Design and implementation of netCDF Markup Language (NcML) and Its GML-based extension (NcML-G_{ML})

Stefano Nativi*, John Caron**, Ethan Davis**, and Ben Domenico**

* University of Florence at Prato
Piazza Ciardi, 25
59100 Prato – Italy
and
IMAA – Italian National Research Council (CNR)
stefano.nativi@pin.unifi.it

** Unidata Program Center
University Corporation for Atmospheric Research
Boulder, CO 80307 - USA
{caron, edavis, ben}@ucar.edu

Abstract

The Network Common Data Form (netCDF) is one of the primary methods of self-documenting data storage and access in the international geosciences research and education community and beyond. NetCDF was designed for use in a networked environment. The recent evolution toward web services approaches to data exchange has focused attention on communication via messages in the defacto standard XML language. XML is a text-based language while netCDF is based on a binary file storage mechanism; thus NcML is a natural augmentation of netCDF with extensions encapsulating descriptions of the structure and content of netCDF objects in an XML form. Since netCDF was designed to be self-documenting, the XML representation of internal netCDF documentation is a natural augmentation of the original netCDF concept. In fact, the netCDF Markup Language (NcML) and NcML-G (NcML-Geography) extensions described in this article have applications beyond merely representing the internal netCDF documentation in the XML language. The NcML coordinate system makes it possible to describe the coordinate system used to represent the netCDF dataset. Furthermore the NcML dataset is a tool for

describing “virtual netCDF” files that may be aggregations of data from several existing netCDF files, or it can represent a target dataset to be created by transforming existing netCDF files into a new form described in the NcML language. The NcML-G extension provides a means for fusing the data models of the traditional netCDF atmospheric science community with those of the GIS community which is of the utmost importance. Bringing the data models and data systems of those communities together will foster an era of interdisciplinary research and education within the geosciences subdisciplines. It will also encourage closer interactions between the geosciences and the societal impacts community. The design and software implementation of the core NcML specification and its extensions are presented and discussed.

Keywords: scientific data markup language, netCDF, GML

1 Introduction

The Network Common Data Form (netCDF) software is widely used throughout the world as a mechanism for storing and accessing scientific data--especially datasets related to the geosciences. The Unidata netCDF web pages [1] list several means for estimating netCDF usage: for example, the netcdfgroup@unidata.ucar.edu mailing list currently contains over 500 addresses in 32 countries. Over 2000 distinct hosts in 55 countries have downloaded the netCDF software distribution since May 1997.

A few of the earth science studies institutions that employ netCDF are: NOAA's Climate Diagnostics Center (CDC), NASA's Halogen Occultation Experiment (HALOE), The global ocean modeling effort at Los Alamos National Laboratory (LANL), Lamont-Doherty Earth Observatory (LDEO) of Columbia University, the National Center for Atmospheric Research (NCAR), the Commonwealth Scientific and Industrial Research Organization (CSIRO) Division of Atmospheric Research in Australia, The Earth Scan Lab, a High-Resolution Picture Transmission (HRPT) ground station at the Coastal Studies Institute, Scripps Institution of Oceanography (SIO), and Sandia National Laboratory.

Beyond the research and education communities, several commercial analysis and data visualization packages have been adapted to access netCDF data. Moreover, netCDF is the vehicle adopted by the Analytical Instrument Association (AIA) to implement the Analytical Data Interchange Protocols for chromatography and mass spectrometry. In addition, the Positron Imaging Laboratories and the Neuro-Imaging Laboratory of the Montreal Neurological Institute have selected netCDF as the data format for their medical image files.

The advent of the web services approach to making data available has increased the emphasis on eXtensible Markup Language (XML) as a means for conveying data and information about data available on the web. NcML was developed to encode information about (though not the data contained by) netCDF files and provide a standard XML dialect by which this information can be shared. NcML does not encode the data from a netCDF file only the metadata about a netCDF file. Thus, NcML provides a powerful complement to the self-documenting nature of netCDF which employs binary file formats and transport mechanisms. NcML has evolved beyond its original goal of simply describing netCDF files, introducing extensions to: explicitly encode part of dataset semantics, generate virtual datasets, mediate with GIS community semantics. NcML-G is an extension of NcML that introduces conventions for specifying information in netCDF files that is characteristic of GIS data (i.e., georeferencing and coverage information); NcML-G_{ML} implements such conventions using the OpenGIS Geography Markup Language (GML) grammar.

2 NetCDF Background

This article is a general description of the netCDF Markup Language (NcML) and a special extension of NcML leveraging Geography Markup Language (GML) grammar (NcML-G_{ML}). To understand NcML, it is important to understand the basics of netCDF which are described in the next section which borrows liberally from the frequently asked questions about netCDF web page [2].

NetCDF is an Application Programming Interface (API) that provides methods for accessing array-oriented data and a freely-distributed collection of software libraries for C, FORTRAN, C++, Java, and perl that implement this interface. The software was developed by Glenn Davis, Russ Rew, Steve Emmerson, John Caron, and Harvey Davies at the Unidata Program Center in Boulder, Colorado and augmented by contributions from other netCDF users [1]. The netCDF libraries define a portable format for representing scientific data. The interface, libraries, and format support the creation, access, and sharing of scientific data.

NetCDF data is:

- *Self-Describing*. A netCDF file includes information about the data it contains.
- *Architecture-independent*. A netCDF file is represented in a form that can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.
- *Directly-accessible*. A small subset of a large dataset may be accessed efficiently, without first reading through all the preceding data.
- *Appendable*. Data can be appended to a netCDF dataset along one dimension without copying the dataset or redefining its structure. The structure of a netCDF dataset can be changed, though this may require copying the dataset.

- *Sharable*. One writer and multiple readers may simultaneously access the same netCDF file.

The XML language and related technologies have similar characteristics : XML is self-describing and architecture-independent; whereas XPath, XQuery, XUpdate, and XPointer, among others support direct access and modifications. Therefore, XML technologies proved to be ideal candidates for netCDF encoding, and it was easy to preserve netCDF capabilities in the NcML specification.

3 NcML Overview

NcML is an XML description of the content and structure of the data stored in a netCDF file [3]. NcML has many of the same characteristics of netCDF. As a text-based language, NcML does not provide direct access to its contents. However, since NcML does not contain the data from a netCDF file, this characteristic is not as important for NcML. Since netCDF is a self-documenting binary form, NcML can consist of the metadata contained in the "header" of the binary netCDF file itself. For those familiar with the toolset associated with netCDF, this most basic form of NcML contains roughly the same information that results from the "ncdump -h" command. In that sense, the content of NcML can be thought of as an XML representation of the netCDF CDL (network Common data form Description Language) [4]. However, later sections of this article will describe NcML features and uses that go beyond XML representations of the metadata within an existing netCDF file. These extensions provide important benefits to users: crucial metadata is often missing from a netCDF file, which can prevent general-purpose visualization and analysis packages from being able to properly use the data in the file. An important example of this is when the geo-referencing coordinate information is missing or in a form that is not understood by a package, a very common case. By using NcML to add the information in a way that the program understands, the file does not need to be rewritten or the program modified, both often impossible in any case. NcML can aggregate different netCDF files together, presenting a single dataset to the user. This feature is often used to stitch time series data together. This makes NcML a declarative language for writing and rewriting netCDF files.

The THREDDS project [Domenico, 2002] uses all these features to make datasets available to the education and research communities that are otherwise not easily accessible.

