Use of a Metadata Documentation and Search Tool for Large Data Volumes: The NGEE Arctic Example

Ranjeet Devarakonda, Les Hook, Terri Killeffer, Misha Krassovski, Tom Boden, and Stan Wullschleger

Climate Change Science Institute
Oak Ridge National Laboratory
1 Bethel Valley Rd, Oak Ridge, TN USA

***Abstract*- The Online Metadata Editor (OME) is a web-based tool to help document scientific data in a well-structured, popular scientific metadata format. In this paper, we will discuss the newest tool that Oak Ridge National Laboratory (ORNL) has developed to generate, edit, and manage metadata and how it is helping data-intensive science centers and projects, such as the U.S. Department of Energy's Next Generation Ecosystem Experiments (NGEE) in the Arctic to prepare metadata and make their big data produce big science and lead to new discoveries.**

Keywords- Mercury Metadata Search Tool, OME, Metadata, FGDC, NGEE Arctic.

I. INTRODUCTION

To optimize data use and application, they should be accompanied by metadata describing their origin, data quality, provenance, syntax, and semantics. Observation data may not be reproducible, therefore it is imperative that initial data be captured and stored correctly. In general terms, metadata represents the details of who, what, when, where, why, and how of the corresponding dataset [1]. eXtensible Markup Language (XML) is the preferred file format for metadata, as it makes it portable and, more importantly, it makes it easy for system discoverability. The newly developed NGEE Arctic Online Metadata Editor (OME) is a Web-based tool that allows researchers and data providers to create and maintain XML files containing key information, or metadata, about their research. Metadata includes information about the specific projects, parameters, time periods, and locations associated with the data. Such information helps put the NGEE research output in context for credible scientific use. In addition, the metadata produced using OME allows other researchers to find these data via metadata clearinghouses like the Mercury Search System [2].

OME is part of ORNL's Mercury software fleet [2]. It was jointly developed to support projects funded by the United States Geological Survey (USGS), U.S. Department of Energy (DOE), National Aeronautics and Space Administration (NASA), and National Oceanic and Atmospheric Administration (NOAA). OME's architecture provides a customizable interface to support project specific requirements. Using this new architecture, the ORNL team developed multiple OME instances including ones for DOE's NGEE Arctic project and Atmospheric Radiation Measurement (ARM) program and for USGS's Core Science Analytics, Synthesis, and Libraries (CSAS&L).

Researchers simply use the ORNL Metadata Editor to enter relevant metadata into a Web-based form. From the information on the form, OME can create an XML file on the server hosting the editor or to the user's personal computer. Researchers can also use the ORNL Metadata Editor to modify existing XML metadata files. OME allows big data centers to produce meaningful, high quality, standards-based, descriptive information about their data products thus helping with data and scientific discovery.

II. IMPLEMENTATION

OME is based on a flexible, service-oriented architecture using Java Spring technology and XMLBeans [3][4]. Using OME, users can perform three tasks—Edit existing metadata in two ways or Create new metadata. To Edit existing metadata, users can (i) load a previously created metadata record from the server running OME by selecting from an alphabetical drop-down list containing all metadata records for that the user (ii) "Load" an existing XML metadata file from their local computer (Figure 1).

978-1-4799-9926-2/15/$31.00 ©2015 IEEE

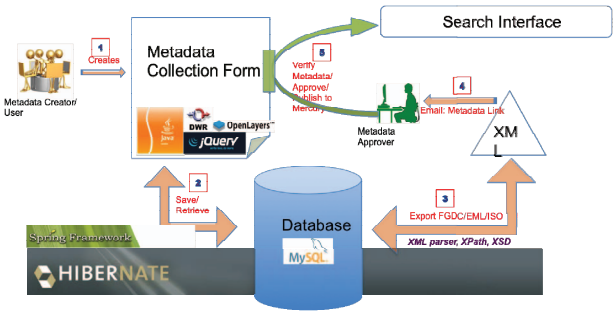

Figure 1. Figure showing the complete workflow to "create" a metadata record within the Online Metadata Editor

To edit an existing metadata record by loading an existing XML file on your local machine, a metadata record by the XML file must be in the Content Standard for Digital Geospatial Metadata (CSDGM) standard format. Lastly, Users can simply (iii) "Create" new Metadata. By choosing this option, a new form is presented to the user (Figure 2). The navigation tabs on the left side represent the 7 major sections [1] in the CSDGM metadata standard.

