16.2

A comparison of common geospatial data model between AS and GIS community

Yuan Ho[1], Jeff Weber and John Caron
Unidata, UCAR

1 Introduction

The complexity and variability of geospatial data formats pose great challenges for both the Atmospheric Science (AS) and the Geographic Information Systems (GIS) communities. AS data models are primarily used to capture and represent information related to complex observed phenomena with spatial and temporal coordinate systems (Nativi). Conventional GIS data models emphasize static representations of reality; they classify all features into vector (discrete features) and raster (continuous features) data models with accurate earth location information to meet the needs of complex and precise spatial queries. The GIS data model has difficulty managing temporal information to represent geographic dynamics, while the AS data model cannot match the richness of GIS geo-referencing information.

There are many research activities exploring the interoperability technologies currently available in the realms of both communities. One of these endeavors is the implementation of the Web Coverage Server specification. A prototype is being developed at Unidata with other groups. Another endeavor is the GIS Demonstration Project conducted at NCAR. The primary motivation of these projects is to ease data semantic transformation and resolve syntactic incompatibility. An ideal data model should allow data interoperability among different user communities. This paper will discuss the possibility of bringing together these two communities by linking the metadata of geospatial data models of two communities.

2 Challenges in Data Models

There are many discipline-specific data models in the AS and GIS communities. It is not feasible to encapsulate all these diverse datasets. Our discussion here focuses on a very small subset of data models which overlap the data models between these two communities. According to Nativi, the overlap of the data models is in the coverage category in most cases, grid-oriented datasets. GIS coverages are two dimensional metaphors for phenomena found on or near a portion of the Earth's surface. Nativi concludes that the coverage structure seems to be the best solution to bridge the GIS and AS data models. As we know, even in the coverage category each community has its own discipline-specific semantics in its data models. We selected a common data model from each community to further discuss the linkage of the two community's needs to develop the concepts to enable users in both communities to share geospatial information. The data models selected are the Network Common Data Format (netCDF) and GeoTIFF. We will discuss the advantages of a conceptual common data model (CDM) through these two implemented data models of the GIS and AS communities.

3. Linking the metadata of netCDF and geotiff to each other's data model

NetCDF metadata contains all the information about the dimensions, attributes, and variables except for the variables data. The netCDF example selected is a standard four-dimensional, multi-variable AVN model output dataset. The metadata information extracted from this dataset has the following structure:

```
netcdf 2003072918_avn-x {
   dimensions:
      record = UNLIMITED ; // (29 currently)
      lat = 73 ;
      lon = 73 ;
      level = 12 ;
      time_len = 21 ;
      valtime_offset = 29 ;
  variables:
      float lat(lat) ;
            lat:long_name = "latitude" ;
            lat:units = "degree_north" ;
      float lon(lon) ;
```


[1] *Corresponding author address*: Yuan Ho, UCAR Unidata, P.O.Box 3000, Boulder CO. 80307-3000; Email: yuanho@ucar.edu.

The Unidata Program Center is sponsored by NSF and managed by UCAR.

```
            lon:long_name = “longitude” ;
            lon:units = “ degrees_east” ;
       float T(record, level, lat, lon);
            T:long_name = “Temperature at isobaric levels” ;
            T:standard_name = “air_temperature” ;
            T:units = “degK” ;
      // global attributes
      :record = "reftime, valtime" ;
      :history = "2003-03-26 16:08:28 - created by gribtocdl" ;
      :title = "126 Wave, 18 Layer Spectral Model Aviation Run" ;
      :Conventions = "NUWG" ;
      :GRIB_reference = "Office Note 388 GRIB" ;
      :GRIB_URL = http://www.nco.ncep.noaa.gov/pmb/docs/on388/ ;
    :version = 0. ;</c></p>

data:

lat = -90, -87.5, -85, -82.5, -80, -77.5, -75, …, 85, 87.5, 90 ;
lon = -30, -25, -20, -15, -10, -5, 0, …, 310, 315, 320, 325, 330 ;
nav_model = "GRIB1" ;
grid_type_code = 0 ;
grid_type ="Latitude/Longitude" ;
grid_name ="Global 5.0 x 2.5 degree grid" ;
grid_center = 7 ;
grid_number = 255 ;
Ni = 73 ;
Nj = 73 ;
La1 = -90 ;
Lo1 = -30 ;
La2 = 90 ;
Lo2 = 330 ;
Di = 5 ;
Dj = 2.5 ;
…
}
```