Within NcML itself, there are currently four parts: a core specification and three extensions, each with its own schema document as noted:

- ***NcML Core Schema*** represents the existing netCDF data model.
- ***NcML Coordinate System Schema*** extends NcML Core Schema to add explicit support for general and geographic coordinate systems.
- ***NcML Dataset Schema*** extends NcML Core Schema to use NcML to define a netCDF file, similar to the **ncgen** command line tool, as well as to redefine, aggregate, and subset existing netCDF files.
- ***NcML Geography (NcML-G) Schemas*** extend the NcML Coordinate System schema to facilitate the use of netCDF datasets by GIS systems; a valuable example is the NcML-G_{ML} schema which uses the grammar introduced by the Open Geospatial Consortium Geography Markup Language (GML).

Figure 1's schema depicts the architecture layers of NcML specifications and its main relationship with netCDF data model. The following sections describe NcML's components, discussing their .model, schema and software implementation aspects.

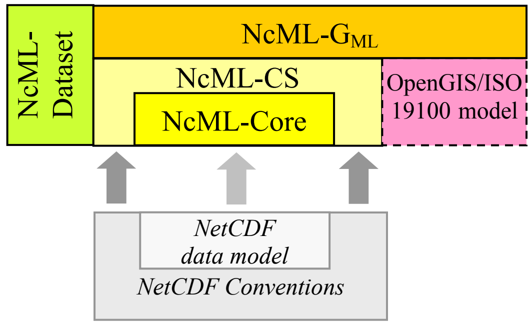

Fig. 1 NcML architecture layers

4 NcML Core

The NcML schema is based on an abstract object model for expressing metadata associated with generic netCDF data and is very closely related to the netCDF data model. As noted above, NcML's goal is to describe netCDF files in a standard, extensible, and sharable manner.

4.1 Object Model

The NcML Core object model includes the following types of objects:

- **NetCDF**, a netCDF dataset (for example, a netCDF file, an aggregate of netCDF files, or a subset of a netCDF file)
- **Dimension**, a named index of specified length
- **Variable**, a multidimensional array of specified type indexed by 0 or more dimensions
- **Attribute**, a name-value pair of specified type

Figure 2 shows the relationships among the objects.

4.2 XML Schema

The object model was encoded in XML, introducing an XML schema. The diagram depicted in Figure 3 represents the NcML-Core schema in XSD format following the netCDF Metadata Object Model (release 1.0). The *netcdf* element is a container of variable content (i.e. *variables*, *dimensions* and *attributes*), the <choice> group element was chosen to implement such aggregation relationship. That is appropriate because the NcML Specification Group controls the specification evolution.

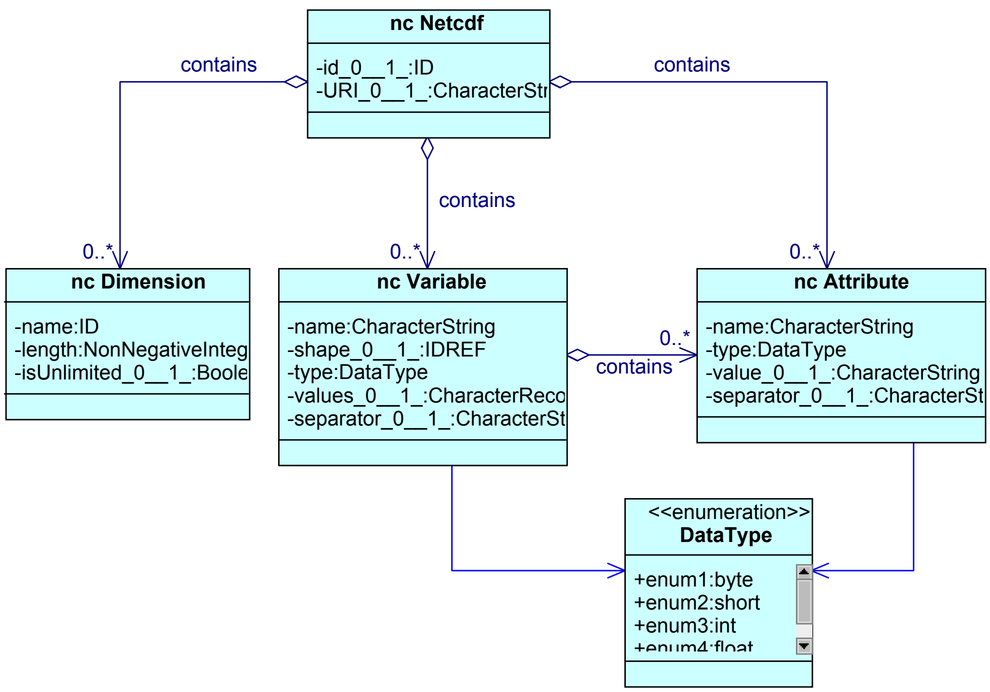

Fig. 2 NcML Core conceptual model

A simple example of an NcML document is shown below. Those familiar with netCDF know that it easily recognises the encoding of the netCDF data model concepts: the document's main sections are separated for clarity:

```
<?xml version="1.0" encoding="UTF-8"?>
<netcdf xmlns="http://www.ucar.edu/schemas/netcdf"
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```

```
          xsi:schemaLocation="http://www.ucar.edu/schemas/netcdf
http://www.unidata.ucar.edu/schemas/netcdf.xsd"
          uri="P:\packages\netcdf\ncml\examples\example.nc">

          <dimension name="time" length="2" isUnlimited="true"/>
          <dimension name="lat" length="3"/>
          <dimension name="lon" length="4"/>

          <attribute name="title" type="string" value="Example Data"/>

          <variable name="rh" type="int" shape="time lat lon">
                    <attribute name="long_name" type="string" value="relative humidity"/>
                    <attribute name="units" type="string" value="percent"/>
          </variable>

          <variable name="T" type="double" shape="time lat lon">
                    <attribute name="long_name" type="string" value="surface temperature"/>
                    <attribute name="units" type="string" value="degC"/>
          </variable>

          <variable name="lat" type="float" shape="lat">
                    <attribute name="units" type="string" value="degrees_north"/>
          </variable>

          <variable name="lon" type="float" shape="lon">
                    <attribute name="units" type="string" value="degrees_east"/>
          </variable>

          <variable name="time" type="int" shape="time">
                    <attribute name="units" type="string" value="hours"/>
          </variable>

