

# Metadata For Data Products in LEAD

2008-10-28

Scott Jensen

scjensen@cs.indiana.edu

Metadata is added for data products based on the LEAD Metadata Schema (LMS). As data products are added to a user's workspace (account) the hierarchy and metadata will be displayed in the workspace browser:

The screenshot displays the 'MyLead Workspace Browser' interface. The top bar shows the user's name, 'Hello, Scott Jensen, you are in your Personal Workspace', and a search button. The main area is split into two panels. The left panel, titled 'PERSONAL WORKSPACE', shows a hierarchical tree of projects and experiments. The right panel, titled 'INFORMATION FOR CURRENTLY SELECTED ITEM', displays detailed metadata for the selected file 'ad2008011715.nc'. The metadata includes fields such as Name, Type, GUID, Description, Owner, Create time, Purpose, Progress, Update, Access Constraint, Use Constraint, Theme, and a list of keys for the CF-1.0 theme.

Field	Value
Name	ad2008011715.nc
Type	file
GUID	urn:uuid:196d2d29-fa44-4beb-a568-b5b9f4a0aae0
Description	Real-time meteorological data assimilations with CONUS coverage at 10km resolution produced hourly by CAPS at OU. The List of contents provides the OPeNDAP URLs for the files within the collection. They have a form: http://lead.unidata.ucar.edu/cgi-bin/nph-dods/test-data/ADAS/OU/ad_date.nc where date has the form: YYYYMMDDHH and indicates the hour for which the data assimilation is valid.
Owner	/C=US/O=National Center for Supercomputing Applications/CN=Anne Wilson
Create time	Unknown null
Purpose	Scientific research and education
Progress	Complete
Update	Continually
Access Constraint	none
Use Constraint	none
Theme	DatasetTypes.lead.org key = ADAS
Theme	CF-1.0 key = projection_x_coordinate key = projection_y_coordinate key = height key = geopotential_height key = time key = x_wind key = y_wind key = upward_air_velocity key = air_potential_temperature key = air_pressure key = specific_humidity key = x_wind key = y_wind key = upward_air_velocity key = air_potential_temperature

When a data file is selected in the workspace browser, such as ADAS input file in the example above, the metadata for that file is displayed on the right-hand side of the screen. The metadata for all data products in LEAD is based on the LEAD Metadata Schema (LMS). There is a minimal amount of metadata that is required to validate against the LMS, but data products need some additional metadata for the Workflow Composer to be able to find them when constructing workflows. In addition, since all of the metadata will be displayed and searchable, having metadata beyond the minimum needed to validate against the schema makes the data more useful.

When adding metadata to the metadata catalog programmatically, there is a java utility in LEAD (that is available as a jar) called the Minimal LEAD Metadata Utility. This utility allows you to build an XML document that contains the fields necessary to validate against the LMS. The constructor for the only class in this utility takes three schema elements as inputs – the origin, the title, and the abstract. In the API documentation these are referred to as the DN, name, and description. The remaining required fields are

populated with their default values. To include additional metadata beyond these required fields, call the `getLeadResourceDoc()` method and it will return the XML bean which is an instance of the `LeadResourceDocument` class from the LMS.

Following is a list of the required fields (and defaults where applicable). Details as to the meaning of the fields can be found in the FGDC workbook. The LEAD schema is a profile of the FGDC.

LEADresource / resourceID<sup>1</sup>  
LEADresource / data / idinfo / citation / origin  
LEADresource / data / idinfo / citation / pubdate (*default is current date*)<sup>2</sup>  
LEADresource / data / idinfo / citation / title  
LEADresource / data / idinfo / descript / abstract  
LEADresource / data / idinfo / descript / purpose (*default is same as abstract*)  
LEADresource / data / idinfo / status / progress (*default is "In work"*)  
LEADresource / data / idinfo / status / update (*default is "Unknown"*)  
LEADresource / data / idinfo / acconst (*default is "Managed by MyLEAD"*)  
LEADresource / data / idinfo / useconst (*default is "Unknown"*)  
LEADresource / data / idinfo / keywords / theme / themekt<sup>3</sup> (*default is "leadproject.org"*)  
LEADresource / data / idinfo / keywords / theme / themekey<sup>3</sup> (*default is a URI for LEAD*)  
LEADresource / data / metainfo<sup>4</sup>

1. The resource ID will be added by the myLEAD Agent when a data product is registered if it does not already have a resourceID.
2. The minimal metadata utility will also populate the pubtime (which is not a required field) with the current local time and the UTC offset.
3. The LMS requires one instance of the element theme, which contains a themekt and at least one themekey. For data products, a theme keyword will be needed for the data type (see below) so that element would meet this requirement, but the minimal metadata utility will add the defaults listed above.
4. The metainfo element has three required child elements, but these relate to what metadata standard was used to describe the data. In the case of LEAD, this will always be the LMS, so the minimal metadata utility always populates these fields. Please see the LMS for the individual elements if you want further details.

There are additional metadata fields that must be populated for each data product. As mentioned above, these additional fields are required for the Workflow Composer to find the data products when constructing workflows. The additional fields required for each data type are the temporal bounds (the `rngdates` element within the `geospatial` element – see below), the spatial bounds (bounding which is also within the `geospatial` element), a theme element for the data type (listed separately below for each data type), and the format name (the `formname` element within the `distinfo` element).

