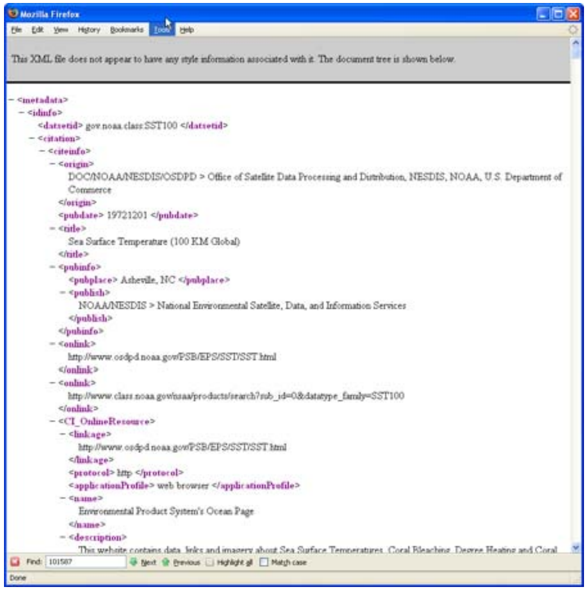
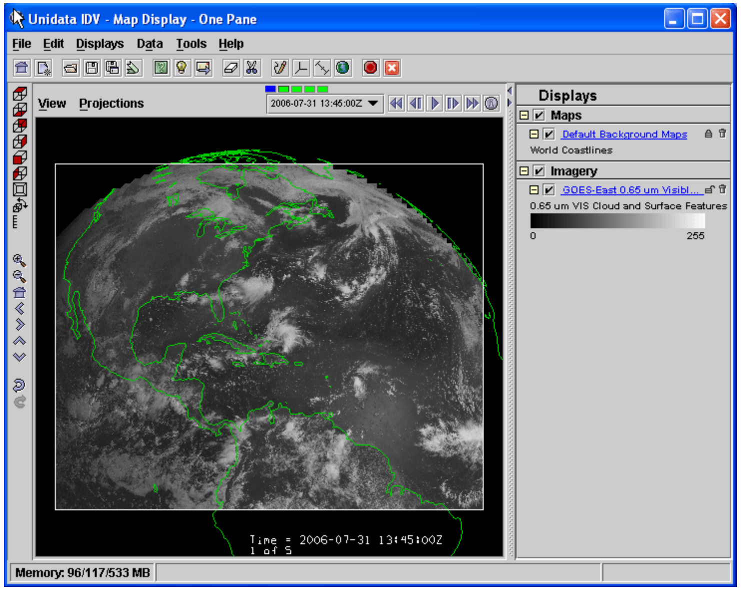
Standard Metadata in Scientific Data Formats