</netcdf>
```

More detailed information on the core model and schema, as well as annotated examples, are available at the NetCDF/NcML site [3].

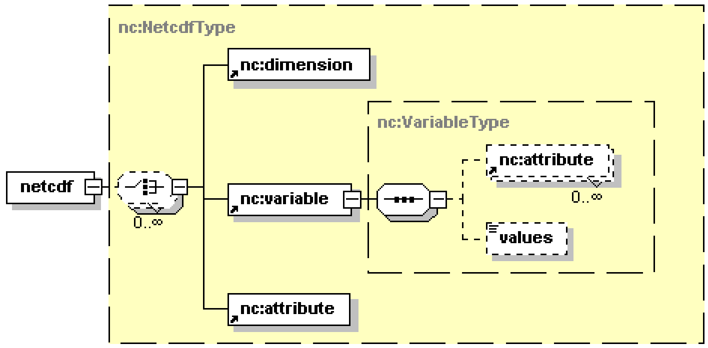

Fig. 3 NcML-Core schema in xmlspy® notation

4.3 Software implementation

The latest NcML Core schema (release 1.0) contains a reference implementation in the netCDF Java library, version 2.1 [5]; it is a Java interface to netCDF files. The library consists of a core (i.e. ucar.nc2 and ucar.ma2) which supports writing netCDF metadata into NcML. The library includes a netCDF interface to OPeNDAP (DODS) datasets as an option. Another optional part uses the NetCDF Markup Language (NcML) to allow the definition of virtual netCDF datasets, and to extend the netCDF data model to include general coordinate systems. The library is freely available and the source code is released under the GNU Lesser General Public License (LGPL). The NCAR community portal contains a prototype web application that extracts NcML metadata from any network-retrievable netCDF file and displays it as an HTML web page [6].

5 NcML Dataset extension

The NcML Dataset extends core NcML to define a new netCDF file. The purpose of the NcML Dataset is to allow:

- Metadata to be added, deleted, changed and in general redefined.
- Variables to be renamed, added, deleted, and restructured.

- Information from multiple netCDF files to be combined.

The netCDF file defined by an NcML Dataset document is called a *netCDF Dataset* and can be a virtual dataset: it does not have to exist as a file on a disk. A *netCDF dataset* is a generalization of a netCDF file. A netCDF Dataset Definition document uses the NcML Dataset XML schema which is an extension of the NcML Core schema.

An NcML Dataset document makes it possible to define a netCDF form that may or may not exist as a file. For those familiar with the netCDF toolset, the NcML Dataset can be thought of as an XML representation similar to netCDF CDL; it is also possible to use an NcML Dataset to redefine, aggregate, and subset existing netCDF files. An NcML Dataset document can also be used to specify a new form into which an existing netCDF file is to be transformed; thus it can be used by various tools in the conversion from one set of netCDF conventions to another. In essence, the NcML Dataset specification provides the ability to define a virtual netCDF object.

5.1 Object Model

The following extensions are made to the core NcML to create an *NCML Dataset:*

1. ***netcdf*** is extended to allow *readMettadata, aggregation* and nested *netcdf* elements where:
    - ***readMetada*** enables reading all or some of the variables, attributes, and dimensions from another netCDF dataset, as well as rename them.
    - ***aggregation*** enables aggregating variables from nested netcdf elements into a single logical variable.
2. ***values*** is extended to allow the definition of regularly spaced values using a start and increment.

Figure 4 shows the NcML Dataset object model. The pale blue objects are the entities defined in the NcML Core model.

5.2 XML Schema

This model was encoded in XML, introducing an XML schema; a representation of it is shown in Figure 5. Several types are redefined and derived by extension from the Core schema. Although redefinition may lead to conflict and hinder extensibility, in this case, we can assume it is safe because the schema evolution is controlled. Details on the models and schema, tutorials and examples are available at the NetCDF/NcML site [3].

5.3 Software implementation

The latest NcML Dataset schema is in release 0.6 alpha; it has a reference implementation in the netCDF Java library, version 2.1 [5]. The optional dataset package of the library (i.e. *ucar.nc2.dataset* package) allows the creation of virtual netCDF datasets, including aggregating individual netCDF files using the extension rules specified in the NcML-Dataset extension model and schema. The same package implements access to netCDF Datasets through the standard Java netCDF API.

The abstractions and XML syntax defined by the NcML Dataset specification are language independent, and in the future a C library implementation could be created to provide equivalent functionality.

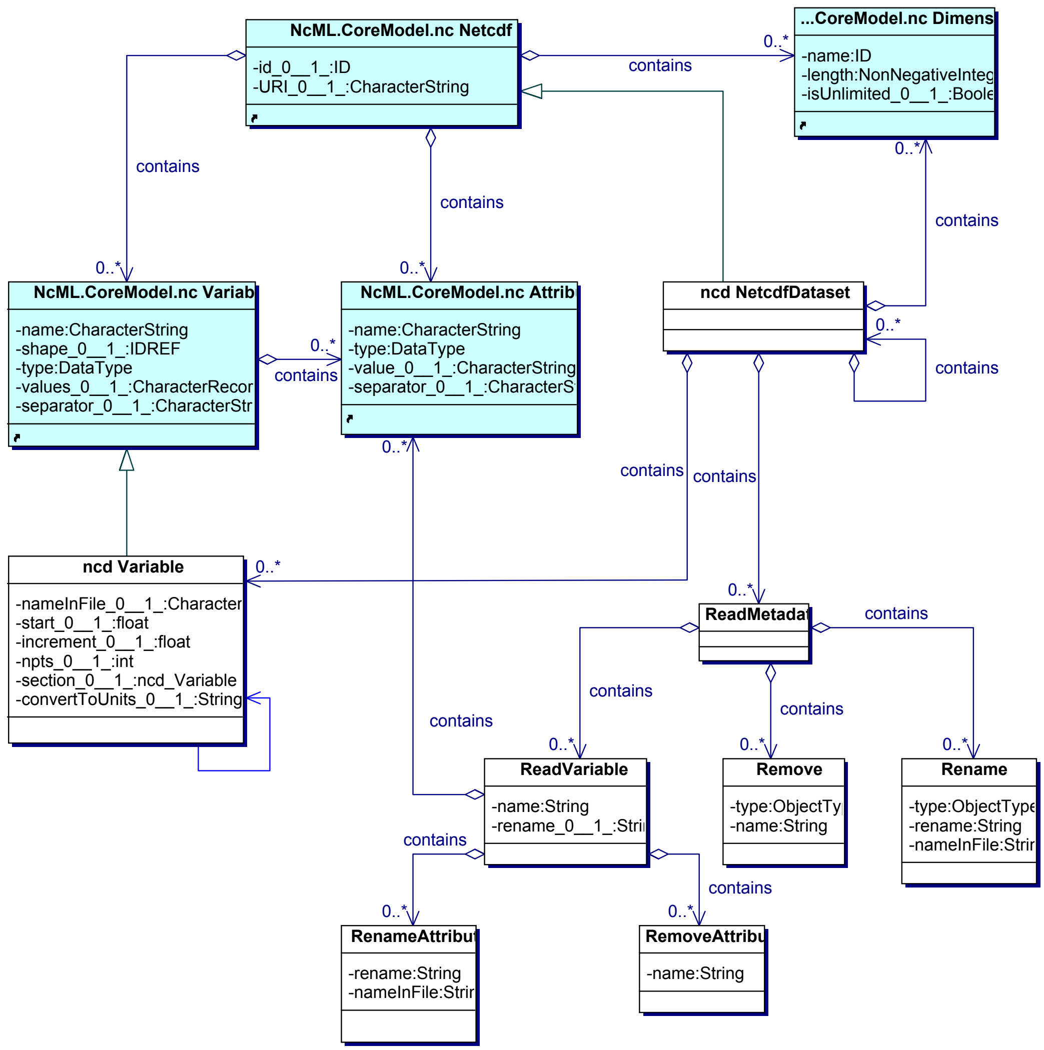

Fig. 4 NcML Dataset conceptual model

6 NcML Coordinate System extension

