gml:remarks>
				<gml:srsName>OceanPressureRefSys</gml:srsName>
				<gml:usesCS>
					<gml:LinearCS gml:id="RefSys05Coords">
						<gml:remarks>The coordinate system is linear with a single axis.</gml:remarks>
						<gml:csName>DepthCoords</gml:csName>
						<gml:usesAxis>
							<gml:CoordinateSystemAxis gml:id="RefSys05CoordAxis1" gml:uom="udunits.xml#dbar">
								<gml:axisName>depth</gml:axisName>
								<gml:axisAbbrev>r</gml:axisAbbrev>
								<gml:axisDirection>Positive downwards.</gml:axisDirection>
							</gml:CoordinateSystemAxis>
						</gml:usesAxis>
					</gml:LinearCS>
				</gml:usesCS>
				<gml:usesEngineeringDatum>
					<gml:EngineeringDatum gml:id="RefSys05Datum">
						<gml:remarks>Defines depths in pressure units from sea surface.</gml:remarks>
						<gml:datumName>Sea surface</gml:datumName>
						<gml:anchorPoint>0</gml:anchorPoint>
					</gml:EngineeringDatum>
				</gml:usesEngineeringDatum>
			</gml:EngineeringCRS>
		</gml:definitionMember>
	</ReferenceSystemDefinitions>
	<!--=================================================================-->
	<!--===== Local phenomenon definitions. =====-->
	<!--=================================================================-->
	<PhenomenonDefinitions gml:id="PhenDefs">
		<gml:name>TestDatasetPhenDefs</gml:name>
		<gml:definitionMember>
			<om:Phenomenon gml:id="rainfall">
				<gml:description>Liquid precipitation measured with raingauge.</gml:description>
				<gml:name codeSpace="http://ndg.nerc.ac.uk">rainfall</gml:name>
			</om:Phenomenon>
		</gml:definitionMember>
	</PhenomenonDefinitions>
```

```xml
<!--=================================================================-->
<!--===== Indirect value array definitions. =====-->
<!--=================================================================-->
<NetCDFExtract gml:id="feat04times">
	<arraySize>10000</arraySize>
	<fileName>radar_data.nc</fileName>
	<variableName>time</variableName>
</NetCDFExtract>
<NetCDFExtract gml:id="feat04azimuth">
	<arraySize>10000</arraySize>
	<fileName>radar_data.nc</fileName>
	<variableName>az</variableName>
</NetCDFExtract>
<AggregatedArray gml:id="feat05cruisetrack">
	<arraySize>2 50</arraySize>
	<aggType>new</aggType>
	<aggIndex>1</aggIndex>
	<component>
		<NetCDFExtract>
			<arraySize>50</arraySize>
			<fileName>cruisetrack.nc</fileName>
			<variableName>alat</variableName>
		</NetCDFExtract>
	</component>
	<component>
		<NetCDFExtract>
			<arraySize>50</arraySize>
			<fileName>cruisetrack.nc</fileName>
			<variableName>alon</variableName>
		</NetCDFExtract>
	</component>
</AggregatedArray>
<NetCDFExtract gml:id="feat05times">
	<arraySize>50</arraySize>
	<fileName>cruisetrack.nc</fileName>
	<variableName>atime</variableName>
</NetCDFExtract>
<ArrayGenerator gml:id="feat05depths">
	<arraySize>3001</arraySize>
	<expression>0:2:6000</expression>
</ArrayGenerator>
<!--=================================================================-->
<!--===== Feature instances =====-->
<!--=================================================================-->
<gml:FeatureCollection>
	<gml:boundedBy>
		<gml:EnvelopeWithTimePeriod srsName="urn:EPSG:geographicCRS:4326">
			<gml:lowerCorner>-10 15</gml:lowerCorner>
			<gml:upperCorner>30 65</gml:upperCorner>
			<gml:timePosition>1998-01-01</gml:timePosition>
			<gml:timePosition>2003-12-31</gml:timePosition>
		</gml:EnvelopeWithTimePeriod>
	</gml:boundedBy>
	<gml:featureMember>
		<PointFeature gml:id="feat01">
			<gml:description>Temperature reading from thermometer.</gml:description>
			<PointDomain>
				<domainReference>
					<Position srsName="urn:EPSG:geographicCRS:4979" axisLabels="Lat Long h"
uomLabels="degree degree m">
						<location>0.1 1.5 25</location>
						<time>2000-08-13T13:51:10</time>
					</Position>
				</domainReference>
			</PointDomain>
```

