Atmospheric Composition Variable Standard Name Convention

Suggested Citation

Silverman, M., and G. Chen, 2022. Atmospheric Composition Variable Standard Name Convention. NASA Earth Science Data and Information System Standards Coordination Office. https://doi.org/10.5067/DOC/ESCO/ESDS-RFC-043v1

Status of this Memo

This RFC document describes the standard for atmospheric composition standard names, including the structure and components of the standard name and instructions on how to generate a standard name.

Distribution of this memo is unlimited.

Change Explanation

This RFC does not update or change a previous RFC.

Copyright Notice

Abstract

The development of this convention for atmospheric composition variable standard names was motivated by the need to use the ICARTT file format standards V2.0 for data reporting from suborbital field studies. ICARTT V2.0 requires a standard variable name entry as part of the variable definition, which was introduced to mitigate challenges in dealing with an ever-increasing large variety of variables from the NASA suborbital field campaigns. Through test implementations in several suborbital field studies, the standard name has demonstrated its role in supporting findable, accessible, interoperable, and reusable (FAIR) principles for data management and stewardship.

The number of data product variables reported during suborbital field campaigns has increased more than tenfold over the last forty years. Typically, the only requirement for variable names has been that they be unique within a data provider's data file. With the rise in the number of variables and lack of constraints in how they are named, the complexity for data users to search and for distributed active archive centers (DAACs) to archive and distribute the data has increased. Examination of the current naming systems (Network Common Data Form Climate and Forecast (CF) Metadata Conventions, and Geoscience Standard Names Ontology (GSN) now the Scientific Variables Ontology (SVO)) has highlighted several limitations. First, only a quarter to a third of atmospheric composition variables have standard names, with the largest missing category being hydrocarbon measurements. Second, some of the vocabulary and definitions are not suitable for atmospheric composition variables. Third, optional qualifiers limit

a consistent structure, which poses problems for discoverability and interoperability. Furthermore, there is no established agile process to address these limitations for NASA suborbital field campaigns.

Due to these limitations, and additional concerns that the variable names sometimes only provide a partial perspective about the measurement or intended physical phenomenon without context in the subject field, a new standard naming system for atmospheric composition variables has been developed. These atmospheric composition variable standard names are designed to unambiguously identify each suborbital field study variable and to enhance data usability, interoperability, and discoverability. These standard names are constructed using four required components: measurement category (MeasurementCategory), core name (CoreName), acquisition method (AcquisitionMethod), and descriptive attributes (DescriptiveAttributes), which are separated by underscores; this consistent structure allows the standard names to be readily deconstructed. Each of the four required components are governed by a list of controlled vocabulary. This structure is similar to the format used for the standard names in the Climate and Forecast Metadata Conventions (CF) and the Geoscience Standard Names (GSN) ontology. The MeasurementCategory and CoreName components provide the basic identification of the measurand and can be used to conduct a broad search to identify all data product variables of the same physical quantity or shared property from different instruments and/or field studies. The AcquisitionMethod identifies the sampling geometry that was used for the measurement, while the DescriptiveAttributes are intended to provide necessary description to support research use or narrow down the search for data of interest. The use of a controlled vocabulary and predictable and decomposable structure is to ensure data usability, machine-actionability, and interoperability.

This document describes the need for standard variable names as well as instructions on how to construct an atmospheric composition variable standard name using a controlled structure and vocabulary for different types of data produced during atmospheric composition suborbital field studies. The lists of controlled vocabulary comprising the standard names are intended to support and improve the data reporting and archiving from NASA suborbital field studies, and it is expected that they will need to be updated in response to advances in atmospheric composition measurements. The Practical Guide for Atmospheric Composition Variable Standard Names and Controlled Vocabulary document is currently located at the Suborbital Science Data for Atmospheric Composition data repository, https://www-air.larc.nasa.gov/missions/etc/AtmosphericCompositionVariableStandardNames.pdf [1]. This document will be maintained by the ASDC DAAC in conjunction with field campaign data management teams. Along with supporting the use of ICARTT 2.0 file format standard, the standard names can be used in broader applications, such as variable attributes in netCDF/HDF formats.

