



**Atmospheric
Science
Data Center**

The Atmospheric Science Data Center (ASDC) at NASA Langley Research Center is responsible for the ingest, archive, and distribution of NASA Earth Science data in the areas of radiation budget, clouds, aerosols, and tropospheric chemistry. The ASDC specializes in atmospheric data that is important to understanding the causes and processes of global climate change and the consequences of human activities on the climate. The ASDC currently supports more than 44 projects and has over 1,700 archived data sets, which increase daily. ASDC customers include scientists, researchers, federal, state, and local governments, academia, industry, and application users, the remote sensing community, and the general public.

The ASDC's strategic plan for fiscal year 2013 and beyond defines six goals that emphasize the vision and support the mission and values of the ASDC. The ASDC Semantic Infrastructure is central to realizing the first of these strategic goals: expanding the ASDC's customer base by improving access to ASDC data.

The ASDC will strive to expand beyond its existing customer base by increasing accessibility to a broader, worldwide market; through the use of innovative technologies, the ASDC will enhance data access capabilities and develop plans to share data with new user communities.

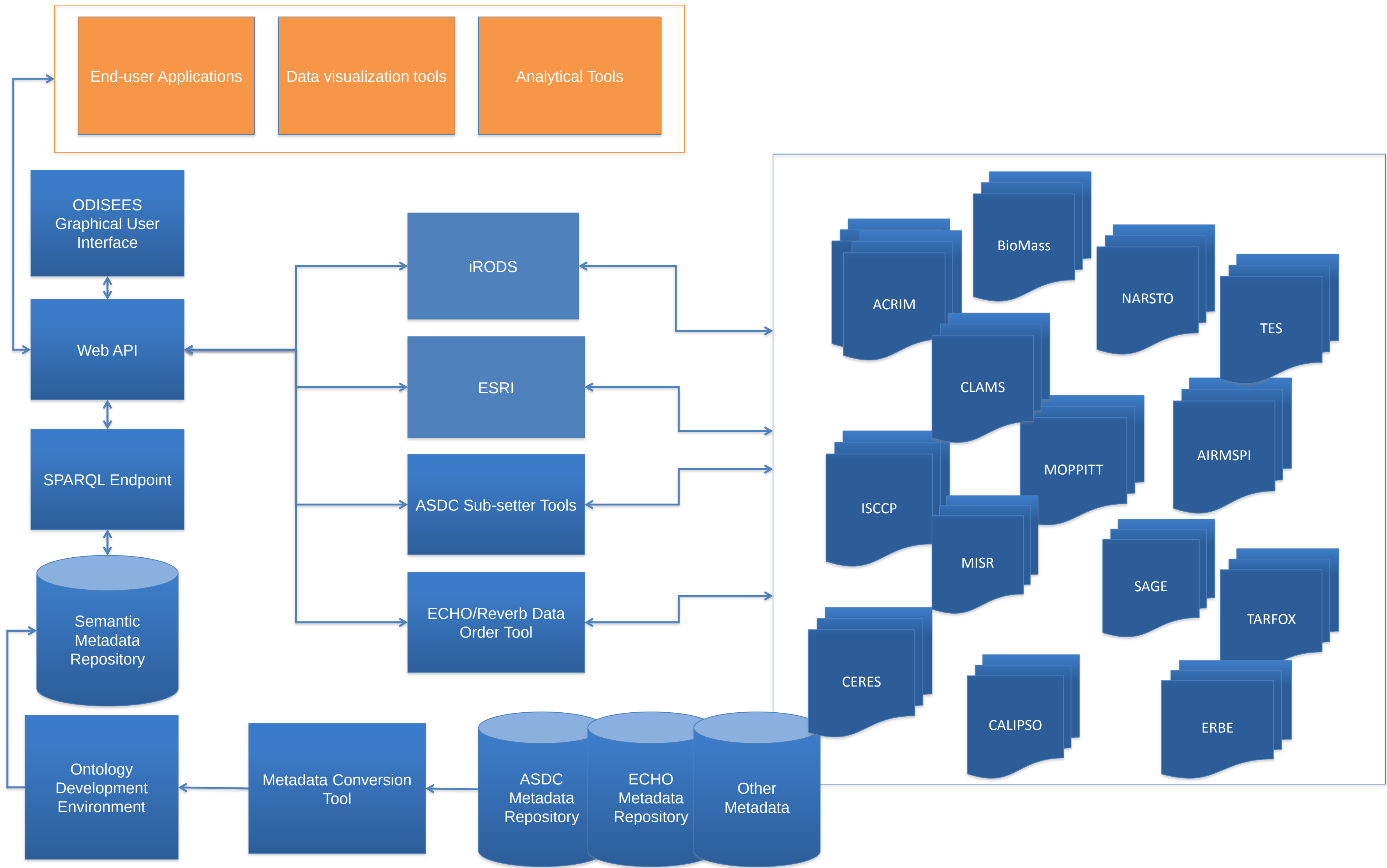
The corpus of ASDC data comprises a widely heterogeneous set of highly complex products. This presents two significant challenges to the ASDC in making its data accessible to and usable by a broad range of prospective customers: 1) How to help customers discover, among the multitude of available options, those data products that are most suitable for their particular purposes; and 2) How to help customers more easily use those products appropriately.

