Geophysical Research Abstracts
Vol. 20, EGU2018-11792, 2018
EGU General Assembly 2018

DataFed: A Federated Data System for Spatio-Temporal Data

Rudolf Husar
Washington University, Energy Environment and Chemical Engineering, St. Louis, United States (rudolf.husar@gmail.com)

DataFed is a distributed web-services-based computing environment for accessing, processing, and visualizing atmospheric data in support of environmental science and management. The data system facilitates the access and flow of spatio-temporal data from provider to users. 'Wrapper' components, non-intrusively wrap heterogeneous, distributed datasets such that they are accessible by standards-based GIS web services: OGC WCS and WMS services. The mediator components (also web services) map the heterogeneous data into a universal spatio-temporal data model. The processing of raw data is performed by reusable web-services for filtering, aggregation, fusion and other data processing services. Complete application programs are created by chaining the services using a data flow language. Chained web services are used to provide homogeneous data views (e.g., geospatial, time views) using a global multi-dimensional data model. The web browser-based user interface allows point and click navigation and browsing the XYZT data space. Currently, the key applications of DataFed are for exploring and analyzing spatial pattern of atmospheric pollutants, seasonal, weekly, diurnal cycles and frequency distributions for exploratory air quality research. Currently, the federated data pool consists of about 30 datasets originating from globally distributed data providers offering surface-based air quality data, satellite observations, emissions data as well as regional and global-scale air quality models. Since 2008, DataFed has been used to support the US Federal EPA and US States in the implementation of air quality standards. The data system was developed at Washington University in St. Louis and also used in research at universities in the US and Asia.