Table of Contents

1 Introduction and Motivation

Ideally, a variable name readily identifies the physical phenomenon that is the subject of a set of measurements, or some other data set, and serves as a link to its mathematical representation, e.g., an equation or a system of equations. It is common for instrument scientists to use their intended measurable quantity as the data variable name, many of which later propagate to the rest of the scientific community via journal publications and textbooks. Datasets (i.e., collections of variables) gathered within the field of atmospheric science are used in a variety of ways, including computational modeling, interpretive analysis, measurement intercomparisons, and validation studies. To enable the use of datasets from different sources (i.e., data products from different measurement assets and/or models), like merge data products, in these various research activities, the data needs to be findable, accessible, interoperable, and reusable (FAIR) [1] across multiple computational resources.

Since the 1980s, over 30 major tropospheric airborne field campaigns have been conducted by NASA and partner agencies to investigate atmospheric composition. The number of variables measured during this time has increased more than tenfold. In 1992, the Transport and Atmospheric Chemistry near the Equator-Atlantic (TRACE-A) campaign measured approximately 50 variables, in comparison to 2019 where the Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ) campaign had over 600 variables reported. Hydrocarbon compounds and aerosol properties make up the largest component of these new variables. The large number of measured variables and use of various variable names for the same measured quantity (especially when considering multiple field campaigns) increases the complexity for distributed active archive centers (DAACs) to distribute the data using NASA's common metadata repository (CMR) [2] through EarthdataSearch [3] and for data users to search for and find their data products of interest.

Variable names are given by the instrument scientists during suborbital atmospheric composition field studies. Since there have not been many guidelines related to variable names, and the only requirement has been that they are unique within a data provider's data file, variable names and structures can vary significantly, even for the same type of variable. For example, one scientist may report chloroform as "CHCl3" while another scientist reports it as "chloroform". Some instrument scientists also add additional information to their variable names to indicate units and/or instruments (e.g., H2O_ppmv for water vapor mixing ratio reported in parts per million by volume and H2O_DLH for the diode laser hygrometer (DLH) measurement of water vapor mixing ratio). This practice can satisfy variable name uniqueness, at least within a field campaign, which is necessary especially when there is more than one instrument to measure the same physical quantity. It is not uncommon to have multiple instruments measure mission critical variables on a single sampling platform and in a single field study. Even the same instrument scientist may choose a different variable name for the same measurement during a different field campaign (Figure 1).

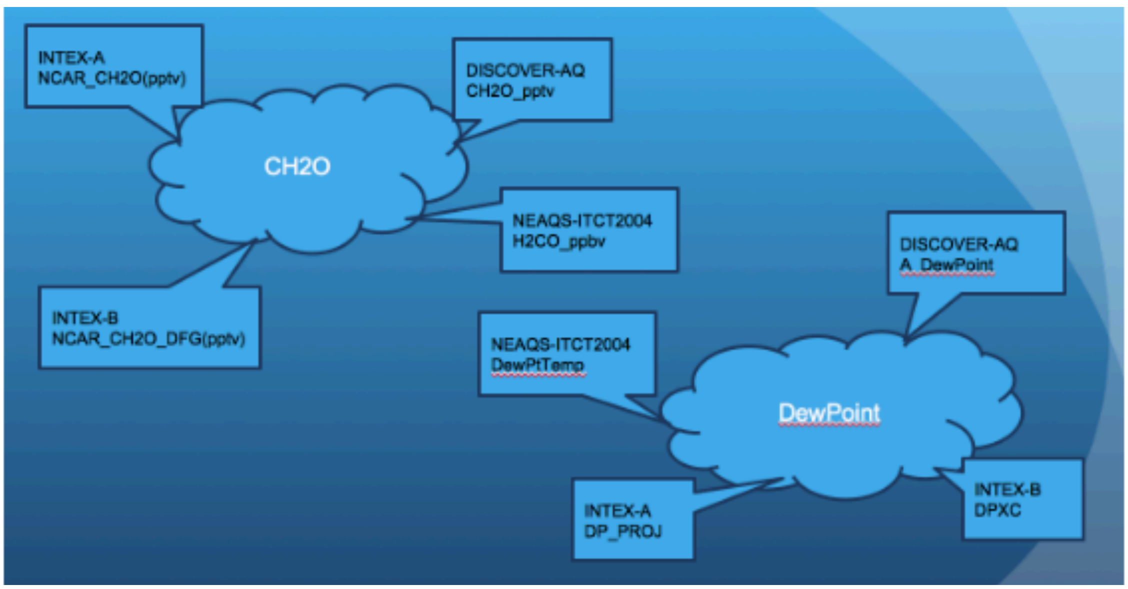

Figure 1: Example of formaldehyde (CH_2O) and dew point variable names across multiple campaigns. In each box, the first line is the campaign name and the second is an example of an investigator-given variable name.