Distinct data products may differ significantly in terms of the circumstances under which the data were collected and processed, even when their subject matter is very similar. Understanding the differences is critical to using the data effectively. To make ASDC data accessible to a broad range consumers, it must provide prospective users with enough information to quickly and meaningfully compare and evaluate the data products offered.

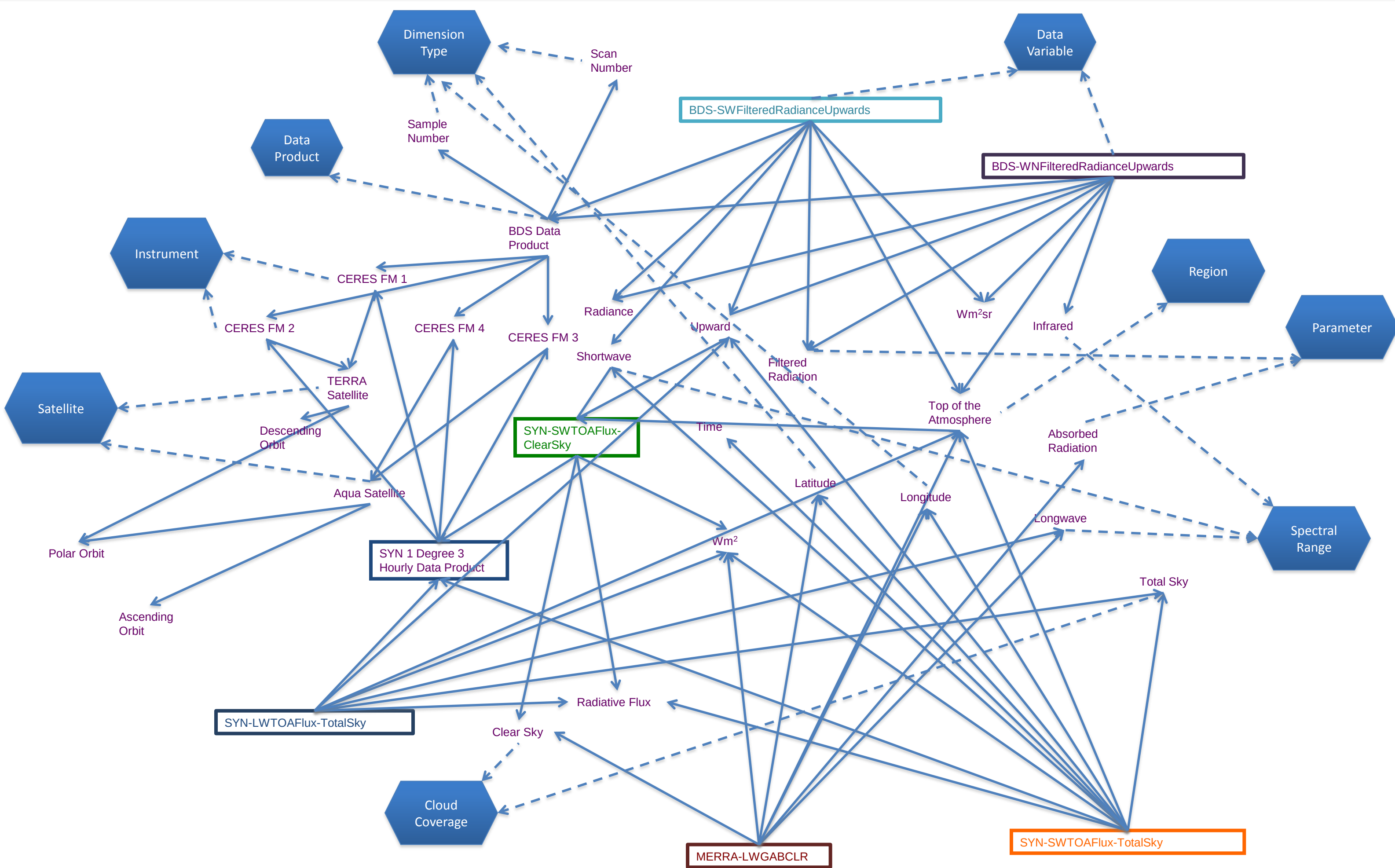
Data formats and structures can also differ significantly from one product to the next. Applications that display and analyze data need access to federated and semantically disambiguated data.