Ted.Habermann@noaa.gov

GEOSS Harmonization - 2009



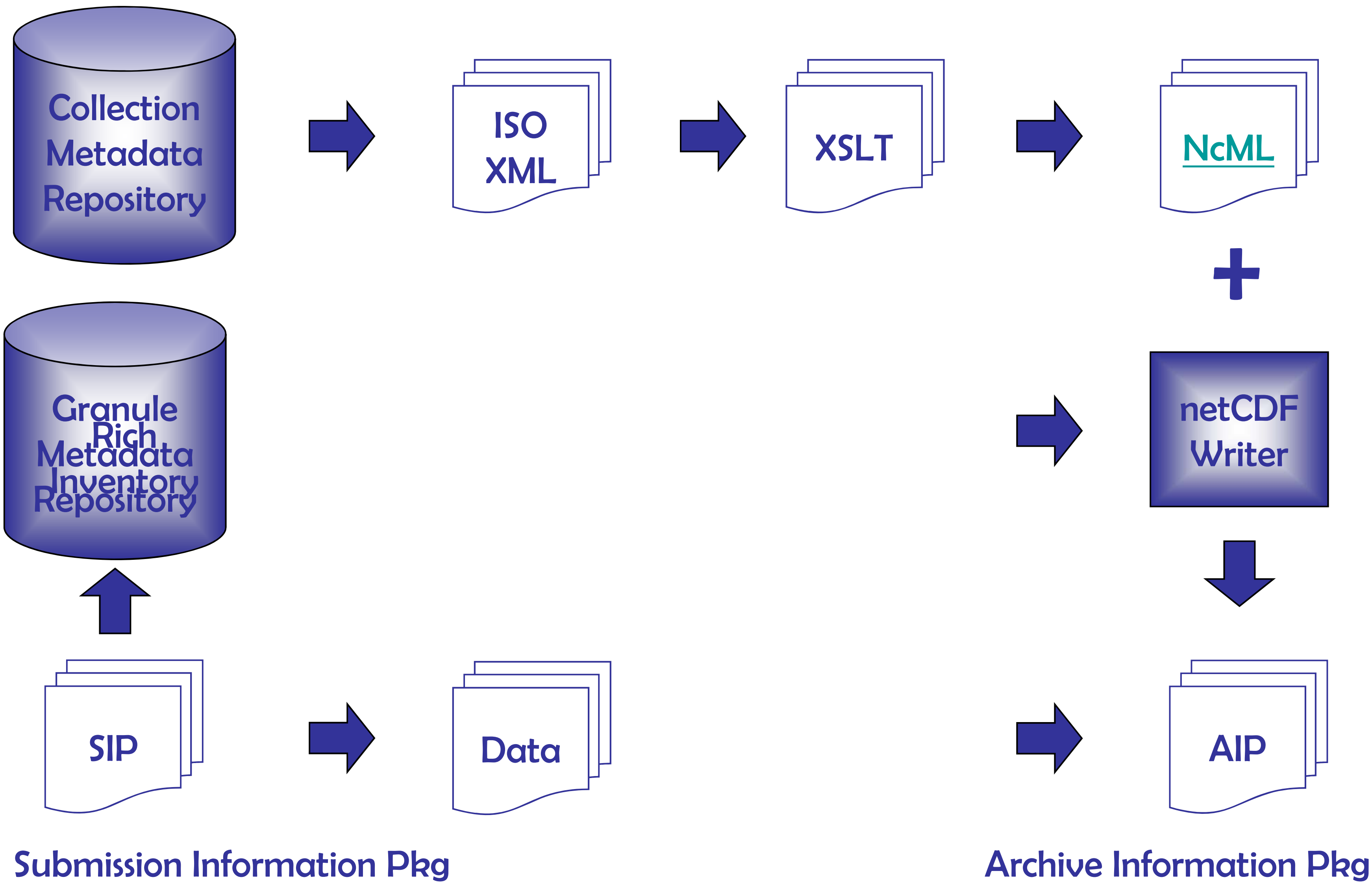


Observations of Metadata in netCDF Files

The obvious difference in the quantity and quality of metadata available in these two netCDF files prompted me to begin thinking about how we can take advantage of improving content in metadata repositories as we write data into scientific data formats.

Metadata Provided with a Level 3 SST Product From CLASS / SWFSC

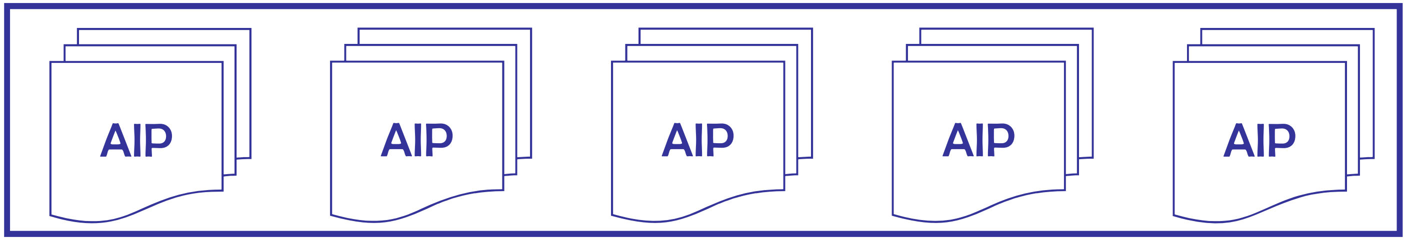
Metadata in this Process



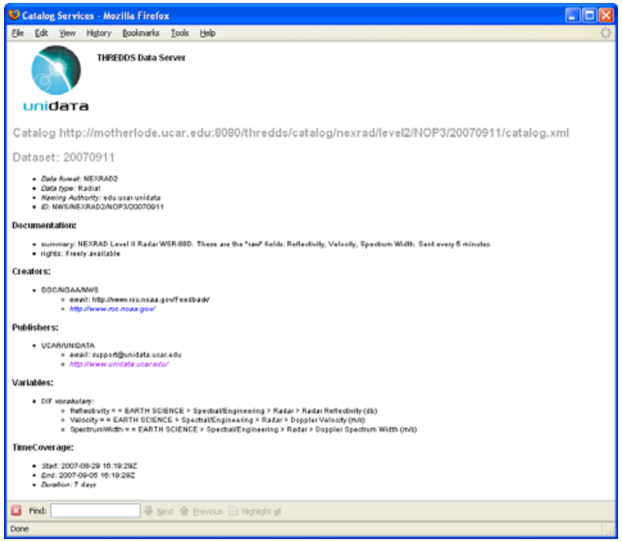
Archive Information Package Content


```
<attribute name="Conventions" type="String" value=""CF-1.0""/>
<attribute name="Metadata_Conventions" type="String" value="“ISO, DublinCore"/>
<attribute name="“ISO_Link" type="String" value="URL"/>
<attribute name="institution" type="String" value="NOAA / NGDC"/>
<attribute name="“iso_publisher_name" type="String" value="NOAA / NGDC"/>
<attribute name="dc_publisher" type="String" value="NOAA / NGDC"/>
<variable name="Analysis Temperature">
  <attribute type="string" name="iso_long_name" value="Analysis Temperature"/>
  <attribute type="float" name="“iso_valid_min" value="-850"/>
  <attribute type="float" name="“iso_valid_max" value="610"/>
  <attribute type="string" name="“iso_units" value="degrees Celsius * 10"/>
  <attribute type="float" name="min" value="150"/>
  <attribute type="float" name="max" value="220"/>
  <attribute type="string" name="units" value="Celsius/>
  <attribute type="float" name="scale_factor" value="0.1/>
</variable>
```