While it is important to have uniqueness within a dataset from one study, particularly when data are reported in ASCII format (e.g., International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) format), the variable name differences make it difficult for users to locate and interact with a particular variable across multiple data sets. This is especially true for data users who are not familiar with the field missions and who would, therefore, not know that they need to search for "A_DewPoint" in one field campaign and "DPXC" in another to find data on dew point, for example. One effective solution to this problem, identified by the Earth Science Data System (ESDS) ICARTT Refresh Working Group, is to introduce variable standard names that can be used as tags for each data variable [4]. This will allow similar measurements (e.g., dew point) to be categorized and located across field campaigns, regardless of what the instrument scientists gave as the variable name. Based on the successful implementation of using standard names in the now-defunct NASA Toolsets for Airborne Data (TAD) web application, the use of standard variable names was adopted in the NASA suborbital data format standard ICARTT V2.0 (https://cdn.earthdata.nasa.gov/conduit/upload/6158/ESDS-RFC-029v2.pdf) [5]. These standard names are intended to enable data discovery (i.e., findable and accessible) and support data ingest across different studies (i.e., interoperable and reusable). By providing a common terminology, all users (instrument scientists, science team members, researchers, students, and DAACs) will be able to find, distribute, and use atmospheric composition data in an efficient manner. These standard names will be used to provide support for discoverability within the Sub-Orbital Order Tool (SOOT) [6].

The Atmospheric Composition Variable Standard Names Convention has been successfully implemented in and accepted by science team members (from instrument scientists to modelers) in FIREX-AQ, the Cloud, Aerosol and Monsoon Processes – Philippines Experiment ($CAMP^2EX$), the Aerosol Cloud Meteorology Interactions over the Western Atlantic Experiment (ACTIVATE), and the Dynamics and Chemistry of the Summer Stratosphere (DCOTSS) field campaigns. While the current naming convention has been used for these campaigns, over one hundred new data product variables have been added to the list of controlled vocabulary in the Practical Guide for Atmospheric Composition Variable Standard Names and Controlled Vocabulary [7] during this time, reinforcing the idea that the controlled vocabulary must be a living document.

The standard names allow data files to be readily grouped based on their physical and/or chemical properties and assigned to their respective data product and/or collections. This allows a DAAC to ingest and archive the data collected more efficiently, along with distributing it to the public. The use of standard names has helped the Atmospheric Science Data Center (ASDC) streamline their data ingest and archival process by saving time assigning data to its respective data product, or collection.

1.1 Known Naming Conventions

A few different variable standard naming systems exist currently, but they do not adequately cover the needs of atmospheric composition data products variables, and specifically the needs of the NASA suborbital atmospheric science field studies (i.e., field campaigns involving aircraft, ground sites, ships, and mobile labs), which often include several newly-developed research grade measurements.

One of the approved standards recommended for use in NASA Earth Science Data Systems is the Network Common Data Form (netCDF) Climate and Forecast (CF) Metadata Conventions [8]. The CF "standard is intended for use with climate and forecast data, for atmosphere, surface and ocean, and designed with model-generated data particularly in mind" [8]. While many principles of the CF convention can be extended beyond their originally intended uses, it has been increasingly recognized by various communities that the CF standard name and its structure have three main shortcomings when handling atmospheric composition variables involved in suborbital field studies: measurand coverage and description; structure and vocabulary; and usability. CF standard names only cover a quarter to a third of the atmospheric composition relevant variables submitted in suborbital field campaigns. Since the CF standard names are designed for model and forecast data, much of the specific vocabulary and terms used are not suitable for describing atmospheric composition data products, leading to insufficient or inaccurate information to support research use. The construction of CF names is governed by a set of guidelines that allows for the addition of qualifiers to a base standard name using underscores. However, the open-ended nature of this functionality allows a range of possibilities for CF standard names that poses a problem for discoverability and interoperability. Lastly, the usability of CF is difficult for atmospheric composition variables; some CF standard names are ambiguous and could refer to several different variables, whereas for other variables, more than one CF standard name could be used. More detail on the challenges of using CF for suborbital campaigns can be found in Appendix B.