The format name (`formname`) is on the following path in the LMS:

LEADresource / data / distinfo / stdorder / digform / digtinfo / formname

The syntax of the `distinfo` element that the format name is contained in also requires that there be a `distrib` child of `distinfo` that contains contact information. If possible, it would also be desirable to include the value for the size of the data product along with the

format name (the transize element).

**All products must have spatial and temporal bounds such as the following:**

```
<lead:geospatial>
  <lead:idinfo>
    <lead:timeperd>
      <fgdc:timeinfo>
        <fgdc:rngdates>
          <fgdc:begdate>20081027</fgdc:begdate>
          <fgdc:begtime>150000000</fgdc:begtime>
          <fgdc:enddate>20081027</fgdc:enddate>
          <fgdc:endtime>160000000</fgdc:endtime>
        </fgdc:rngdates>
      </fgdc:timeinfo>
      <fgdc:current>unknown</fgdc:current>
    </lead:timeperd>
    <lead:spdom>
      <fgdc:bounding>
        <fgdc:westbc>-136.6</fgdc:westbc>
        <fgdc:eastbc>-59.05999999999999</fgdc:eastbc>
        <fgdc:northbc>48.5</fgdc:northbc>
        <fgdc:southbc>19.64603</fgdc:southbc>
      </fgdc:bounding>
    </lead:spdom>
    <lead1:vertdom>
      <lead1:lowerb>-200.0</lead1:lowerb>
      <lead1:upperb>20200.0</lead1:upperb>
    </lead1:vertdom>
  </lead:idinfo>
</lead:geospatial>
```

## Data Product Specific Metadata

### Meteorology Aviation Routine report (METAR)

```
<fgdc:theme>  
  <fgdc:themekt>DatasetTypes.lead.org</fgdc:themekt>  
  <fgdc:themekey>METAR</fgdc:themekey>  
</fgdc:theme>
```

*Also has the format name (NetCDF)*

### Upper Air: Rawinsondes

```
<fgdc:theme>  
  <fgdc:themekt>DatasetTypes.lead.org</fgdc:themekt>  
  <fgdc:themekey>Rawinsonde</fgdc:themekey>  
</fgdc:theme>
```

*Also has the format name (NetCDF)*

### North American Model (NAM)

```
<fgdc:theme>  
  <fgdc:themekt>VisTools.lead.org</fgdc:themekt>  
  <fgdc:themekey>IDV</fgdc:themekey>  
</fgdc:theme>  
<fgdc:theme>  
  <fgdc:themekt>DatasetTypes.lead.org</fgdc:themekt>  
  <fgdc:themekey>NAM</fgdc:themekey>  
</fgdc:theme>
```

*Also NAM property with elements for offset and resolution*

*Also has the format name (GRIB)*

### NEXRAD Level II

```
<fgdc:theme>  
  <fgdc:themekt>VisTools.lead.org</fgdc:themekt>  
  <fgdc:themekey>IDV</fgdc:themekey>  
</fgdc:theme>  
<fgdc:theme>  
  <fgdc:themekt>DatasetTypes.lead.org</fgdc:themekt>  
  <fgdc:themekey>NEXRAD Level II</fgdc:themekey>  
</fgdc:theme>
```

*Also has the format name (NEXRAD2)*

### NEXRAD Level III

Currently we have no NEXRAD Level III data in the data catalog

### GOES: Visible and Infrared Imagery

Currently we have no GOES data in the data catalog (the option is disabled in the selection box)

### **Non-NAM NCEP Gridded Analyses and Forecasts**

Currently we have no data of this type in the data catalog (the option is disabled in the selection box)

### **ARPS Data Analysis System (ADAS)**

```
<fgdc:theme>  
  <fgdc:themekt>DatasetTypes.lead.org</fgdc:themekt>  
  <fgdc:themekey>ADAS</fgdc:themekey>  
</fgdc:theme>
```

*Currently this usually has extensive CF-1.0 keywords added by UNIDATA  
Also has the format name (NetCDF)*

### **Unidata Steered WRF (WRF)**

```
<fgdc:theme>  
  <fgdc:themekt>VisTools.lead.org</fgdc:themekt>  
  <fgdc:themekey>IDV</fgdc:themekey>  
</fgdc:theme>  
<fgdc:theme>  
  <fgdc:themekt>DatasetTypes.lead.org</fgdc:themekt>  
  <fgdc:themekey>WRF</fgdc:themekey>  
</fgdc:theme>
```

*Also has the format name (GRIB-1)*

### **Wind Profiler**

```
<fgdc:theme>  
  <fgdc:themekt>DatasetTypes.lead.org</fgdc:themekt>  
  <fgdc:themekey>wind profiler</fgdc:themekey>  
</fgdc:theme>
```

*Also has the format name (NetCDF)*

### **ACARS**

Currently we have no ACARS data in the data catalog (the option is disabled in the selection box)

### **RASS**

Currently the RASS data in the data catalog is listed as wind profiler data.

### **NWS National Digital Forecast Data Base (NDFD) Grids**

Currently we have no data of this type in the data catalog (the option is disabled in the selection box)