Metadata

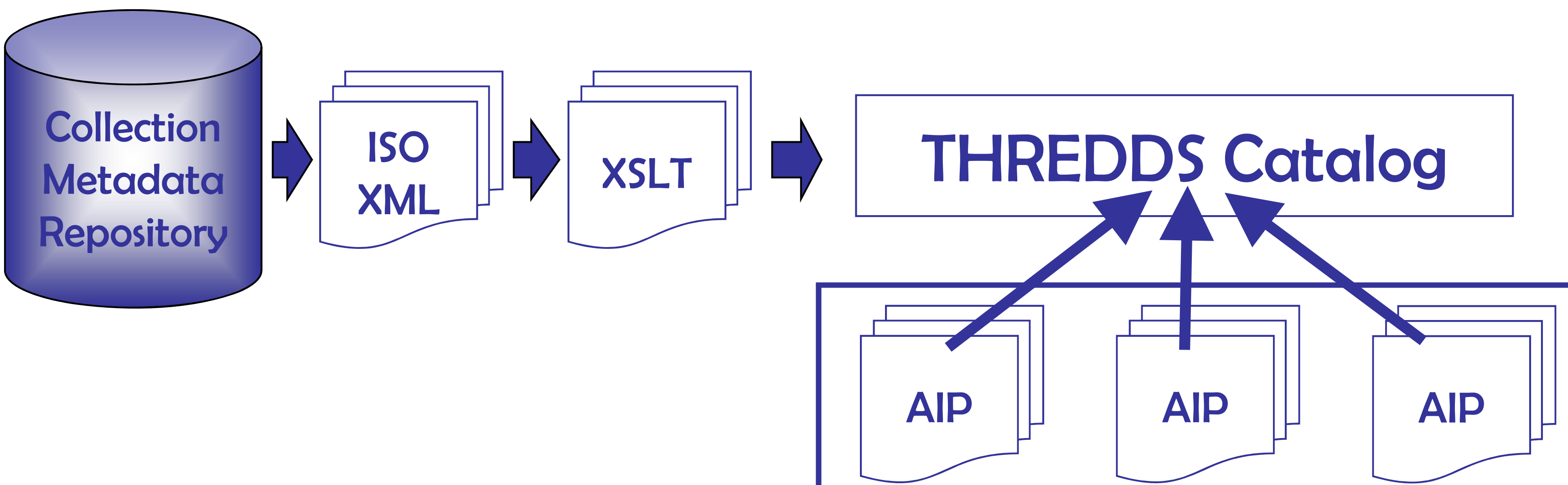


THREDDS Metadata

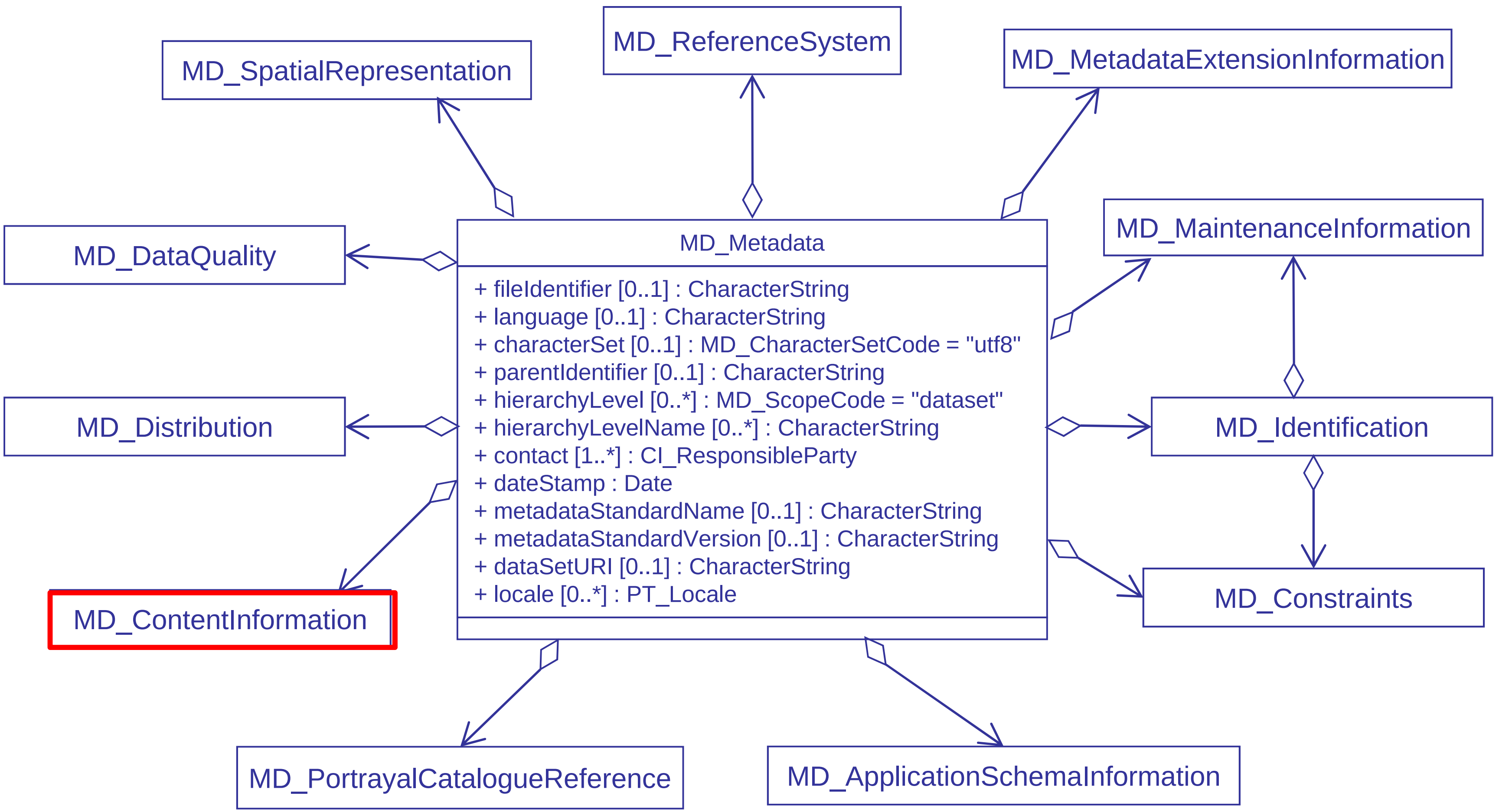


1) The THREDDS Catalog Generator can extract selected collection or granule metadata from the files and include it in the catalog.