Another current list of standard names is the Geoscience Standard Names Ontology (GSN) now known as the Scientific Variables Ontology (SVO) [9], which is based on and extended from the Community Surface Dynamics Modeling System (CSDMS) [10] standard names. The GSN standard names are a set of variable names using a series of rules and controlled vocabulary designed to avoid ambiguous variable names and domain-specific terminology, use generic or already-standardized object names, are human and machine readable, and standardized. The standard names are constructed through pairings of object names and quantity names (identification of measurement concept used to quantify the object in some way) and uses CF names for atmospheric chemistry (a key part of atmospheric composition studies). However, due to the reliance on CF names, which as mentioned previously do not fully capture the range of data product variables currently reported in atmospheric field campaigns nor the attributes necessary to describe measurement sampling and data reporting, these standard naming schemes are not suitable for suborbital atmospheric composition field studies.

Since the beginning of development of the Atmospheric Composition Variable Standard Names, a new convention called the InteroperAble Descriptions of Observable Property Terminology (I-ADOPT) Framework Ontology has been developed. The goal of this framework is to "facilitate interoperability between existing variable description models" [11]. The framework includes six key recommendations for the description of observable properties and/or variables [11]. Out of the six recommendations, the Atmospheric Composition Variable Standard Names currently follow the first four recommendations, which emphasize both human and machine readability, data usability, decomposition approach, and use of controlled vocabulary. The remaining two recommendations relate specifically to alignment with the I-ADOPT Framework. One recent implementation of this is the Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS) vocabulary [12]. ACTRIS is highly compatible, with only minor ontological differences, to the Atmospheric Composition Variable Standard Names. The main difference is that the newly released ACTRIS vocabulary is mostly tailored to surface network observations with routine measurements, while most NASA suborbital field studies involve research grade measurements, which often leads to new data variables and need an agile process to provide adequate coverage.

1.2 Additional Requirements of Standard Names

Along with the limitations described above, there are two additional important aspects that the atmospheric composition variable standard names need to address.

First, atmospheric composition variables sometimes can only provide a partial perspective of the intended physical phenomenon or quantity. For example, the number size distribution of aerosol particles is used in the assessment of aerosol particle microphysical and optical properties. In theoretical description, aerosol particle size is defined as the geometric size, but this cannot be readily measured by instruments commonly used in the field studies. What is typically reported in datasets are optical size (based on particle scattering properties), mobility size, or aerodynamic size. These measurements are related to, but different from the geometric size, and this difference can have a substantial impact depending on the measurement and application. Thus, the standard names need to contain attributes that provide information relative to the measurements to ensure data usability.

Second, some variable names commonly used in literature are unlikely to be understood by users outside of the subject area by simply reading the variable name. However, these terms can be used by users to get a precise definition and to learn about the subject area and the measurement, since they are widely used by the research community (e.g., in journal publications). A goal of the atmospheric composition variable standard names is to be accessible to and provide context for all users, while still remaining connected to those in the subject area. Therefore, it's essential to use terminology commonly used in the atmospheric science community.

For example, the terms SSA, CCN, and NOy do not convey a clear message about the physical quantities and processes they represent without additional definitions:

- SSA: single scattering albedo, a ratio of scattering to extinction, of aerosol particles within an observed size range and at a particular wavelength
- CCN: a proxy of concentration of cloud condensation nuclei measured at set instrument supersaturation levels
- NOy: a proxy of the total reactive nitrogen species, which is a lumped measurement of reactive nitrogen species including NO, NO2, HNO3, PAN, HONO, NO3, N2O5, organic nitrates, and possible particulate nitrates

To address the limitations in the current set of standards available and the additional concerns described above, a new system of standard names has been formed. The new system of standard names attempts to strike a good working compromise between completeness (or explicitness) and generality. The following section provides a set of instructions for the construction of atmospheric composition variable standard names for different types of data product variables conducted during suborbital field studies of atmospheric composition. It is the intent that these lists of controlled vocabulary [7] will be used to support and improve the data reporting, usability, and interoperability from NASA suborbital field studies by fulfilling the variable standard name requirement, in which each data product variable is tagged with a standard name, as required in the ICARTT v2.0 file format standard. It is noted that the application of the atmospheric composition variable standard names is not limited to the ICARTT V2.0 data format. For example, they can also be used as a variable attribute in netCDF or HDF files.

