



## ODM2 (Observation Data Model): The EarthChem Use Case

Kerstin Lehnert (1), Lulin Song (1), Leslie Hsu (1), Jeffrey S Horsburgh (2), Anthony K Aufdenkampe (3), Emilio Mayorga (4), David Tarboton (2), and Ilya Zaslavsky (5)

(1) Columbia University, Lamont-Doherty Earth Observatory, Palisades, United States (lehnert@ldeo.columbia.edu), (2) Civil and Environmental Engineering, Utah State University, Logan, UT, United States, (3) Stroud Water Research Center, Avondale, PA, United States, (4) Applied Physics Laboratory, University of Washington, Seattle, WA, United States, (5) San Diego Supercomputer Center, University of California San Diego, San Diego, CA, United States

PetDB is an online data system that was created in the late 1990's to serve online a synthesis of published geochemical and petrological data of igneous and metamorphic rocks. PetDB has today reached a volume of 2.5 million analytical values for nearly 70,000 rock samples. PetDB's data model (Lehnert et al., G-Cubed 2000) was designed to store sample-based observational data generated by the analysis of rocks, together with a wide range of metadata documenting provenance of the samples, analytical procedures, data quality, and data source. Attempts to store additional types of geochemical data such as time-series data of seafloor hydrothermal springs and volcanic gases, depth-series data for marine sediments and soils, and mineral or mineral inclusion data revealed the limitations of the schema: the inability to properly record sample hierarchies (for example, a garnet that is included in a diamond that is included in a xenolith that is included in a kimberlite rock sample), inability to properly store time-series data, inability to accommodate classification schemes other than rock lithologies, deficiencies of identifying and documenting datasets that are not part of publications.

In order to overcome these deficiencies, PetDB has been developing a new data schema using the ODM2 information model (ODM=Observation Data Model). The development of ODM2 is a collaborative project that leverages the experience of several existing information representations, including PetDB and EarthChem, and the CUAHSI HIS Observations Data Model (ODM), as well as the general specification for encoding observational data called Observations and Measurements (O&M) to develop a uniform information model that seamlessly manages spatially discrete, feature-based earth observations from environmental samples and sample fractions as well as in-situ sensors, and to test its initial implementation in a variety of user scenarios. The O&M model, adopted as an international standard by the Open Geospatial Consortium, and later by ISO, is the foundation of several domain markup languages such as OGC WaterML 2, used for exchanging hydrologic time series. O&M profiles for samples and sample fractions have not been standardized yet, and there is a significant variety in sample data representations used across agencies and academic projects. The intent of the ODM2 project is to create a unified relational representation for different types of spatially discrete observational data, ensuring that the data can be efficiently stored, transferred, catalogued and queried within a variety of earth science applications. We will report on the initial design and implementation of the new model for PetDB, and results of testing the model against a set of common queries. We have explored several aspects of the model, including: semantic consistency, validation and integrity checking, portability and maintainability, query efficiency, and scalability. The sample datasets from PetDB have been loaded in the initial physical implementation for testing. The results of the experiments point to both benefits and challenges of the initial design, and illustrate the key trade-off between the generality of design, ease of interpretation, and query efficiency, especially as the system needs to scale to millions of records.