GeoTIFF is a common raster file format. Its specification is an extension to the TIFF specification. It uses a small set of reserved TIFF tags to store a broad range of geo-referencing information, catering to geographic as well as projected coordinate system needs. A TIFF file begins with a file header and then a series of image file directories. Each image file directory (IFD) describes the core image properties of a raster image, including the geospatial coordinate system. IFD has both structural metadata (i.e. TIFF header) and geo-spatial metadata (i.e. GeoTIFF header). From the above netCDF metadata structure a variety of information is transported to the IFD structure:

```
 NewSubfileTypeTag = 2
 ImageWidthTag = 73
 ImageLengthTag = 73
 BitsPerSampleTag = 8
 CompressionTag = 1
 StripOffsetsTag = 8
 OrientationTag = 1
 SamplesPerPixelTag = 1
 RowsPerStripTag = 30
 StripByteCountsTag = 21316
 PlanarConfigurationTag = 1
 SMinSampleValue = 233.5;
 SMaxSampleValue = 318.7;
 PlanarConfigurationTag = 1
 ResolutionUnitTag = 1
 PageNumberTag = 0
 ModelTiepointTag       = (0.0, 0.0, 0.0, -180.0, 90.0, 0.0)
 ModelPixelScaleTag      = (5.0, 2.5, 0.0)
 GeoKeyDirectoryTag:
      GTModelTypeGeoKey        =  2      (Geographic)
        GTRasterTypeGeoKey       =  1      (RasterPixelIsArea)
        GeographicTypeGeoKey     =  4326   (WGS_84)
        GeoPrimeMeridianGeoKey   =  8901   (Greenwich)
        GeoAngularUnitsGeoKey    =  8      (Degree)
```

The data semantic transformation between netCDF and GeoTIFF shown in this example is rather complicated due to the mismatch in metadata granularity. The geo-referencing information in netCDF is kept as simple as possible; therefore the semantic information in netCDF is insufficient. There are few obvious one to one relations between netCDF attributes and geotiff tags. We need to do the transformations from index space to a latitude-longitude geo-referencing coordinate system. Adding standard geo-referencing metadata information is necessary to fully construct the semantic metadata of geotiff. For example, associating the data dimensions to the geo-referencing coordinates, i.e. mapping the lat and lon to ModelTiepointTag, we not only need to extract the lat and lon data value, we also need to rotate the data coordinate to (-180, 180) longitude range and flip the data to (90, -90) latitude range, which is the conventional geo-referencing coordinate system. Defining the GeoKeyDirectoryTag is based on the common geographic coordinate system WGS_84. This information, along with other associated geo-referenced metadata, is not contained in the netCDF dataset. ImageWidthTag and ImageLengthTag are two of those few tags mapping directly from netCDF attributes. SMinSampleValue and SMaxSampleValue are extracted from the data value field. In this exercise we realize the syntactic metadata, i.e. the information about the data types and structures, are easier to fit into the other's requirement. However the semantic metadata, i.e. the information about the content of the data, what the variables mean, their ranges, and so on, are quite difficult to communicate between the AS and GIS communities

4 A conceptual common data model

Within each discipline the semantic objects of netCDF and GeoTIFF required to describe a dataset are sufficient and unique, but it is a challenge to relate the geo-referencing information when the dataset needs to be distributed to both communities. A netCDF translator can provide the linkage of metadata between an AS data model and a GIS data model. A crucial factor for this translator to work efficiently is the standardization of metadata conventions in the AS data model which implies a conceptual common data model. A conceptual common data model provides a framework to overcome the different cultures and histories of collecting and archiving the data, and the differences in conceptualizations and interpretations of geographic world.

There are several key elements in this conceptual data model:

1) Unambiguous syntaxes cross the boundaries of AS and GIS communities
2) Consistent geo spatial representations
3) Standard set of attributes
4) Standard APIs for data and metadata transformation
5) Adding geo-referencing capabilities to AS data model

Effective and efficient access to data is critical to the development of a geospatial data model in each community. We believe implementing these key elements provides a logical framework for enhancing data interoperability between AS and GIS.

5 Summary and conclusions

There are many approaches to exploring data interoperability between the AS and GIS communities. We believe better understanding of the relationship of data structures, usages, or semantics between the two communities is essential for success. This paper discusses the linkage of metadata of two data models, from AS and GIS. We explore the possibility of identifying a conceptual CDM. While it is our hope of to bridge the gap of semantic metadata between AS and GIS , we realize a CDM can only answer the compatibility issue of syntactic metadata. However, for the data interoperability challenge between the AS and GIS, a conceptual data model implementation in both communities can ease the transformation of semantic metadata. Each data model can link to another, and AS data can be provided to a broad GIS user community.

6 References

Stefano Nativi, “Differences among the data models used by the geographic information systems and atmospheric science communities”, AMS 2002.

Terri L. Betancourt and Olga Wilhelmi, “Final Report on the GIS Demonstration Project”.

John Caron and Stefano Nativi, “NetCDF to GIS: the Good, the Bad, and the Ugly”, AMS 2004.