Simply using netCDF is not sufficient to make data "self-describing" and represent its semantics. The names of variables and dimensions should be meaningful and conform to relevant conventions; dimensions should have corresponding coordinate variables where sensible; global attributes should provide dataset ancillary information following standard specifications. For that purpose, several groups have defined additional conventions and styles for netCDF data.

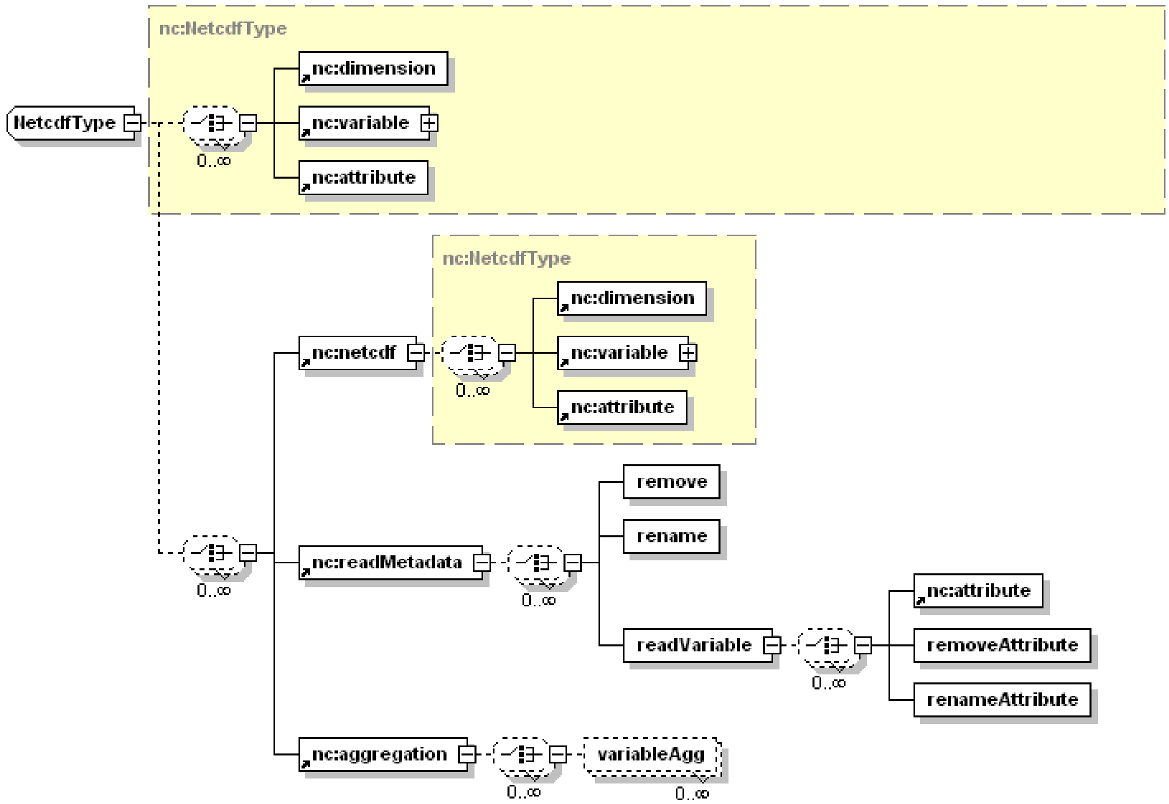

Fig. 5 *NcML-Dataset schema in xmlspy® notation*

Descriptions of these conventions, as well as examples incorporating them can be accessed from the netCDF Conventions site [7]. The NcML Coordinate System (NcML-CS) extends the NcML Core model to add explicit support for general and geographic coordinate systems. It can be used to describe existing netCDF conventions for coordinate variables, and is also able to capture more complex coordinate systems and transformations. NcML-CS provides a standard encoding which eases the sharing of coordinate system information in the netCDF community. The NcML-CS model was tested for several popular netCDF conventions, such as: AWIPS, COARDS, CF-1, NUWG and WRF model.

6.1 Object Model

The NcML-CS object model extends the NcML core model to add explicit support for general and geographic coordinate systems.

Summary of extensions to Core:

1. **netcdf** is redefined to allow coordinateSystem and coordinateTransform elements.
2. **variable** is redefined to allow the coordinateSystem attribute, which lists the coordinate systems for this variable.
3. **coordinateAxis** extends variable, adding the units attribute, as well as optional attributes axisType, positive and boundaryRef
   - **axisType** attribute identifies geographic coordinate axes. It has the possible values of GeoX, GeoY, GeoZ, Height, Pressure, Lat, Lon, or Time.
   - **positive** attribute is only used for vertical coordinate axes, and gives the direction of increasing values. It has the value up or down.
   - **boundaryRef** attribute is the name of a variable that is the boundary of this coordinate axis.
4. **coordinateSystem** is a list of coordinateAxis elements, and optional coordinateTransform elements.
5. **coordinateTransform** represents a mathematical transformation of one coordinate system to another. It has a list of name/value pairs, called parameters, and an optional transformType that has the value Projection or Sigma.

The schema of Figure 6 depicts the NcML-CS object model; pale blue objects are part of the NcML core package.

6.2 XML Schema

The NcML-CS model was encoded in XML, introducing an XML schema; it is essentially the external XML representation of the netCDF data model extended by coordinate system conventions. An annotated representation of the schema is shown in Figure 7. Netcdf Type and VariableType types are redefined and derived by extension from the Core schema. Although, redefinition could lead to conflicts and hinder extensibility, here again we can assume it is safe because the schema evolution is controlled.

Types related to geographic coordinate system information, may be revised in future specifications by resorting to GML constructs. Details on the models and schema, and NcML document examples for several conventions are available at the netCDF/NcML site [3]

6.3 Software implementation

The latest NcML Coordinate System schema in XML is in release 1.0 beta. It has a reference implementation in the netCDF Java library, version 2.1 [5]. As previously stated, the optional dataset package of the library (i.e. *ucar.nc2.dataset* package) allows the creation of virtual netCDF datasets; a dataset sub-package (i.e. *ucar.nc2.dataset.conv* package) extracts coordinate systems semantics from netCDF files using netCDF conventions and can write the retrieved information to XML using the NcML-CS extension.

7 NcML Geography

There is an increasing emphasis on Earth systems approaches to interdisciplinary research among the subdisciplines in the geosciences. Much of that research involves integrated studies of scientific datasets from each subdiscipline. As just one example, atmospheric scientists and hydrologists are joining forces to develop a better understanding of water flow in drainage basins. Such studies can benefit from coherent access to radar data from the atmospheric science domain and streamflow information from hydrology. One natural application of this sort of

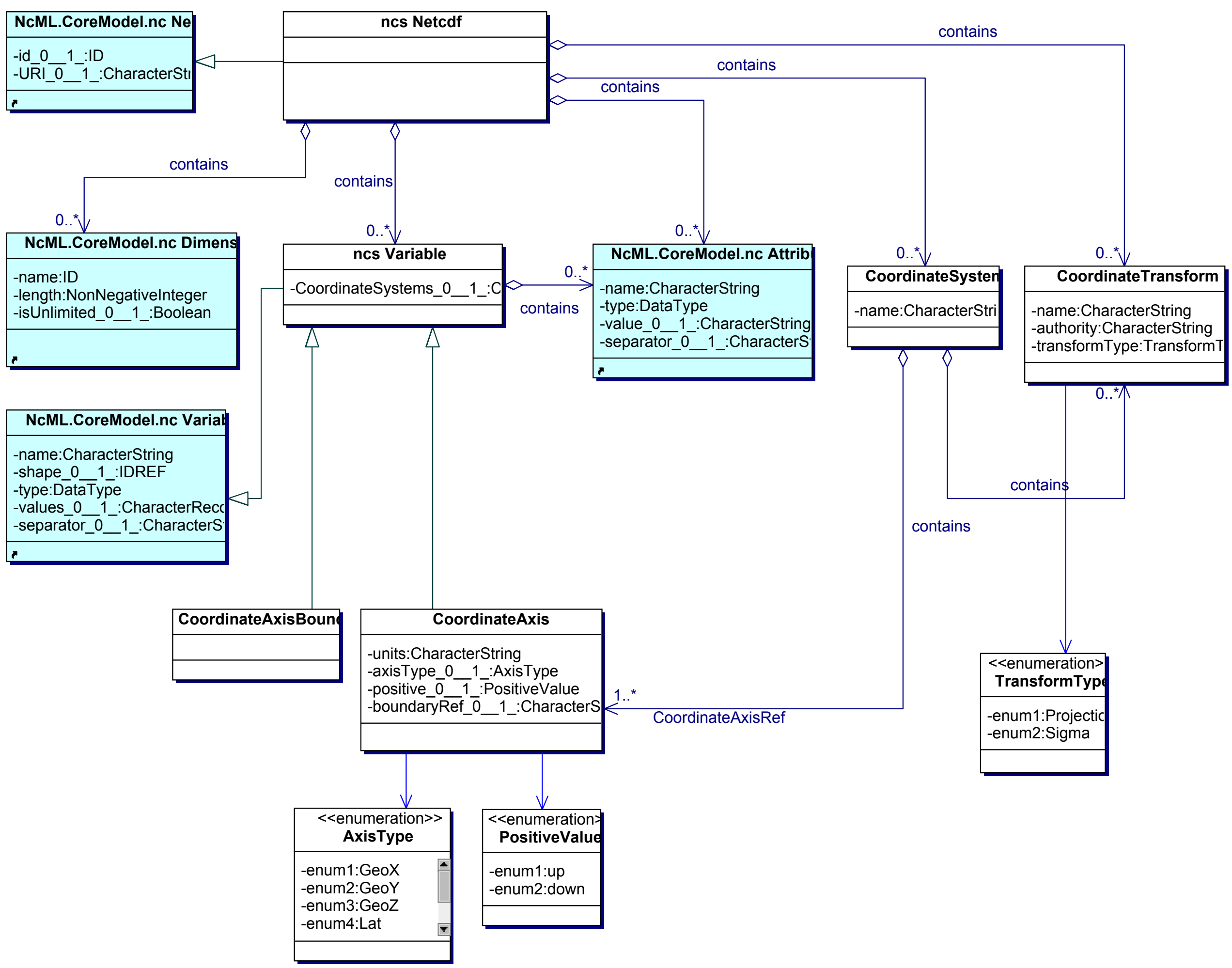

Fig. 6 NcML-CS conceptual model