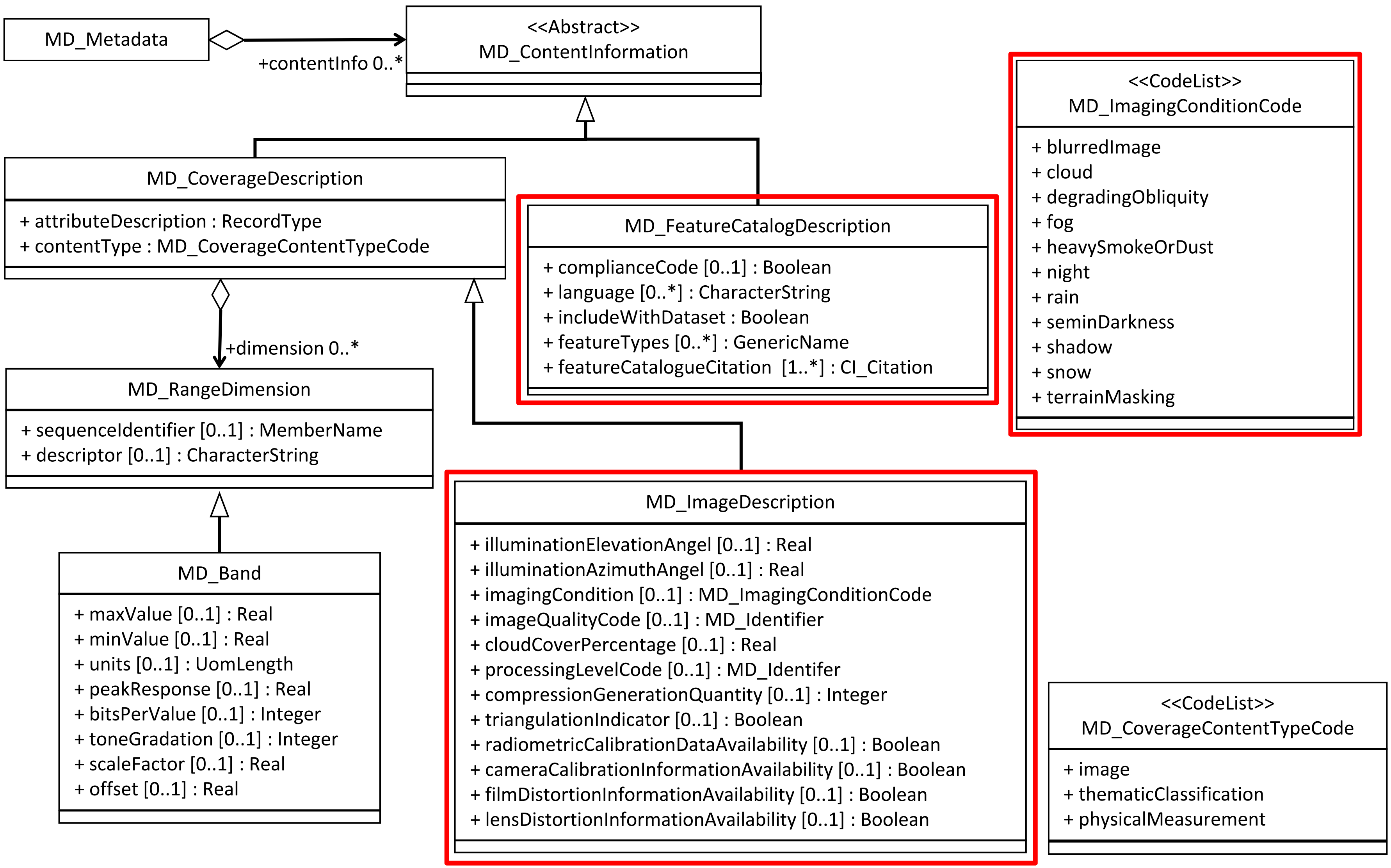
2) The Metadata_Link attributes could be extracted to support a THREDDS Metadata Service.

