Leveraging OGC, FDGC, and CF Metadata to Populate a Discoverable Semantic Data Model Suitable for the Model Web

The Earth, Life, and Semantic Web (ELSEWeb) project consists of a semantically-enabled service oriented environment supporting Data-to-Model interactions that streamline the flow of heterogeneous geographic, social, and geological data into a set of modeling services, which generate predictions of species proliferation. The specific set of models targeted in ELSEWeb serve as part of the University of Kansas’ Lifemapper system, and each of these models ingest stacks of geospatial data known as “scenario layer sets” similarly to other Model Web applications such as eHabitat. In these modeling environments, scenario layer sets provide modeling algorithms with information about the environment from which a given species or habitat spread will be predicted. Prior to the numerous transformations (e.g., sub-setting, adjusting resolutions, changing projections, and converting formats) required to shape input data into forms suitable for Lifemapper ingestion, users must first identify relevant data which will comprise layer sets. This typically entails that users both identify input data sources such as DAACs or data warehouses as well as analyze the plethora of sources’ meta-data descriptions, to determine if the data is relevant for their modeling needs thereby slowing the process of submitting model generation requests.

The ELSEWeb infrastructure aims to alleviate this potential bottleneck by “wrapping” with published geospatial datasets with a uniform metadata description using the Web Ontology Language (OWL) that supports automated discovery and automated transformation planning with the Semantic Automated Discovery and Integration (SADI) framework, which also comprises the ELSEWeb infrastructure. Currently, ELSEWeb has translated 5700 specific meta-data desriptions employed at UNM’s Earth Data Analysis Center (EDAC) to generate semantic Linked Data that The linked data view of EDAC data provides information about: the entity being measured, the specific measurement, the sensor or method used to derive the measurement, the coverage (both temporal and spatial) and the Web reference to the dataset itself. To generate these semantic linked-data descriptions, ELSEWeb relies on a family of meta-data descriptions including: (1) OGC WCS getCapabilities which provides data endpoints, (2) FGDC, which provides information about the sensor or derivation method, (2) Climate Forecast (CF) conventions for describing both the kinds of measurement and the entity of interest; the CF labels are contained within the FGDC metadata. ELSEWeb demonstrates that this set of information is suitable for automated service orchestration clients (such as SHARE) to discover, transform, and serve EDAC linked data into Lifemapper models, provided declarative descriptions of what kinds of model users want to generate.

Currently, ELSEWeb linked-data registry encompasses 1700 EDAC services, that can be automatically discovered, integrated and transformed into layer sets suitable by Lifemapper, providing suers with a combinatoric data space from which to generate models.