research is in the analysis and prediction of flash floods. Infrastructure and demographic data from the societal impacts community is also important. . Traditionally these communities have used completely different data storage and access systems. As noted earlier, netCDF is in widespread use in the atmospheric sciences. However, the hydrology and societal impacts communities traditionally have used Geographic Information Systems (GIS) for storing, accessing, and analyzing their data.

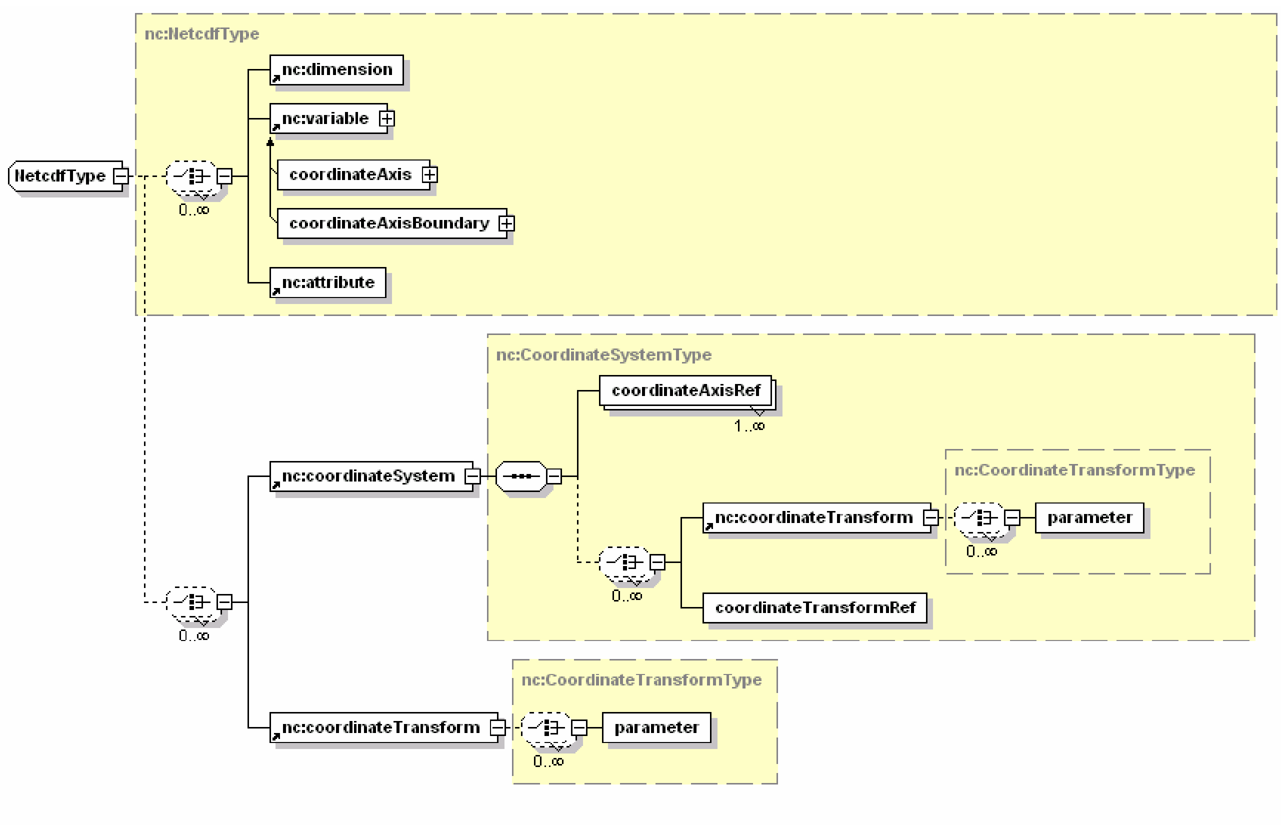

Fig. 7 *NcML-CS schema in xmlspy® notation*

The main objective of the NcML Geography (NcML-G) extension is to address issues such as these and facilitate the interoperability between the atmospheric science and the GIS communities. This will simplify the data access difficulties that currently hinder integrated interdisciplinary studies between the two communities. NcML-G can be considered a piece of middleware to provide the research and education communities with atmospheric sciences legacy datasets, through GIS community tools.

Moreover, NcML-G will enable the atmospheric sciences community to more readily ascertain the societal impacts of extreme weather events and provide a better means of communicating those impacts to the people affected.

7.1 Object model

The NcML-G model addresses the need to interconnect the atmospheric science data model with the GIS data model to achieve interoperability between them. NcML-G extends the NcML-CS data model to add explicit support for information characteristic of GIS data. It contains:

1. dataset general extension information;
2. referencing information for spatial and temporal coordinate systems;
3. information describing the geometry of coverage derived from the netCDF dataset.

NcML-G introduces a set of supplementary concepts which explicitly formalizes these essential GIS facets of many netCDF datasets

7.1.1 Atmospheric Science Datasets as GIS Coverages

The difficulty of achieving this interoperability in the real world of data stems from the fact that the atmospheric science dataset model is fundamentally different from the GIS data model [Nativi et al., 2004]. In particular, most atmospheric science data is based on a composite approach, whereas most GIS data is organized according to a geo-relational approach [Molenaar, 1991]. The composite approach is in essence a "bottom up" means of organizing data, proceeding from individual measurement values to aggregated entities made up of those measurement values. On the other hand, the geo-relational approach to data organization is "top down" proceeding from meaningful aggregation entities to their actual measurements content. GIS systems work with two fundamental geographic data types: features and coverages. Features generally represent geometric entities on the surface of the earth (e.g., rivers, streets, lakes, and tracts of land). The characteristics of those entities are usually stored in a formal DBMS. Coverages, however, can be used to map composite data of the sort found in satellite images, radar observations, or the output of forecast models. In fact, GIS considers a coverage a special case of feature. For the purposes of NcML-G model, we consider the netCDF model as the reference composite model for the atmospheric science community (see Figure 8), and the standard ISO TC211 Coverage models [ISO, 2001] as the reference geo-relational model for the GIS community (see Figure 9). Interoperability was achieved mediating and integrating models concepts to achieve the NcML-G model depicted in Figure 10.

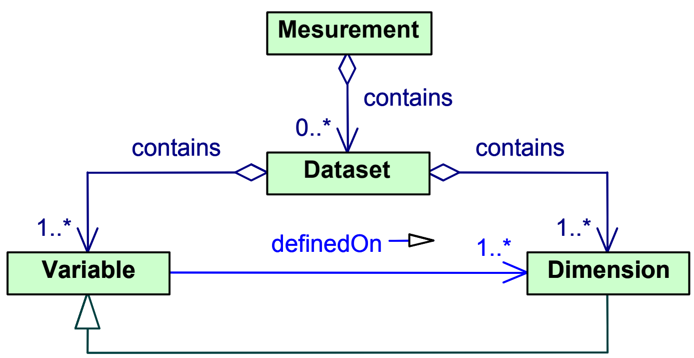

Figure 8 Atmospheric Science Composite model

As shown in Figure 10, there are four main extension steps to prepare netCDF data for GIS-based applications :

1. generate **geo-extent information** for a netCDF dataset. It captures the dataset's general extension information. This is useful for online catalogue applications, for incorporating

georeferenced datasets into digital libraries and other discovery systems, and for classifying and filtering datasets;

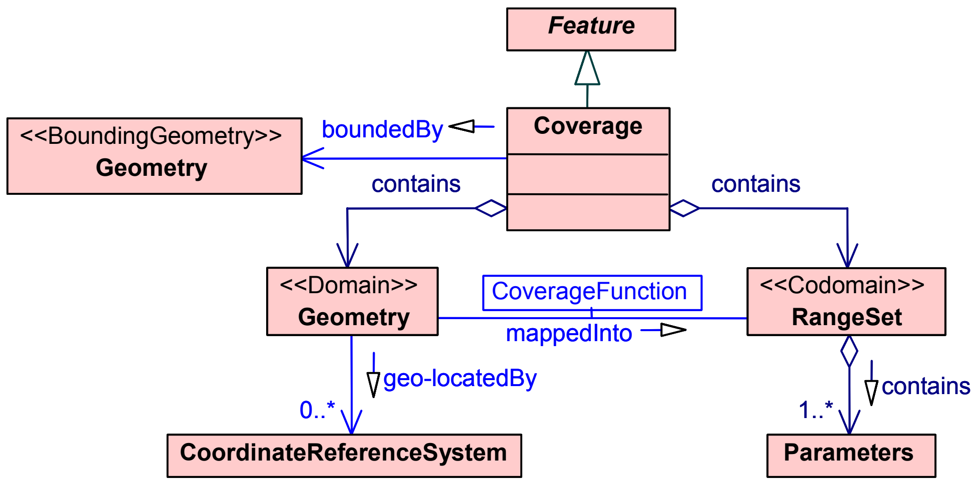

Figure 9 Coverage geo-relational simplified model

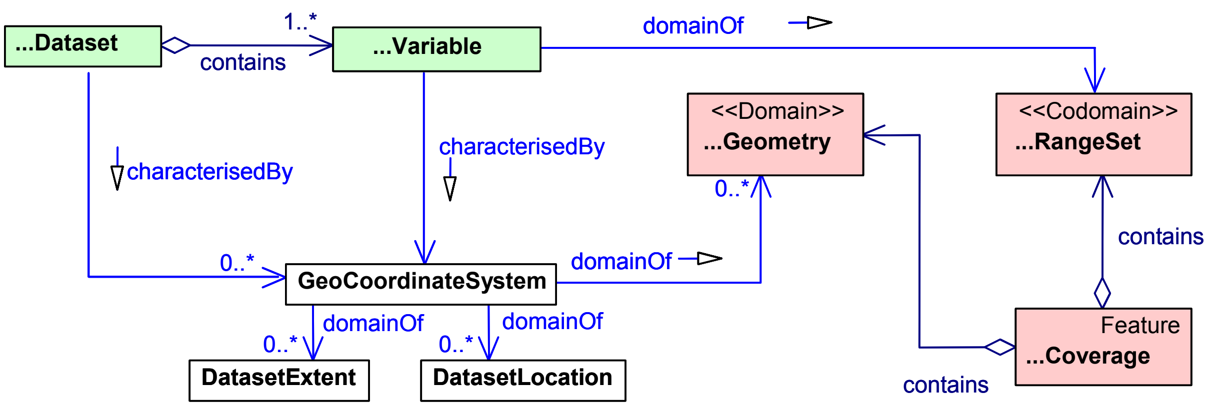

Figure 10 The NcML-G model

2. generate **geo-location information** for a netCDF dataset. It captures the referencing information for spatial and temporal coordinate systems. This information is useful for representing dataset values on referenced maps and for overlaying them on geo-referenced layers (e.g. themes).