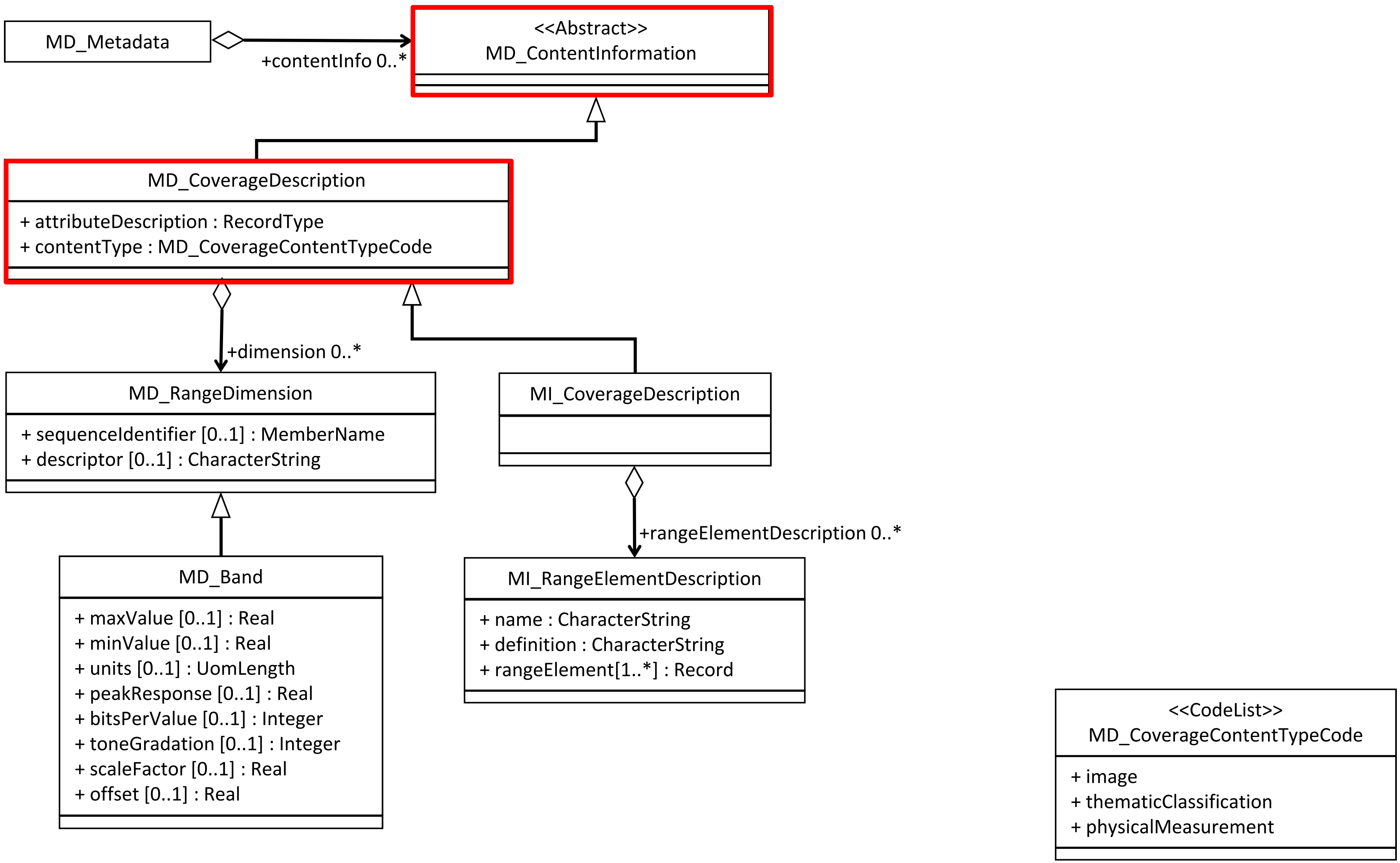


MD_ContentInformation



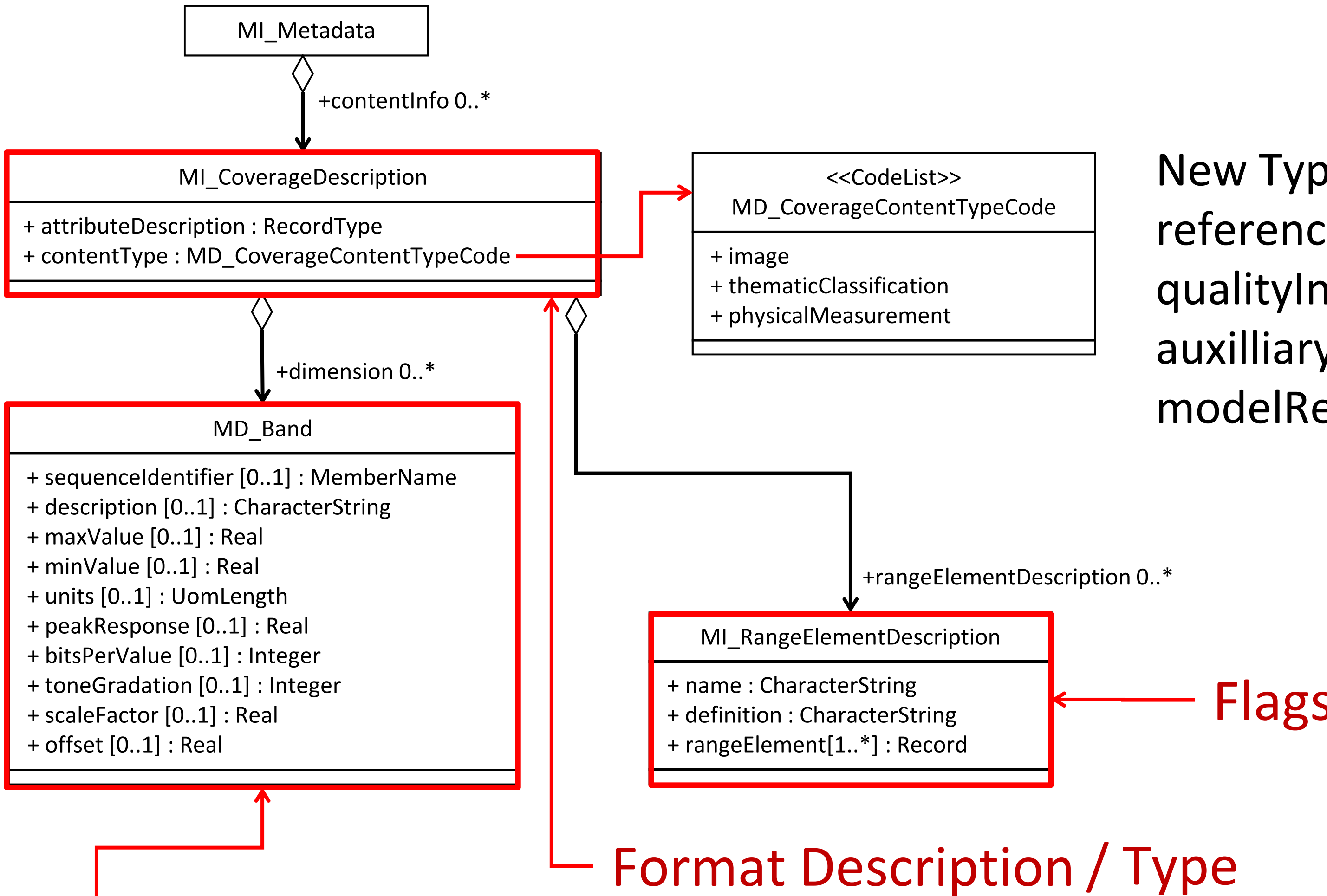
MD_ContentInformation





MI_ContentInformation 19115-2 (Grid Coverages)

MI_CoverageDescriptions

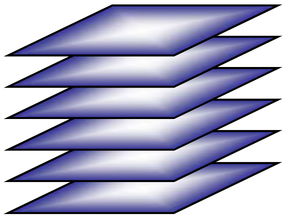


New Types:
referenceInformation
qualityInformation
auxilliaryData
modelResult

Layers, Member = Unique Identifier = Parameter Name

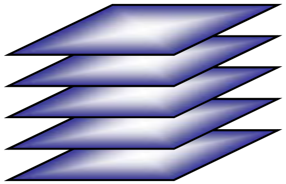
http://motherlode.ucar.edu:8081/thredds/cdmremote/fmrc/NCEP/GFS/CONUS_191km/files/GFS_CONUS_191km_20091111_0000.grib1?req=ncml

Simple multi-layered dataset (SST50)



Physical Measurements:

Analysis Temperature (sea_surface_temperature)	contentInfo 1
Gradient X+, X-, Y+, Y-, Average	
Ice Field	

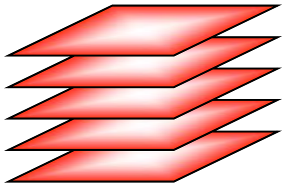


Auxilliary Data:

Physiographic Desc. (land_binary_mask)	contentInfo 2
Spatial Covariance X+, X-, Y+, Y-	contentInfo3

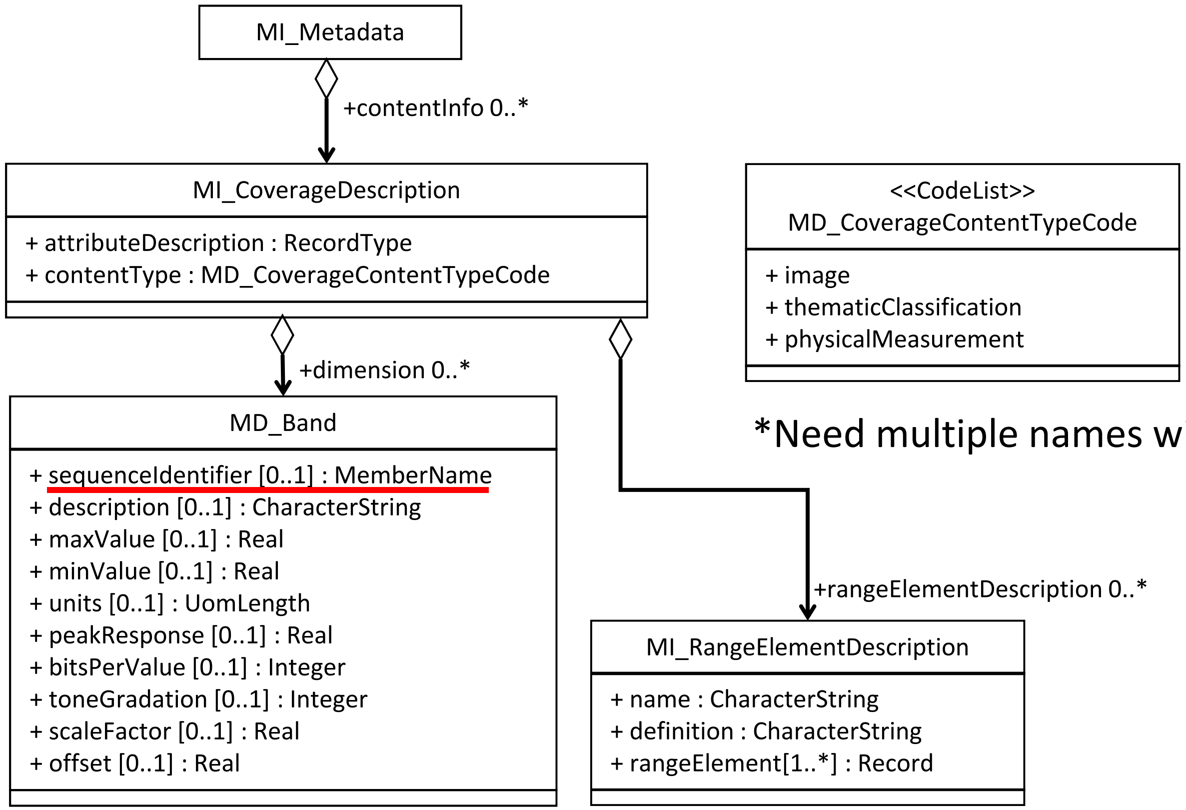
Quality Reports:

Lineage / Processing	dataQuality 1
Processing Parameters 1	dataQuality 2
Processing Parameters 2	dataQuality 3
Mission History	dataQuality 4



Quality Information:

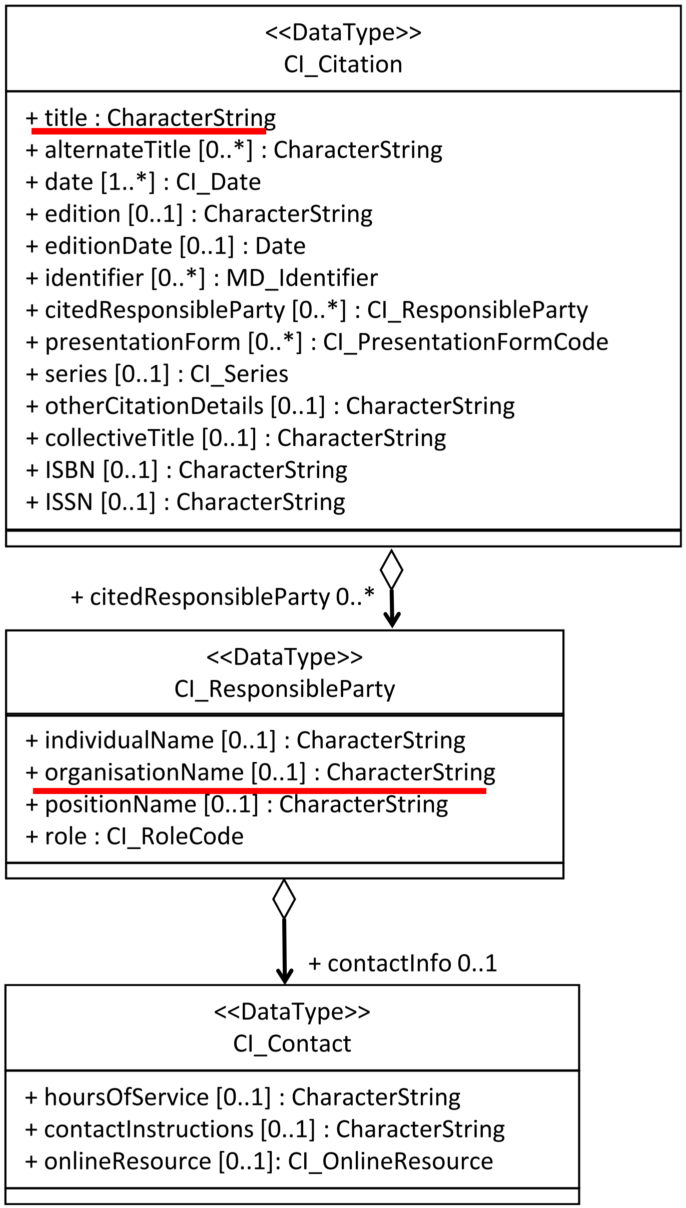
Age Of Most Recent Observation	dataQuality 5
Number of Observations	dataQuality 6
Class 1 Coverage	dataQuality 7
Reliability	dataQuality 8

