

Presenter: Beth Huffer

AUTHORS:

E. Huffer, M.M. Little, J. Kusterer

Title: ODISEES: A New Paradigm in Data Access

Abstract:

As part of its ongoing efforts to improve access to data, the Atmospheric Science Data Center has developed a high-precision Earth Science domain ontology (the “ES Ontology”) implemented in a graph database (“the Semantic Metadata Repository”) that is used to store detailed, semantically-enhanced, parameter-level metadata for ASDC data products. The ES Ontology provides the semantic infrastructure needed to drive the ASDC’s Ontology-Driven Interactive Search Environment for Earth Science (“ODISEES”), a data discovery and access tool, and will support additional data services such as analytics and visualization.

The ES ontology is designed on the premise that naming conventions alone are not adequate to provide the information needed by prospective data consumers to assess the suitability of a given dataset for their research requirements; nor are current metadata conventions adequate to support seamless machine-to-machine interactions between file servers and end-user applications. Data consumers need information not only about what two data elements have in common, but also about how they are different. End-user applications need consistent, detailed metadata to support real-time data interoperability.

The ES ontology is a highly precise, bottom-up, queriable model of the Earth Science domain that focuses on critical details about the measurable phenomena, instrument techniques, data processing methods, and data file structures. Earth Science parameters are described in detail in the ES Ontology and mapped to the corresponding variables that occur in ASDC datasets. Variables are in turn mapped to well-annotated representations of the datasets that they occur in, the instrument(s) used to create them, the instrument platforms, the processing methods, etc., creating a linked-data structure that allows both human and machine users to access a wealth of information critical to understanding and manipulating the data. The mappings are recorded in the Semantic Metadata Repository as RDF-triples.

An off-the-shelf Ontology Development Environment and a custom Metadata Conversion Tool comprise a human-machine/machine-machine hybrid tool that partially automates the creation of metadata as RDF-triples by interfacing with existing metadata repositories and providing a user interface that solicits input from a human user, when needed. RDF-triples are pushed to the Ontology Development Environment, where a reasoning engine executes a series of inference rules whose antecedent conditions can be satisfied by the initial set of RDF-triples, thereby generating the additional detailed metadata that is missing in existing repositories.

A SPARQL Endpoint, a web-based query service and a Graphical User Interface allow prospective data consumers – even those with no familiarity with NASA data products - to search the metadata repository to find and order data products that meet their exact specifications. A web-based API will provide an interface for machine-to-machine transactions.