3. generate one or more **coverages** from a netCDF dataset along with information describing them. Specifically: coverage RangeSet elements are generated from netCDF variable elements and coverage Geometry is generated from the Geo-Coordinate System elements that characterise netCDF variables. These RangeSet elements are useful for expressing dataset components as logical geo-relational features which allows integration of netCDF variable elements with other GIS features.
4. link the NcML variables elements with the corresponding coverage range sets.

7.1.2 The achieved extension model

.Defining a geo-coordinate system for netCDF data is of great importance for achieving Interoperaperability with the GIS data model. Hence, the NcML-CS model is the natural starting specification for introducing the required elements, and the Coverage concept must be introduced; the ISO TC 211 [ISO, 2001] coverage model was used for this purpose.
Figure 11 is a schematic representation of the NcML-G object model which extends the NcML-CS one: yellow objects extend the NcML coordinate system type introducing general extension and referencing properties; orange objects describe referenced and unreferenced coverage.
Five object packages constitute the achieved model for the NcML-G extension (depicted in different colours):

1. **GISExtension package**: contains objects describing referenced and unreferenced geocoordinate system.
2. **GeoDomainExtent package**: contains objects describing the spatio-temporal extent of the dataset.

3. **GeoDomainLocation package**: contains objects describing the spatio-temporal coordinate reference system of the dataset.
4. **UnreferencedCoverage package**: contains objects describing unreferenced coverages.
5. **ReferencedCoverage package**: contains objects describing referenced coverages.

According to GIS semantics, a coverage associates a position within a spatial/temporal domain to a value of a defined data type. A coverage is a function from a spatial/temporal domain to an attribute domain. The domain can be referenced or unreferenced. Figure 12 depicts the object model for both NcML-G unreferenced coverages (the CoverageGeoCoordSystem object and the three yellow objects) and referenced coverages (all objects in the figure).

7.2 XML schema

7.2.1 NcML-G_{ML}

The NcML-G model was encoded in XML using the grammar introduced by the Open Geospatial Consortium GML version 3.0 [OGC, 2003a]; such implementation is called NcML-G_{ML}. This is not the only encoding. The authors developed another one which is loosely based on the XML grammar introduced by ISO TC211 to implement their model specifications.

The choice of using the GML grammar makes it possible, for the GIS Community, to seamlessly use netCDF datasets represented consistent with the NcML-G model. Indeed, GML conforms to the GIS Community semantics and is a well accepted interoperability language for GIS. OGC GML is a XML grammar written in XML Schema for the modelling, transport, and storage of geographic information.

GML version 3.0 is in the ISO standardization process as ISO/TC211/PT 19136 [ISO, 2003]. GML 3.0 provides a rich variety of objects for describing geographical information -including geographic features, coordinate reference systems, geometry, topology, time, units of measure and generalized values. The GML encodes the geomatics standard models introduced by the OGC and ISO TC211 initiatives. NcML-G$_{ML}$ is based on the following GML 3.0 specifications: feature, temporal, coordinateReferenceSystems, grids and coverages.

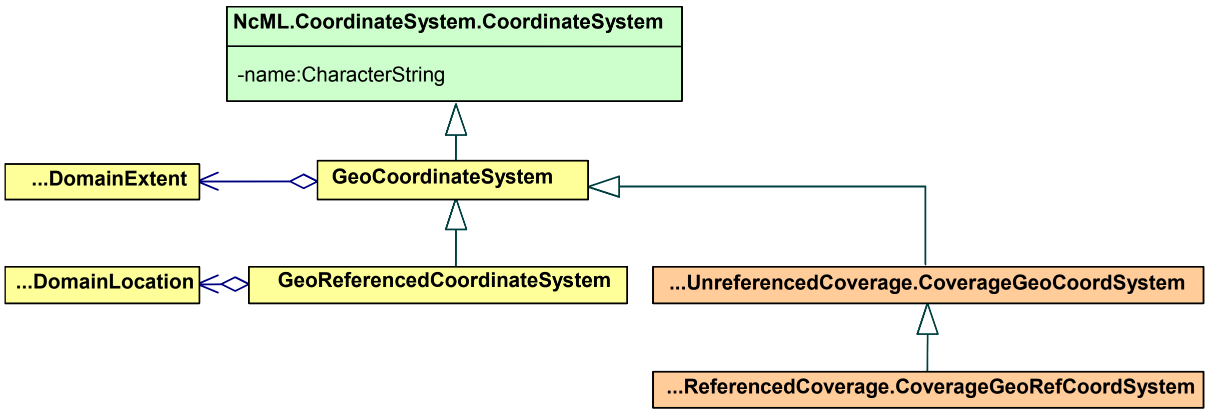

Figure 11 The NcML-G object model

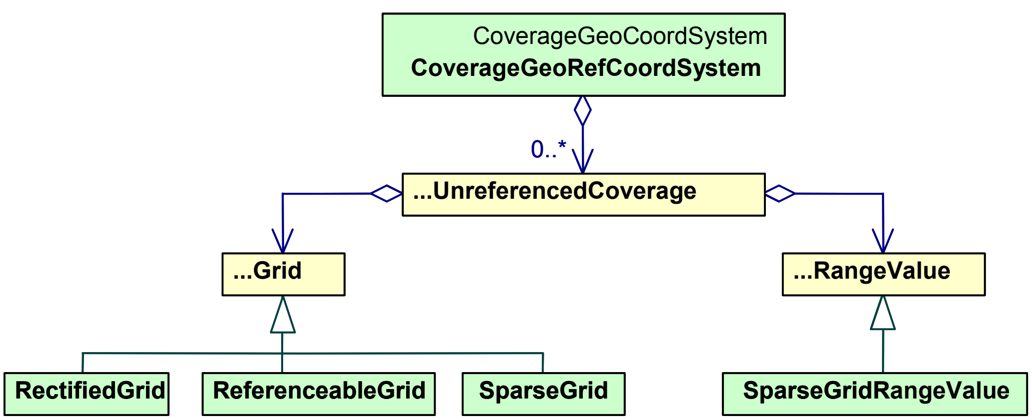

Figure 12 The NcML-G model for referenced coverage

An annotated representation of the schema of the NcML-G$_{ML}$ Geo-Coordinate System and Grid Coverage elements is shown in Figure 13 and 14, respectively. XML Name Spaces reports the origin of the sub-elements depicted in the Figures. The NcML-G$_{ML}$ geoCoordinateSystemType was derived by extension from the CoordinateSystemType defined in the NcML-CS schema. An instance of geoCoordinateSystemType was derived by element substitution from an instance of CoordinateSystemType. The same approach was followed for defining the geoRefCoordinateSystemType and its instance element.

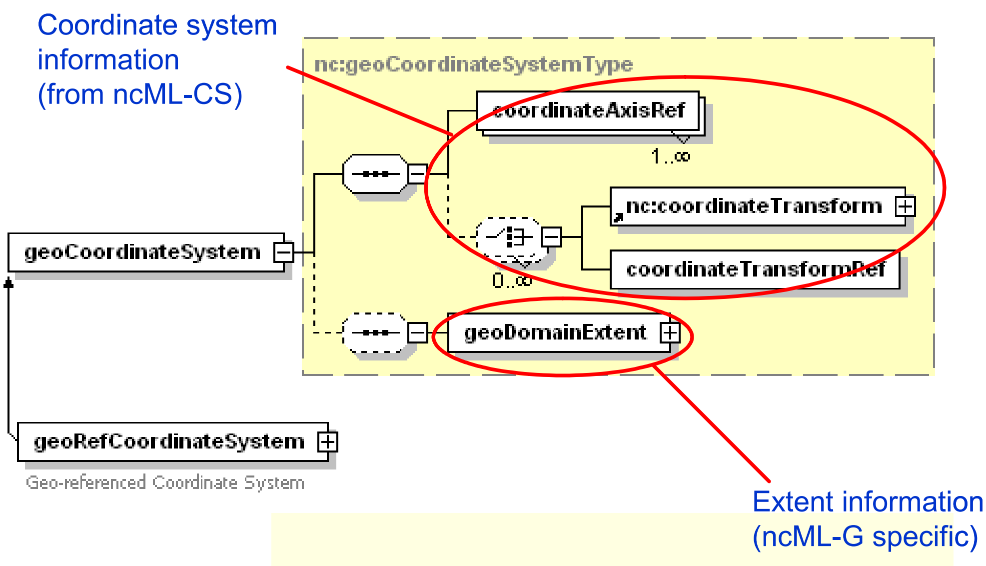

Figure 13 The NcML-GML Geo-Coordinate System schema in xmlspy® notation

The following example demonstrates the generation of a GIS coverage from a netCDF dataset. The coverage has a geo-referenced grid geometry:

```
<netcdf xmlns="http://www.ucar.edu/schemas/netcdf" xmlns:gml="http://www.opengis.net/gml"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.ucar.edu/schemas/netcdf ……\…\ReferencedCoverage.xsd" id="filename"
uri="http://www.ucar.edu.filename">
<!--   NcML Core elements   -->
       <dimension name="x" length="4"/>
       <dimension name="y" length="3"/>
       <variable name="x" type="int">
             <values separator=";">12;22;22;20</values>
       </variable>
       <variable name="y" type="int">
             <values separator=";">44;44;45</values>
       </variable>
```

```
<variable name="temperature" type="double" shape="x,y">
      <values
separator=",">237.6,258.7,260.2,277.5,270.4,269.5,266.5,267.4,269.6,268.0,268.4,268.1 </values>
</variable>
<variable name="WV" type="double" shape="x,y">
      <values separator=",">4.6,4.7,5.2,5.3,5.4,6.8,6.1,6.4,6.6,6.5,6.0,5.8</values>
</variable>
```

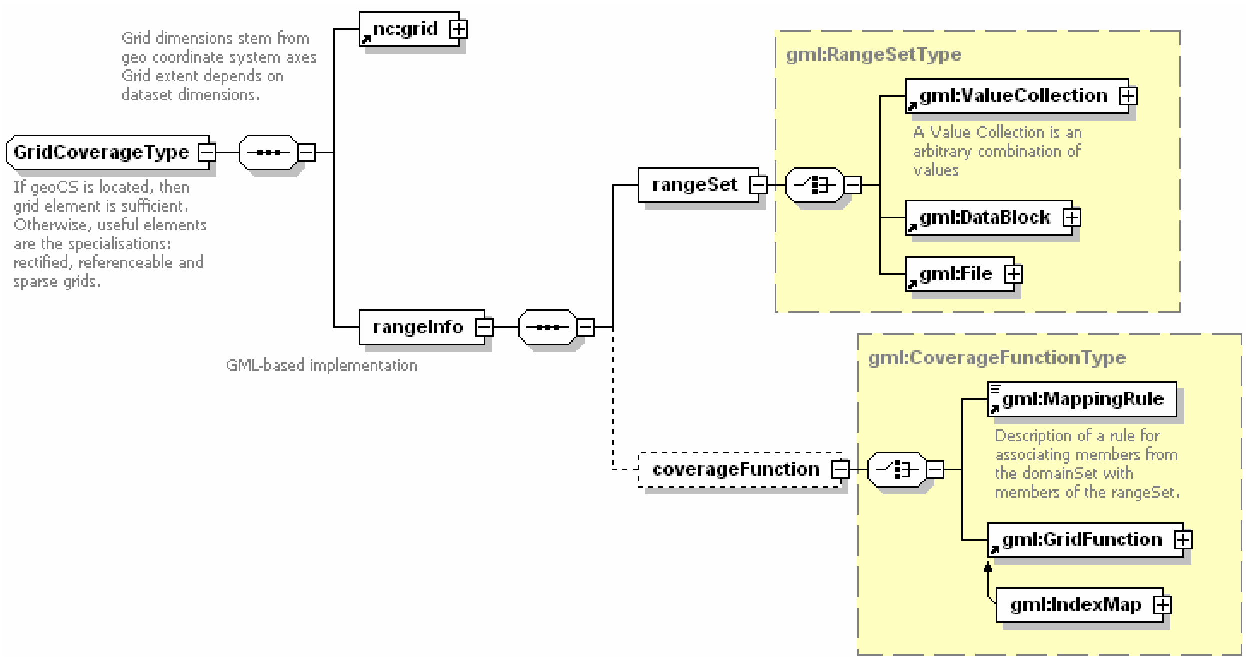

Figure 14 The NcML-GML Grid Coverage schema in xmlspy® notation