Semantic technologies – ontologies, triple stores, reasoners, linked data – offer functionality for addressing this issue. Ontologies can provide robust, high-fidelity domain models that serve as common schema for discovering, evaluating, comparing and integrating data from disparate products. Reasoning engines and triple stores can leverage ontologies to support intelligent search applications that allow users to discover, query, retrieve, and easily reformat data from a broad spectrum of sources.

The Atmospheric Science Data Center's Semantic Infrastructure will provide semantically enhanced metadata to support data exploration and discovery to enable prospective consumers to quickly and easily find the data products that are best suited for their requirements, and schema mapping to enable automated data integration



The ASDC Earth Science Ontology models relationships among data products, parameters and file structures. The ontology serves as a “lingua franca” to facilitate mappings among disparate data vocabularies and data models, illuminating both the similarities and the differences among available data products. These mappings facilitate data federation and integration, and intelligent search applications such as ODISEES (Ontology-Driven Interactive Search Environment for Earth Science).



ASDC Earth Science Ontology

The **Semantic Metadata Repository** is a complex component that consists of a graph database and the Earth Science ontology. The graph database is a commercial, off-the-shelf product called “AllegroGraph,” developed by Franz, Inc.

The SPARQL Endpoint enables queries against the ODISEES Metadata Repository using SPARQL (SPARQL Protocol and RDF Query Language), an established W3C standard. The SPARQL queries are generated by the Web API in response to user input and passed to the Metadata Repository via the SPARQL Endpoint.

The **ODISEES GUI** provides parameter-based search and query interfaces that can quickly and intuitively guide prospective data consumers to the NASA data resources that satisfy their exact criteria. The GUI features an interactive browser that allows users to specify the physical parameters they are interested in, and in response, displays a list of the variable/data set pairs that contain values for the specified parameters.

For each result, the GUI links users to additional information about the instruments, satellites, climate models, algorithms and other metadata relevant to the using and understanding that NASA product. From the ODISEES GUI, users can push a data order directly to various data access systems.

The Web API provides an interface between the SPARQL Endpoint, the ODISEES GUI, and the data access systems. This web service queries against the Semantic Metadata Repository, even when the user has no familiarity with the ontology structure or the vocabulary used therein. As users provide input via the ODISEES GUI, the Web API translates the input into well-formed SPARQL queries, passes the queries to the ODISEES SPARQL Endpoint, receives the query results, and pushes them out to the GUI in JSON or XML formats. The Web API can support a variety of machine-to-machine transactions with the Semantic Metadata Repository.

The Metadata Conversion Tool (MCT), along with **the Ontology Development Environment**, will convert metadata from existing sources into the OWL-based language of the ES Ontology. The MCT will import metadata from existing resources, convert it into an RDF-compatible format, and semantically enhance it by translating the native terminology of the original metadata to the terminology used by the ASDC Earth Science Ontology. Inference rules are used to generate additional metadata needed to support intelligent search and data integration.

- The ontology represents data products, instruments, satellites, science and the essential characteristics of the data to which values can be assigned.
- Domain objects are modeled in detail, including information about their potential relationships to other domain elements and to data variables and products.
- Data variables are represented as a set of relationships - expressed as RDF triples - that together describe the variable in a machine-readable format that provides sufficient detail to enable scientists to assess, with reasonable precision, the nature of the information encoded by such variable.
- Semantic mappings are based on shared characteristics, rather than similarity of names. Two objects are equivalent if they share an appropriate set of essential characteristics. Essential characteristics can be determined by the individual data user.
- The model supports data discovery and integration, and can provide summaries of information to help users determine if a given data product satisfies his or her requirements.
- The model supports schema mapping to enable data federation and integration. Users can easily determine how disparate data products differ in model and format and how they can be rendered compatible.