2 Atmospheric Composition Variable Standard Name Format

The atmospheric composition variable standard names are constructed using four required components, each governed by controlled vocabulary: measurement category (MeasurementCategory), core name (CoreName), acquisition method (AcquisitionMethod), and descriptive attributes (DescriptiveAttributes), with each component separated by an underscore:

Atmospheric Composition Variable Standard Name =
MeasurementCategory_CoreName_AcquisitionMethod_DescriptiveAttributes

This format is designed to enable FAIR principles through data discovery, distribution, interoperability and use (or re-use), by accurately describing all variables from different sources while using a consistent and predictable format. For data discovery (findability), the MeasurementCategory and CoreName (i.e., CoreName themselves and their respective descriptions) can be used to conduct a broad search to identify all data product variables of the same physical quantity from different instruments and/or field studies. The AcquisitionMethod then identifies the sampling geometry used for the measurement, enhancing the data's reusability, while the DescriptiveAttributes can be used to narrow down the search for data of interest. The descriptive attributes provide necessary description to support research use while also enhancing the data's reusability (or usability). The standard name is meant to be interpreted as a whole for data usability. For each of these components, a list of controlled vocabulary exists to maintain interoperability [7].

2.1 MeasurementCategory

The MeasurementCategory broadly groups all standard names into one of thirteen categories (Gas, AerComp, AerMP, AerOpt, CldComp, CldMicro, CldMacro, CldOpt, Met, GasJValue, AquJValue, Platform, and Rad) based on shared properties and/or types of measurements. See Table 1 for a full description of each MeasurementCategory. Each measurement category encompasses either objects or properties of objects, in the case of aerosol and cloud particles. By dividing variables into thirteen categories, the standard name is able to represent each group with a clearly defined set of descriptive attributes and vocabularies. The types and/or number of attributes within each category are constant and have been tailored to each type of measurement (e.g., aerosol particle optical property vs. aerosol particle composition) or medium (e.g., trace gas vs. aerosol particle); therefore, MeasurementCategory is defined by the measurement medium and type of measurements. The format for every standard name within a MeasurementCategory is consistent (i.e., they have the same number and type of descriptive attributes). For example, all 'Gas' standard names follow the format, Gas_CoreName_AcquisitionMethod_MeasurementSpecificity_Reporting, whereas all meteorology standard names have the following format, Met_CoreName_AcquisitionMethod_None. Additionally, the MeasurementCategory provides uniqueness as using only CoreNames could be ambiguous (e.g., a particle number concentration could be describing cloud or aerosol particles).

Table 1: Measurement Category Descriptions and Associated Attributes

MeasurementCategory	Description	Number of Descriptive Attributes	DescriptiveAttributes
Gas	Abundance, relative abundance, or properties of specific trace gas compounds or a group of trace gases measured or reported as one lumped quantity	2	MeasurementSpecificity, Reporting
AerMP (Aerosol Particle Microphysical Properties)	Aerosol microphysical properties of particles not segregated by chemical composition, e.g., abundance, relative abundance, size, and size distribution	4	MeasurementRH, SizingTechnique, SizeRange, Reporting
AerComp (Aerosol Particle Composition)	Aerosol chemical (or composition) properties (including as a function of particle size), i.e., abundance or relative abundance of the chemical components, bulk chemical characteristics, and mixing state	3	SizingTechnique, SizeRange, Reporting
AerOpt (Aerosol Optical Properties)	Intensive and extensive optical properties and optical hygroscopicity of all sampled aerosol particles or a subgroup of aerosol particles	4	MeasurementRH, WL, SizeRange, Reporting
CldComp (Cloud Particle Composition)	Cloud particle chemical composition, ratio of compositions, and chemical characteristics	3	SizingTechnique, SizeRange, Reporting
CldMicro (Cloud Microphysical Properties)	Cloud particle abundance, size, and size	3	SizingTechnique, SizeRange, Reporting

	distribution as well as phase information		
CldMacro (Cloud Macrophysical Properties)	Cloud macrophysical properties, e.g., spatial coverage	0*	None
CldOpt (Cloud Optical Properties)	Intensive and extensive optical properties of cloud particles	1	WL
Met	Meteorology parameters	0*	None
GasJValue	Gas phase photolytic rate coefficients	3	MeasurementDirection, SpectralCoverage, Products
AquJValue	Aqueous phase photolytic rate coefficients	3	MeasurementDirection, SpectralCoverage, Products
Platform	Measurement platform (e.g., aircraft, ship, motor vehicles) position, attitude, and navigational parameters	0*	
Rad (Radiation Measurements)	Radiance, irradiance, and actinic flux measurements	1	WLMode

* While no descriptive attributes exist for these measurement categories, '_None' must be used in place of the DescriptiveAttribute.

