6-21-2011

Developing Metadata for the DRInet Repository

Jake R. Carlson
Purdue University, jakecar@umich.edu

Developing Metadata for the DRInet Repository

Jake Carlson - Data Services Specialist - Purdue University Libraries / Distributed Data Curation Center - jrcarlso@purdue.edu

What is DRInet?

DRInet is a regional-scale drought information network. It seeks to advance the state of collection and synthesis of local and regional scale data in multiple domains that contribute to the national drought maps by providing an interoperability framework for making such data and assessment information available to a wide range of stakeholders.

Use Cases & Functionality of DRInet

- Assessing droughts & their impacts
- Determining drought abasement strategies
- Improve forecasting and monitoring of drought indicators
- Linkages between casual factors of drought.
- Synthesizing information for decision making
- Assessing variability and context in drought triggers
- Assessing causality and relation between drought and regional air &water quality
- Impact of drought on plant disease
- Economic (Impact) Assessment
- Drought education

Intended Audiences

- Climatologists
- Economists
- Town Managers
- Farmers
- Hydrologists
- Policy Makers
- Agriculture Experts
- County Extension Educators

Role of Metadata in DRInet

- **Discovery** – to enable searching and browsing of data within DRInet
- **Description** – to enable the user to understand the nature, purpose and potential uses of the data.
- **Define Relationships** between data set – to identify how a particular data set is related to another data set in DRInet, and to identify how a particular data set fits into DRInet as a whole.
- **Dissemination** – to enable others to find the data from outside of the DRInet portal. Metadata should be made harvestable by search engines, union catalogs and other external agencies as appropriate.

GCMD Visit GCMD website

Directory Interchange Format (DIF) Writer's Guide, 2010. Global Change Master Directory. National Aeronautics and Space Administration. [http://gcmd.nasa.gov/User/difguide/].

Sample DIF Metadata Record for DRInet Data

Entry ID - The unique document identifier of the metadata record.

<Entry_ID> [automatically generated] </Entry_ID>

<Data_Set_Citation>
<Dataset_Creator>Subramania I. Sritharan i</Dataset_Creator>
<Dataset_Title>Real time data collection of water quality parameters for the Massie Creek, Ohio from October 1, 2010 to October 20, 2010.</Dataset_Title>
<Dataset_Series_Name>Real time data collection of water quality parameters for the Massie Creek, Ohio</Dataset_Series>
<Dataset_Release_Date>[automatically populated from date of upload]</Dataset_Release_Date>
<Dataset_Release_Place>Purdue University, West Lafayette, IN</Dataset_Release_Place>
<Dataset_Publisher>driNET</Dataset_Publisher>
</Data_Set_Citation>

Data Set Citation – Allows the data set producer to be cited properly.

Personnel – Who to contact with questions about the data or its handling in the repository.

<Personnel>
<Role>TECHNICAL CONTACT</Role> <First_Name>Subramania</First_Name> <Last_Name>Sritharan</Last_Name> <Email>sri@centralstate.edu</Email>
<Role>DRInet ADMIN</Role> <First_Name>Lan</First_Name> <Last_Name>Zhao</Last_Name> <Email>lanzhao@purdue.edu</Email>
</Personnel>

ISO Topic – A thematic classification using a controlled vocabulary.

<Parameters>
<Category>EARTH SCIENCE</Category> <Topic>TERRESTRIAL HYDROSPHERE </Topic> <Term>WATER QUALITY/WATER CHEMISTRY</Term> <Variable_Level_1>Water Level</Variable_Level_1> <Variable_Level_2>Water Temperature</Variable_Level_2> <Variable_Level_3>Water ph</Variable_Level_3>
</Parameters>

Parameters – A hierarchical representation of keywords that describe the data set.

<ISO_Topic_Category>Inland Waters</ISO_Topic_Category>

<Data_Center>
<Data_Center_Name> <Long_Name>Drought Research Initiative Network (DRInet)</Long_Name> </Data_Center_Name> <Data_Center_URL> https://drinet.hubzero.org/</Data_Center_URL>
</Data_Center>

Data Center – The agency responsible for distributing the data (DRInet).

Summary – A description of the data set that include its history, development purpose, and use.

<Summary>
<Abstract>YSI Multi-parameter water quality radiosonde (YSI 600 series system) was installed on the left bank of the Massie Creek, OH by the bridge on Wilberforce-Clifton Road around January 2007. The same location (Lat 39°43'22", Long -83°52'58") also has a continuous USGS stream flow monitoring station (Hydrologic Unit 05090202) 1.7 mi upstream from the Clark Run, a designated tributary to the creek. The radiosonde water quality sensor has multiple probes to simultaneously measure water level, temperature, pH, dissolved oxygen, turbidity, and conductivity of the stream in real time. Water quality parameters are remotely monitored using radio-telemetry and processed using a Nexsens iChart 6software. The sampling frequency is 30 minutes. The base station also receives signals from a weather station. The WXT 520 weather station provides a complete sensor interface and data collection platform for simultaneously measuring wind speed and direction, precipitation, barometric pressure, temperature, and relative humidity. </Abstract>
<Purpose>Hydrological data received from the weather station, in stream water quality radiosonde, and the USGS stream flow monitoring station is being used in our water resources management and environmental engineering curricula to teach the principles of watershed hydrology using Massie Creek watershed as an example. Students majoring in Water Resources Management program have been using the data to study water quality on the Massie creek and calculate water quality index at different times of the year as part of the requirements for passing "WRM 3100 Streams and Lakes" course. This data may also be used in a large scale national study on development of drought triggers for agricultural applications. The study will be part of a USDA research project that has recently been awarded to Purdue University through its Agriculture and Food Research Initiative (AFRI) Competitive Grants Program.</Purpose>
</Summary>

Metadata Name – The name and version of the standard used (DIF).

<Metadata_Name> CEOS IDN DIF</Metadata_Name> <Metadata_Version>VERSION 9.8.2</Metadata_Version>

Sensor Name – Identifies the sensors used to generate the data.

<Sensor_Name> <long_name>YSI Multi-parameter water quality radiosonde (YSI 600 series system)</long_name></Sensor_Name>
<Sensor_Name> <long_name>WXT 520 weather station</long_name></Sensor_Name>

Coverage – Information about the temporal and spatial attributes of the data.

<Temporal_Coverage> <Start_Date>2010/10/01</Start_Date><Stop_Date>2010/10/20</Stop_Date> </Temporal_Coverage>

<Spatial_Coverage> <Latitude>39°43'22"</Latitude> <Longitude>- 83°52'58"</Longitude> </Spatial_Coverage>

Location – The named location where the data were generated.

<Location>
<Location_Category>CONTINENT</Location_Category> <Location_Type>NORTH AMERICA</Location_Type> <Location_Subregion1>UNITED STATES OF AMERICA</Location_Subregion1>
<Location_Subregion2>Ohio</Location_Subregion2> <Location_Subregion3>Greene</Location_Subregion3>
</Location>

Originating Center – The agency that originally generated the data.

<Originating_Center>Central State University. Wilberforce, OH</Originating_Center>

DIF Creation & Revision – Tracking information about the metadata record.

<DIF_Creation_Date> 2011-02-11</DIF_Creation_Date>
<Last_DIF_Revision_Date>2011-02-11</Last_DIF_Revision_Date>

PURDUE UNIVERSITY Libraries Access. Knowledge. Success.

National Science Foundation WHERE DISCOVERIES BEGIN