

EARTH OBSERVATION SYMPOSIUM (B1)  
Earth Observation Data Management Systems (4)

Author: Mr. Curt Tilmes

National Aeronautics and Space Administration (NASA), Goddard Space Flight Center, United States,  
Curt.Tilmes@nasa.gov

Dr. Albert Fleig

PITA Analytic Sciences, United States, Albert.J.Fleig@nasa.gov

DATA MANAGEMENT SYSTEMS FOR WELL DEFINED AND REPEATABLE DATA SETS:  
PROVIDING THE INFORMATION NEEDED TO GO FROM IMAGINATION TO REALITY

**Abstract**

There are both scientific and legal reasons for providing the information needed to be able to show exactly how a data set was made and to be able to repeat its production with reasonable expectation of getting the same answer. Current Earth Observation Data Management Systems often do not support or even allow provision of all the information necessary to do this. One science impact of the lack of this information is that future climate scientists will find it difficult or impossible to tell with certainty how much of the difference between a current and a future data set is related to a true geophysical change and how much originates in changes of processing algorithms or data processing system. A legal impact also might result from being unable to explain exactly how a data set was made and what the input data and algorithms used in its production were. We will discuss some of the major elements of the needed information that are not present in most current data management systems and describe a system and operational procedures that automates collection of the information necessary to answer these questions and provides tools for accessing and using this information. This Atmospheric Composition Processing system is being developed to support measurements from the Ozone Mapping and Profiling Suite, OMPS, to be flown on the next generation of U.S. weather satellites and is evolving from components used for processing and reprocessing ozone data from the TOMS and OMI instruments flown from 1978 through the present. The system automatically captures full provenance for all of the elements used in the data set production including algorithm version and source code, input data set version and source, and hardware and processing system software configuration. We will discuss the operational impact of this approach including its impact on performance, storage and operational cost. This is an ongoing development and we will be presenting first results of our system as well as the programmatic implications of maintaining this information on scientists, funding agencies and archival institutions.