2.2 CoreName

The CoreName is a component of the standard name that provides the unambiguous identification of the measurand. Depending on the MeasurementCategory, the reported data product variable and representative CoreName can be an identifier of a trace gas or related property, an aerosol property, cloud property, radiation flux, meteorological parameter, or a sampling platform location and attitude. In some cases, the CoreNames chosen are those that have been commonly used in literature, such as NOy, whereas other times they're an abbreviated version of the physical quantity, such as AOD (Aerosol Optical Depth). As described previously, the terms commonly used in literature are sometimes unlikely to be understood by users outside of the subject area. To enhance the usability, a brief description is provided along with each CoreName. This description provides additional information for when the CoreName chosen isn't explicit in what it means. See Table 2 for examples of CoreNames and description.

For trace gas variables, the names of specific species are a combination of chemical formulas and chemical names. The chemical names used for volatile organic carbon species follow a standard chemical nomenclature, which has been agreed upon by multiple measurement groups. In addition, these names are linked, when applicable, to Chemical Abstracts Service (CAS) numbers, which are unique for each chemical compound. Each CoreName also has a corresponding MeasurementSpecificity (S, M, or NA) to define whether the CoreName represents a specific chemical compound or group of compounds as some instruments do not have sufficient selectivity to determine individual trace gas species. These data are reported as the sum of multiple species or group of species. For these lumped variables, the CoreNames are either those used in literature (e.g., NOy, PNs) or a combination of names for specific compounds (e.g., iButeneAnd1Butene for the sum of Isobutene and 1-Butene) included.

Table 2: Example Trace Gas CoreNames and Description

CoreName	**Description**	**Chemical Formula**	**CAS Number**	**Specificity**
H2	Hydrogen	H_2	1333-74-0	S
O2	Oxygen	O_2	7782-44-7	S
EthONO2	Ethyl nitrate	$C_2H_5NO_3$	625-58-1	S
nPropONO2	n-Propyl nitrate	$C_3H_7NO_3$	627-13-4	S
iPropONO2	Isopropyl nitrate	$C_3H_7NO_3$	1712-64-7	S
x1Butene	1-Butene	C_4H_8	106-98-9	S
iButeneAnd1Butene	Sum of Isobutene and 1-Butene	C_4H_8	N/A	M
Z13Pentadiene	(Z)-1,3-Pentadiene	C_5H_8	1574-41-0	S
E13Pentadiene	(E)-1,3-Pentadiene	C_5H_8	2004-70-8	S

The CoreNames for aerosol variables use or are derived from terminology commonly found in literature.

Table 3: Example Aerosol Microphysical (AerMP) CoreNames and Description

CoreName	**Description**
NumConc	Number concentration of aerosol particles
NonVolatileNumConc	Non-volatile number concentration of aerosol particles
CCN	Cloud condensation nuclei number concentration
MassSizeDist	Mass Size Distribution i.e., mass concentration expressed as a function of aerosol particle size.

Table 4: Example Aerosol Chemical Composition (AerComp) CoreNames and Description

CoreName	Description
Sulfate	Particulate Sulfate Ion
SulfateSizeDist	Particulate Sulfate Ion as a function of particle size
Soot	Soot particles from combustion processes
OrganicAerosol	Particulate organic matter, including carbon and all other elements (e.g., H, O, N) in organic molecules

Table 5: Example Aerosol Optical (AerOpt) CoreNames and Description

CoreName	Description
Absorption	Light absorption coefficient of aerosol particles
SSA	Single Scattering Albedo of aerosol particles
AngstromExponentAbs	Angstrom exponent for absorption coefficients of aerosol particles
AOD	Column-integrated aerosol particle extinction coefficient

Table 6: Example Meteorology CoreNames and Description

CoreName	Description
StaticPressure	Ambient Atmospheric Static Pressure
DewPoint	Temperature to which air must be cooled to become saturated with respect to liquid water (or frost)
H2OMR	Mass mixing ratio of water vapor to dry air mass
UWindSpeed	E-W Horizontal Wind Speed, positive east

2.3 AcquisitionMethod

The acquisition method refers to the sampling geometry of the reported variable. Four acquisition methods are available to choose from as part of the standard name (Table 2).