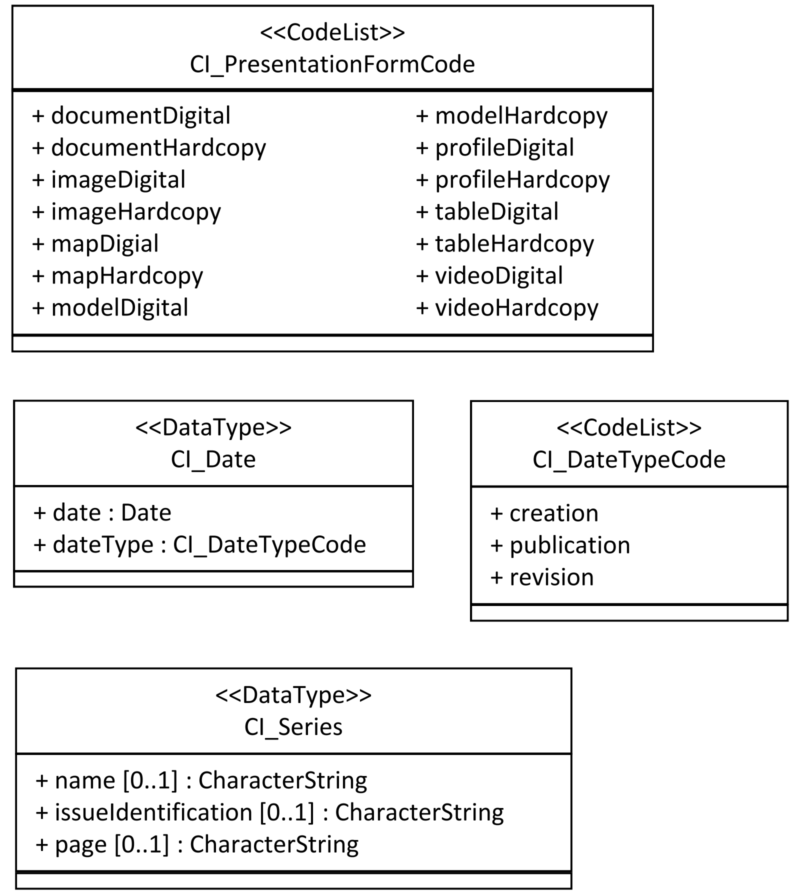


*Need multiple names with sources (MD_Identifiers)

```
<variable type="float" shape="lat lon" name="Optical Thickness">
  <attribute type="string" name="standard_name" value="atmosphere_optical_thickness_due_to_aerosol"/>
  <attribute type="string" name="long_name" value="Optical Thickness"/>
  <attribute type="float" name="valid_min" value="0"/>
  <attribute type="float" name="valid_max" value="2440"/>
  <attribute type="string" name="units" value="Optical Thickness Unit X 1000"/>
</variable>
```

MI_CoverageDescription (Grid Coverages)