```xml
                <gml:rangeSet>
                    <gml:QuantityList uom="udunits.xml#degreesC">10</gml:QuantityList>
                </gml:rangeSet>
                <parameter xlink:href="CFStandardNames.xml#air_temperature"/>
            </PointFeature>
        </gml:featureMember>
        <gml:featureMember>
            <PointSeriesFeature gml:id="feat02">
                <gml:description>January timeseries of raingauge measurements</gml:description>
                <PointSeriesDomain>
                    <domainReference>
                        <Trajectory srsName="urn:EPSG:geographicCRS:4979">
                            <locations>0.1 1.5 25</locations>
                            <times frame="#RefSys01">0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
24 2 26 27 28 29 30 31</times>
                        </Trajectory>
                    </domainReference>
                </PointSeriesDomain>
                <gml:rangeSet>
                    <gml:QuantityList uom="udunits.xml#mm">5 3 10 1 2 8 10 2 5 10 20 21 12 3 5 19 12 23 32 10 8 8 2
0 0 1 5 6 10 17 20</gml:QuantityList>
                </gml:rangeSet>
                <parameter xlink:href="#rainfall"/>
            </PointSeriesFeature>
        </gml:featureMember>
        <gml:featureMember>
            <ProfileFeature gml:id="feat03">
                <gml:description>Vertical wInd profile</gml:description>
                <ProfileDomain>
                    <domainReference>
                        <OrientedPosition srsName="urn:EPSG:geographicCRS:4326">
                            <location>-1 3</location>
                            <time>1999-07-21T10:00:00</time>
                            <direction>
                                <gml:horizontalAngle uom="udunits.xml#degrees">0</gml:horizontalAngle>
                                <gml:verticalAngle uom="udunits.xml#degrees">90</gml:verticalAngle>
                            </direction>
                        </OrientedPosition>
                    </domainReference>
                    <domainComplement>
                        <DirectPositionList srsName="#RefSys02">10 20 30 40 50 60 70 80 90
100</DirectPositionList>
                    </domainComplement>
                </ProfileDomain>
                <gml:rangeSet>
                    <gml:DataBlock>
                        <gml:rangeParameters>
                            <gml:CompositeValue>
                                <gml:valueComponents>
                                    <gml:measure uom="udunits.xml#ms-1"/>
                                    <gml:measure uom="udunits.xml#ms-1"/>
                                </gml:valueComponents>
                            </gml:CompositeValue>
                        </gml:rangeParameters>
                        <gml:doubleOrNullTupleList>2.0 1.1 2.4 0.8 3.3 0.1 2.6 -0.2 5.6 0.1 4.5 1.3 6.3 0.1 5.7 0.9 4.2
1.1 3.2 -0.1</gml:doubleOrNullTupleList>
                    </gml:DataBlock>
                </gml:rangeSet>
                <parameter>
                    <om:CompositePhenomenon gml:id="wind">
                        <gml:name>Vector wind</gml:name>
                        <om:componentPhenomenon xlink:href="CFStandardNames.xml#eastward_wind"/>
                        <om:componentPhenomenon xlink:href="CFStandardNames.xml#northward_wind"/>
                    </om:CompositePhenomenon>
                </parameter>
```

```xml
            </ProfileFeature>
        </gml:featureMember>
        <gml:featureMember>
            <ProfileSeriesFeature gml:id="feat04">
                <gml:description>Scanning radar timeseries.</gml:description>
                <ProfileSeriesDomain>
                    <domainReference>
                        <OrientedTrajectory srsName="urn:EPSG:geographicCRS:4326">
                            <locations>-0.5 2.1</locations>
                            <times xlink:href="#feat04times" frame="#RefSys03"/>
                            <directions>
                                <horizontalAngles uom="udunits.xml#degree" xlink:href="#feat04azimuth"/>
                                <verticalAngles uom="udunits.xml#degree">23</verticalAngles>
                            </directions>
                        </OrientedTrajectory>
                    </domainReference>
                    <domainComplement>
                        <DirectPositionList srsName="#RefSys02">100 200 300 400 500 600 700 800 900
1000</DirectPositionList>
                    </domainComplement>
                </ProfileSeriesDomain>
                <gml:rangeSet>
                    <NetCDFExtract>
                        <arraySize>10 10000</arraySize>
                        <uom>udunits.xml#percent</uom>
                        <fileName>radar_data.nc</fileName>
                        <variableName>hum</variableName>
                    </NetCDFExtract>
                </gml:rangeSet>
                <gml:coverageFunction>
                    <MappingRule scanOrder="+gridI +series"/>
                </gml:coverageFunction>
                <parameter xlink:href="XFStandardNames.xml#relative_humidity"/>
            </ProfileSeriesFeature>
        </gml:featureMember>
        <gml:featureMember>
            <ProfileSeriesFeature gml:id="feat05">
                <gml:description>Cruise CTD salinity</gml:description>
                <ProfileSeriesDomain>
                    <domainReference>
                        <OrientedTrajectory srsName="urn:EPSG:geographicCRS:4326">
                            <locations order="byCoords" xlink:href="#feat05cruisetrack"/>
                            <times xlink:href="#feat05times" frame="#RefSys04"/>
                            <directions>
                                <horizontalAngles uom="udunits.xml#degree">0</horizontalAngles>
                                <verticalAngles uom="udunits.xml#degree">-90</verticalAngles>
                            </directions>
                        </OrientedTrajectory>
                    </domainReference>
                    <domainComplement>
                        <DirectPositionList srsName="#RefSys05" xlink:href="#feat05depths"/>
                    </domainComplement>
                </ProfileSeriesDomain>
                <gml:rangeSet>
                    <AggregatedArray>
                        <arraySize>3001 50</arraySize>
                        <uom>#psu</uom>
                        <aggType>new</aggType>
                        <aggIndex>1</aggIndex>
                        <component>
                            <NetCDFExtract>
                                <arraySize>50</arraySize>
                                <fileName>cruisestation0??.nc</fileName>
                                <variableName>ctdsal</variableName>
                            </NetCDFExtract>
```

```
                        </component>
                    </AggregatedArray>
                </gml:rangeSet>
                <gml:coverageFunction>
                    <MappingRule scanOrder="+gridI +series"/>
                </gml:coverageFunction>
                <parameter xlink:href="http://bodc.ac.uk/PSAL"/>
            </ProfileSeriesFeature>
        </gml:featureMember>
    </gml:FeatureCollection>
</Dataset>
```