```
<!--   NcML-GML elements based on NcML-CS elemnts  -->
<coverageGeoRefCoordSystem name="GeographicSystem">
      <coordinateAxisRef ref="x"/>
      <coordinateAxisRef ref="y"/>
      <geoDomainExtent description="EPSG:4326">
            <spatialExtent>
                  <gml:Envelope>
                        .........
                  </gml:Envelope>
            </spatialExtent>
```

```
            </geoDomainExtent>
            <geoDomainLocation>
                  <gml:GeographicCRS>
                        <crsID>
                              <code>4326</code>
                              <codeSpace>EPSG</code>
                              …….
                        </crsID>
                        .........
                  </gml:GeographicCRS>
            </geoDomainLocation>
            <gridCoverage>
                  <grid dimension="2">
      <limits>
            <gridEnvelope>
                  ....
            </gridEnvelope>
      </limits>
      <axisName>Longitude</axisName>
      <axisName>Latitude</axisName>
                  </grid>
                  <rangeValues>
                        <rangeInfo>
                              <values>
                                    <variable>temperature</variable>
                                    <variable>WV</variable>
                              </values>
                        </rangeInfo>
                  </rangeValues>
            </gridCoverage>
      </coverageGeoRefCoordSystem>
</netcdf>
```

A couple of XML document sections which conform to different NcML part schemas are highlighted.

7.3 Software implementation

The latest NcML-G_{ML} schema in XML is in release 0.2 alpha. Three separate XML schemas have been defined to specify:

1. Geo-Coordinate Systems elements defined from netCDF dataset;
2. Un-referenced Coverage elements;
3. Referenced Coverage elements.

A similar schema specification has been used for testing interoperability between the THREDDS [Domenico, 2002] services and the OGC Web Coverage Service specification [OGC, 2003b].
A Java reference implementation will be released when the schema specification will be more stable: the OGC adopted the new GML 3.1 specification, which is being considered for the new version of the schema. In addition, the level of integration of GML coordinate system grammar into NcML-G_{ML} specification is being revised.

8 Conclusions

The paper has presented NcML and the extensions used to introduce coordinate system and dataset information. In addition, the NcML geography extension, encoded on the top of GML grammar: NcMLG_{ML} was presented. NcML and its extension models were analyzed, models implementations were discussed and several encoding examples were presented.
NcML was conceived as the enabling tool to bring netCDF into the interoperable framework of Web Services and through it to Grid Services. Similarly, NcML-G_{ML} was conceived as the means of bringing netCDF capabilities to the GIS Community, and through it to benefit society at large.

Given a netCDF file, NcML-Core specification provides conventions to express it in XML and therefore leverage the interoperability benefits offered by the Web Services framework. If the netCDF data follows given conventions, it is possible to extend its interoperability using NcML-CS conventions which express general and geographic coordinate system semantics and content. Another interoperability improvement concerns the possibility of generating a "virtual" netCDF dataset which can be defined starting from a given file, or combining it with others. The NcML-Dataset specifications would provide useful conventions to achieve the goal. Eventually, the NcML-G specifications would provide the conventions to "present" the given netCDF data as a set of GIS coverages, enriching it with the appropriate metadata to implement full interoperability with GIS applications. NcML-G is very young, but it has been used for testing interoperability between the THREDDS services and the OGC Web Coverae Service specification which is of great interest for the GIS community.

Acknowledgements

This work has been partially funded by the U. S. National Science Foundation project called THematic Real-time Environmental Distributed Data Services (THREDDS) at the Unidata Program Center of the University Corporation for Atmospheric Research.

References

Domenico, B., Caron, J., Davis, E., Kambic, R., and Nativi, S., 2002. Thematic Real-time Environmental Distributed Data Services (THREDDS): Incorporating Interactive Analysis Tools into NSDL. Journal of Digital Information, Vol. 2 Issue 4 Art. No. 114, 29 May 2002.

ISO 19123, 2001. Schema for coverage geometry and functions, ISO TC211 document N1038, 18 Jan 2001.

ISO 19136, 2003. Geography Markup Language, ISO TC211 document N1432, 25 Apr 2003.

Molenaar, M., 1991. Status and problems of geographical information systems. The necessity of a geo-information theory. ISPRS Journal Programme. Remote Sensing, Elsevier Science Publisher, B.V. pages 85-103, 1991.

Nativi, S., Blumenthal, B., Habermann, T., Hertzmann, D., Raskin, R., Caron, J., Domenico, B., Ho, Y., Weber, J., 2004. Differences among the data models used by the Geographic Information Systems and Atmospheric Science communities, Proceedings of American Meteorological Society – 20th Interactive Image Processing Systems Conference, Seattle (WA), Jan 2004.

OGC (Open Geospatial Consortium, Inc.), 2003a. OpenGIS® Geography Markup Language (GML) Implementation Specification – version 3.00. OpenGIS® project document OGC 02 023r4. 29 January 2003.

OGC (Open Geospatial Consortium, Inc.), 2003b. Web Coverage Service (WCS) - version 1.0.0. OpenGIS® project document OGC 03-065r6. 27 August 2003.

Internet References

[1] Unidata netCDF Group, “netCDF Usage”, *Unidata software: netCDF*, available at http://my.unidata.ucar.edu/content/software/netcdf/usage.html

[2] Unidata netCDF Group, “NetCDF Frequently asked questions”, *Unidata software: netCDF,* http://available at http://my.unidata.ucar.edu/content/software/netcdf/faq.html#whatisit

[3] NcML Specification Group, “NcML Overview”, *Unidata NcML tutorial*, available at www.unidata.ucar.edu/packages/netcdf/ncml/

[4] Unidata netCDF Group, “NetCDF Description Language”, *Unidata software: netCDF*, available at http://www.unidata.ucar.edu/packages/netcdf/guidec/guidec-15.html#HEADING15-0

[5] J. Caron, “NetCDF Java Library”, *Unidata netCDF packages*, available at http://www.unidata.ucar.edu/packages/netcdf-java/

[6]https://dataportal.ucar.edu:8443/xslt.do?xml=http://dataportal.ucar.edu/metadata/acd/mozart/ncml/h0040.ncml&schema=ncml

[7] Unidata netCDF Group, “netCDF Conventions”, *netCDF User's Guide*, available at http://www.unidata.ucar.edu/packages/netcdf/conventions.html