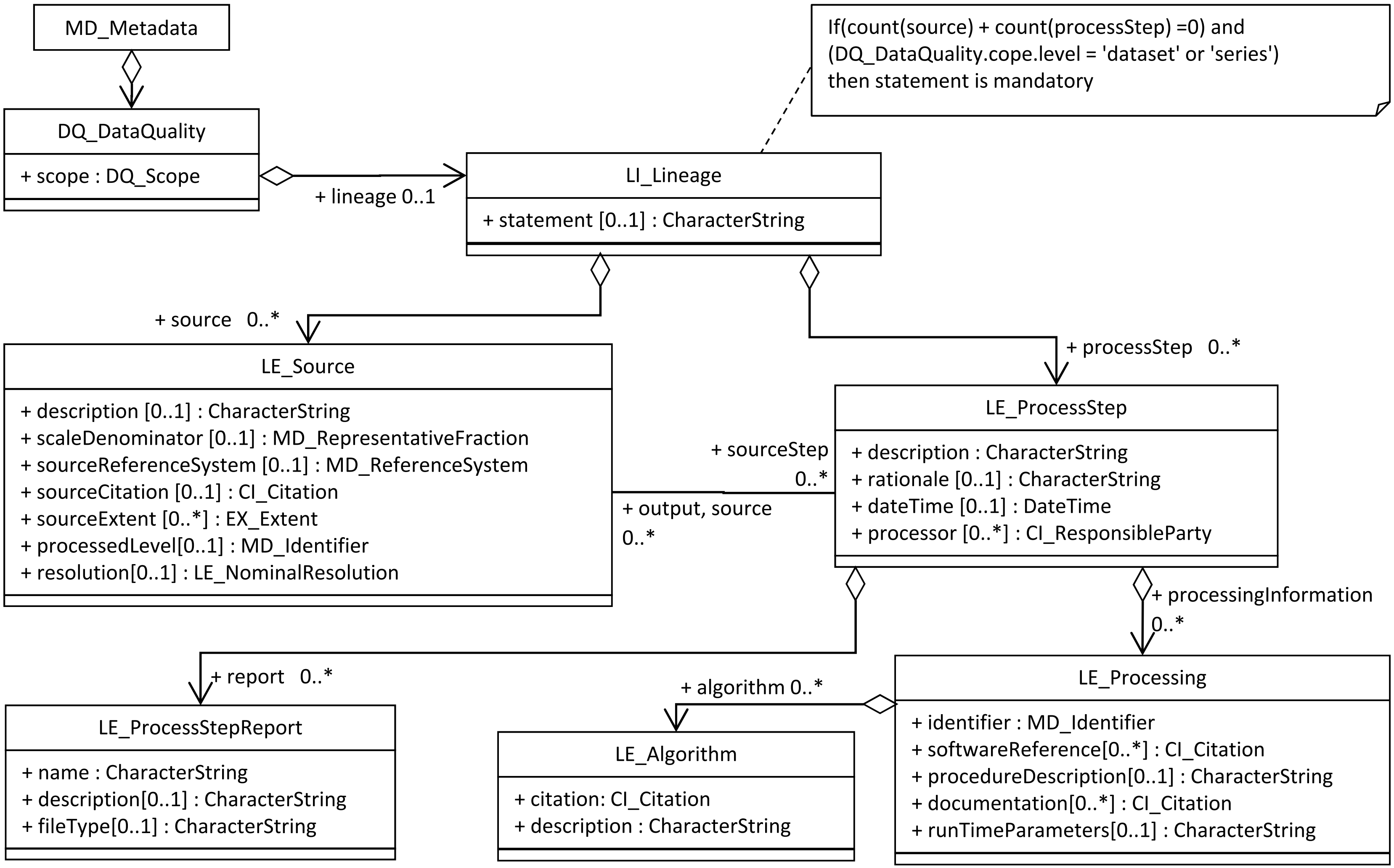




<attribute type="string" name="title“ value="Aerosol Optical Thickness (100 KM)"/>

<attribute type="string" name="institution" value="DOC/NOAA/NESDIS/OSDPD >Office of Satellite Data Processing and Distribution, NESDIS, NOAA, U.S. Department of Commerce"/>

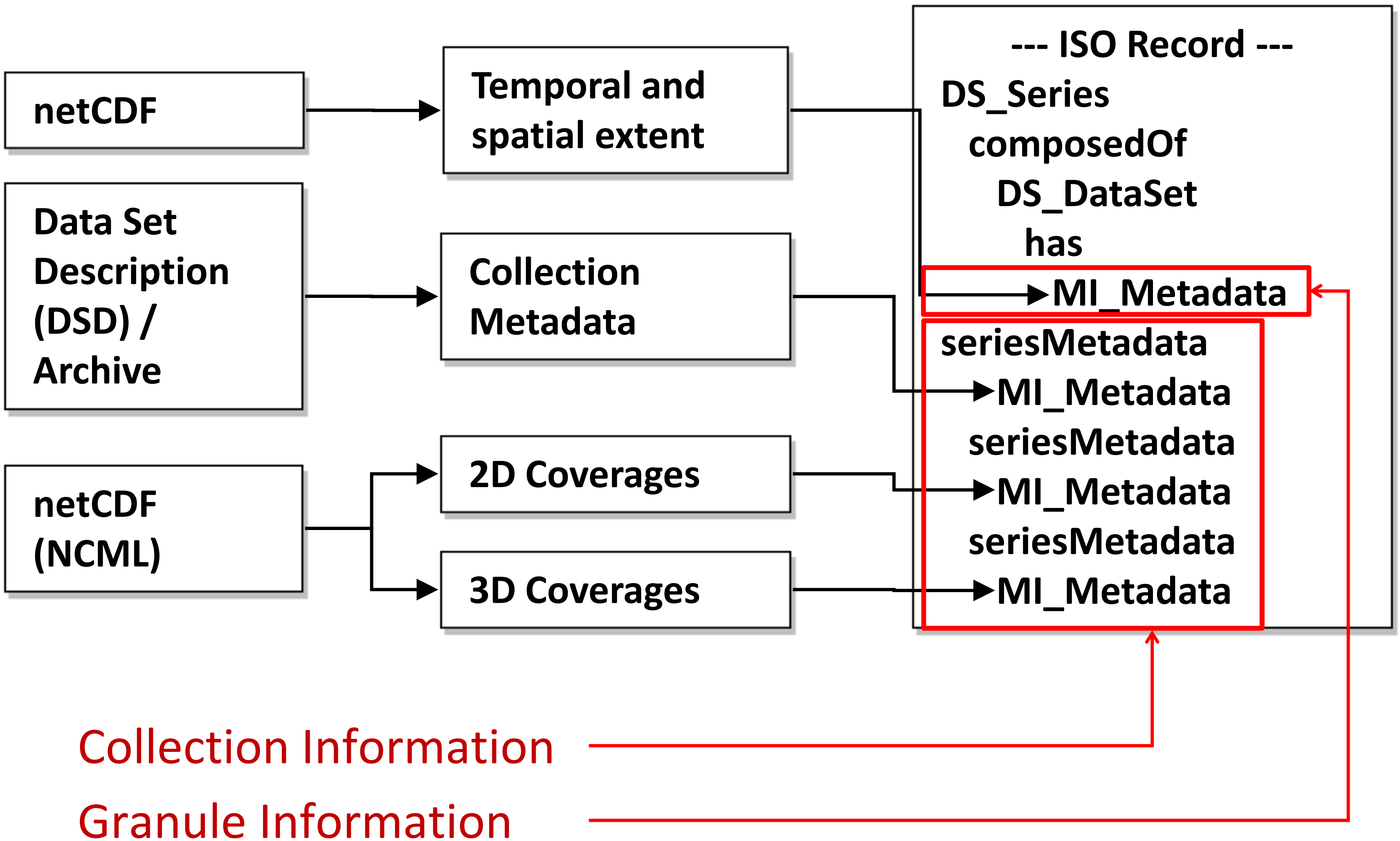
CI_CitationInformation



<attribute type="string" name="history" value=" Process Step: Orbital processing: Input level 1 data, calibration, cloud screening, sst and optical thicknesses computations. Gridded products of aerosol observations are produced weekly from the afternoon satellite data."/>

DQ_Lineage (19115-2)

Combining Collection and Granule Information



Conclusion

Most metadata content is independent of the standards used to represent it because it is designed to answer common questions about scientific data.

We can learn quite a bit by understanding how to represent the same answers in several ways.

Questions or suggestions: ted.habermann@noaa.gov