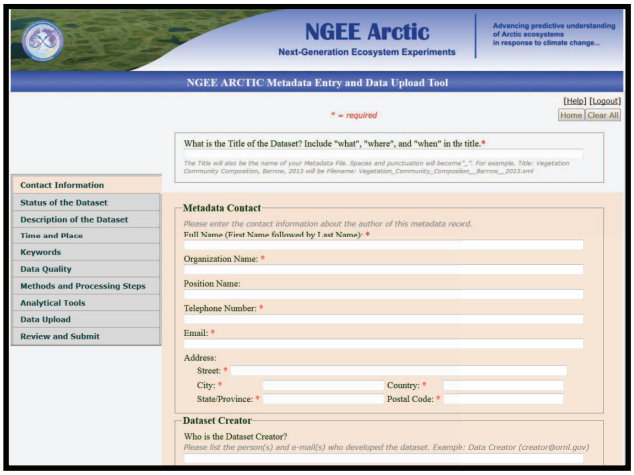

Figure 2. Online Metadata Editor (OME) form for the NGEE Arctic OME instance

Once the metadata records are successfully created and quality checked, they are then passed on to NGEE Arctic's Mercury search engine (Figure 3). The Mercury search engine harvests the metadata, builds a centralized virtual index, and makes them available from the Mercury search user interface. The centralized repository of metadata records with distributed data sources provides extremely fast search results to the user, while allowing NGEE Arctic data providers to advertise and distribute their data while maintaining complete control and ownership of that data.

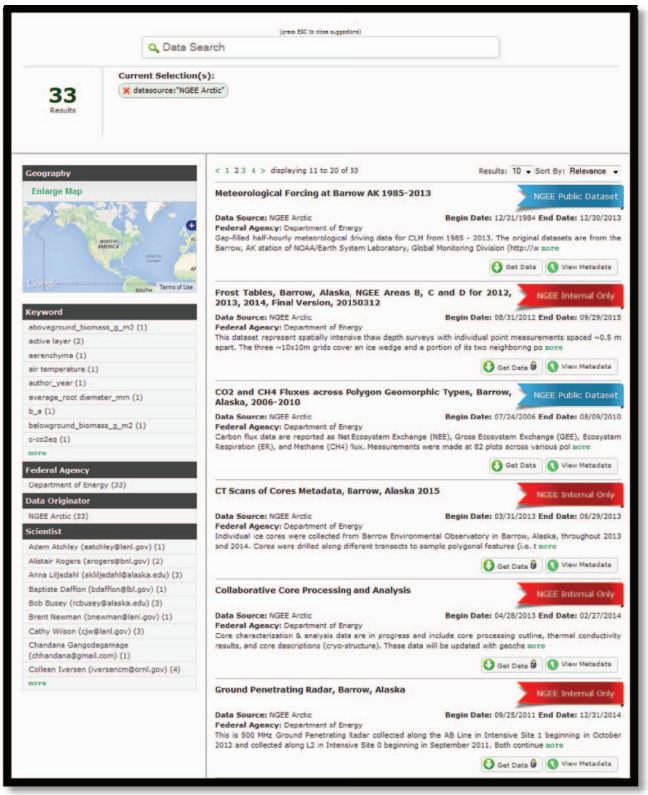

Figure 3. Illustration of the NGEE Arctic Mercury Search Tool

The search engine allows users to perform a variety of fielded, spatial, and temporal searches across these metadata sources. The new version 4.0 of the NGEE Arctic search engine provides marked improvements in search speed, support for additional metadata formats, and integration with OpenLayers and Google Maps APIs. Integration with these APIs enables improved spatial queries including polygon searches and semantic searches using an ontology-based framework. Other notable features of the new NGEE Arctic search engine include facetted- type search capabilities, support for RSS delivery of search results, and an enhanced User Interface.

III. CONCLUSION

In this paper, we discussed why it is important to have standardized, comprehensive metadata associated with data and how OME is helping large scientific projects, such as NGEE Arctic, produce and manage quality, standards-based metadata. We have discussed various options OME presents for researchers, data providers, and users to produce standards-based metadata. Also, how the OME tool provides a gateway for various data intensive projects at ORNL.

ACKNOWLEDGMENTS

The Next-Generation Ecosystem Experiments (NGEE Arctic) project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Oak Ridge National Laboratory is managed by the UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

REFERENCES

[1] Federal Geographic Data Committee. FGDC-STD-001-1998. Content standard for digital geospatial metadata (revised June 1998). Federal Data Committee. Washington, D.C.

[2] Devarakonda, Ranjeet, et al. "Mercury: reusable metadata management, data discovery and access system." Earth Science Informatics 3.1-2 (2010): 87-94.

[3] Spring MVC Form. Available at: http://spring.io/. Last accessed June 6 2014.

[4] Java Architecture for XML Binding. Available at: https://jaxb.java.net/. Last Accessed June 12 2014