

Member Nodes - Task #6047

MNDeployment # 4247 (Deprecated): EDORA (Environmental Data for the Oak Ridge Area)

Validate EDORA content against the ORNL Mercury schema

2014-08-18 23:04 - Chris Jones

Status: Closed	Start date: 2014-08-18
Priority: Normal	Due date:
Assignee: Chris Jones	% Done: 100%
Category:	Estimated time: 0.00 hour
Target version: Operational	
Story Points:	
Description The EDORA MN stores science metadata in a custom format defined by the ORNL DAAC group, which uses FGDC elements and elements specific to the Mercury software system. Validate the content against the most recent schema (attached), and work with Jim and Ranjeet to get the schema properly published under a versioned namespace.	
Subtasks: Task # 6048: Fix 404 NotFound errors on the EDORA MN New Task # 6053: Update science metadata files to include ORNL schema type In Progress	

History

#1 - 2014-08-18 23:29 - Chris Jones

- Status changed from New to In Progress

In checking for content availability, I'm seeing a few 404 NotFound errors:

Walker_Branch_Watershed_Daily_Climate__1993-2010.xml	404
Walker_Branch_Watershed_Daily_Climate__1993-2008.xml	404
Throughfall_Displacement_Experiment__Ecosystem_Model_Intercomparison_Data.xml	404

The error was, for example:

```
<?xml version="1.0" encoding="UTF-8"?>
```

No system metadata could be found for given PID: Walker_Branch_Watershed_Daily_Climate_.xml

I'll ask Jim and Ranjeet to look into this.

#2 - 2014-08-19 23:51 - Chris Jones

- Assignee changed from Chris Jones to Ranjeet Devarakonda

Ranjeet and Jim,

On Aug 14, 2014, at 7:21 AM, Jim Green jgreen@iiaweb.com wrote:

Hi Chris,

The attached xsd file should validate daac, edora and rgd.
I've verified it against all of the science metadata files for these three mercury instances.

Unfortunately, I'm still seeing a number of schema validation errors. I've used two different parsers (xmllint and the Xerces Java Parser), and both encounter the same errors:

```
cjones@ridgway:edora$ xmllint --noout --schema mercury-ornl-v2.xsd results/Map_ORR_Wetlands_Vector.xml
results/Map_ORR_Wetlands_Vector.xml:55: element themekt: Schemas validity error : Element 'themekt': This element is not expected. Expected is (
themekey ).
results/Map_ORR_Wetlands_Vector.xml:92: element computer: Schemas validity error : Element 'computer': Character content other than whitespace
is not allowed because the content type is 'element-only'.
results/Map_ORR_Wetlands_Vector.xml:92: element computer: Schemas validity error : Element 'computer': Missing child element(s). Expected is (
networka ).
results/Map_ORR_Wetlands_Vector.xml:154: element Documentation_Link: Schemas validity error : Element 'Documentation_Link': This element is
not expected. Expected is ( OME_Software_Version ).
results/Map_ORR_Wetlands_Vector.xml fails to validate
```

For instance, the first error involves the `themekt` element. The schema file defines the content of `/metadata/idinfo/keywords/theme` as only allowing one theme keyword thesaurus and multiple theme keywords in the sequence. In many of the science metadata documents, the `themekt` element shows up multiple times.

The second issue is that the content of `/metadata/distinfo/stdorder/digform/digtopt/onlinopt/computer` is element-only, with child elements of `networka` and a grandchild of `networka`. These are missing from the science metadata documents.

There are other errors that I didn't track down, but you get the gist. These content issues will need to be addressed so they fully validate against the schema, so I'm assigning this ticket to Ranjeet for now.

#3 - 2014-08-20 13:51 - Chris Jones

- Assignee changed from Ranjeet Devarakonda to Jim Green

#4 - 2014-08-22 17:06 - Ranjeet Devarakonda

- translation missing: en.field_remaining_hours set to 0.0
- File mercury-ornl-v3.xsd added
- Assignee changed from Jim Green to Chris Jones
- File mercury-ornl-v3.xsd added
- Status changed from In Progress to Closed

Updated schema (mercury-ornl-v3.xsd) to fix all validation issues. Testing using xmllint against DAAC, EDORA and RGD records. See results below:

```
$ xmllint --noout --schema mercury-ornl-v3.xsd /data/Mercury_instances/dataone_mn/edora/harvested/*.xml
/data/Mercury_instances/dataone_mn/edora/harvested/Map_Counties_Surrounding_ORR_Land_Cover_Landsat_NLCD_30m_1992.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_Counties_Surrounding_ORR_Land_Use_Settings_SAMAB_90m_1992.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_Counties_Surrounding_ORR_Wetlands_Vector.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Aspect_2m_1993.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Bridges_Vector.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Buildings_Vector.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Clinch_River_Vector.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Digital_Elevation_Model_2m_1993.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Geologic_Formations_Vector_Coverage.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Land_Cover_LandsatTM_25m_1994.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Land_Cover_LandsatTM_30m_1984.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Land_Cover_LandsatTM_NLCD_30m_1992.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Land_Cover_LandsatTM_SAMAB_30m_1992.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Land_Cover_SPOT_60m_1987.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Pine_Beetle_Salvage_Vector_1999.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Pine_Beetle_Spots_Vector_1999.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Soil_Inventory_Vector.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Streams_Vector.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Map_ORR_Wetlands_Vector.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Oak_Ridge_TNDaily_and_Monthly_Climate_Data_1947-2000.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/ORNL_Free-Air_CO2_Enrichment(FACE)Experiment_Data.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Throughfall_Displacement_ExperimentEcosystem_Model_Intercomparison_Data.xml
validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Throughfall_Displacement_Experiment_Data_Report_CDIAC_NDP-078A.xm
l validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_AmeriFlux_CO2_Water_Vapor_and_Energy_Exchange_1995-
1998.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_Atmospheric_Deposition_Data.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_Daily_Climate1993-2008.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_Daily_Climate1993-2010.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_Monthly_Climate_Data_1951-1994.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_Precipitation.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_Stream_Chemistry.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_Stream_Discharge_and_Annual_Runoff.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_Vegetation_Inventory_1967-1997.xml validates
/data/Mercury_instances/dataone_mn/edora/harvested/Walker_Branch_Watershed_Vegetation_Inventory_1967-2006.xml validates
```

#5 - 2014-09-29 21:25 - Laura Moyers

- Target version changed from *Deploy by end of Y1Q1* to *Deploy by end of NCTE*

#6 - 2014-11-24 14:57 - Laura Moyers

- Target version changed from *Deploy by end of NCTE* to *Operational*

Files

mercury-ornl-v2.xsd	15.7 KB	2014-08-18	Chris Jones
mercury-ornl-v3.xsd	17.4 KB	2014-08-22	Ranjeet Devarakonda