Table 7: Controlled Vocabulary for AcquisitionMethod

AcquisitionMethod	Description
InSitu	Measurement in close proximity to a location of interest, typically near the instrument or sampling platform

AcquisitionMethod	Description
VertCol	Measurement of a vertically integrated column, where the column measured is nominally perpendicular to the earth's surface
SlantCol	Measurement of a vertically integrated column, where the column measured is not nominally perpendicular to the earth's surface (e.g., from a sun-tracking instrument)
Profile	Measurement of vertically resolved profile

2.4 DescriptiveAttributes

The descriptive attributes are components of the standard name that provide measurement and/or data reporting information relevant for data use and faceted data search, particularly when comparing results obtained with other methods of observations. The number and types of descriptive attributes are dependent on the MeasurementCategory (Table 1), but constant within each MeasurementCategory. For example, all trace gas standard names have two required descriptive attributes, MeasurementSpecificity and Reporting, whereas all aerosol optical property standard names have four required descriptive attributes, wavelength (WL), MeasurementRH, SizeRange, and Reporting. Within each of these descriptive attribute fields, e.g., WL, SizeRange, etc., there is a controlled list of terminology (7) that can be used. While some of the descriptive attribute categories overlap, e.g., Reporting, which is a descriptive attribute for trace gas, aerosols, and clouds, the list of controlled terms within that attribute is different. The use of the Reporting attribute reflects the practice that one measurement can be reported in different ways for different applications.

It is possible for a MeasurementCategory to not require any DescriptiveAttributes, e.g., Meteorology. In this case, "None" is used as the value for this attribute, *Met_StaticTemperature_InSitu_None*. There are also scenarios when a descriptive attribute for a MeasurementCategory might not be applicable, in which case the DescriptiveAttribute would be "None". For example, dimensionless variables such as SSA don't need a reporting attribute as in *AerOpt_SSA_InSitu_Green_RHd_Bulk_None*. A full description of each DescriptiveAttribute can be found in Appendix C.

2.5 Atmospheric Composition Variable Standard Name Examples

The following examples present a standard name for each MeasurementCategory using the controlled vocabulary for that category.

Gas_CoreName_AcquisitionMethod_MeasurementSpecificity_Reporting
Example of an in-situ measurement of CO2 gas reported in molar fraction with respect to dry air: **Gas_CO2_InSitu_S_DMF**

AerMP_CoreName_AcquisitionMethod_MeasurementRH_SizingTechnique_SizeRange_Reporting
Example of an in-situ measurement of aerosol particle number size distribution reported at reduced relative humidity derived from an aerodynamic sizing technique for coarse-mode aerosols at standard temperature and pressure:
AerMP_NumSizeDist_InSitu_RHd_Aerodynamic_Coarse_STP

AerComp_CoreName_AcquisitionMethod_SizingTechnique_SizeRange_Reporting
Example of an in-situ measurement of organic aerosols particles derived using a vacuum aerodynamic technique for accumulation-mode aerosol particles reported as mass concentration at standard temperature and pressure:
AerComp_OrganicAerosol_InSitu_VacuumAerodynamic_Accu_MassSTP

AerOpt_CoreName_AcquisitionMethod_WL_MeasurementRH_SizeRange_Reporting
Example of an in-situ measurement of absorption measured at a red wavelength under reduced humidity conditions with a bulk aerosol particle size range reported in ambient conditions:
AerOpt_Absorption_InSitu_red_RHd_Bulk_AMB

CldMicro_CoreName_AcquisitionMethod_SizingTechnique_SizeRange_Reporting
Example of an in-situ measurement of cloud particle number size distribution derived from an optical sizing technique measuring droplets being reported at ambient conditions:
CldMicro_NumSizeDist_InSitu_Optical_Drop_AMB

CldComp_CoreName_AcquisitionMethod_SizingTechnique_SizeRange_Reporting
Example of an in-situ measurement of the mass concentration of sodium in cloud water derived from a chemical technique where the particle measurement is not size resolved reported at ambient conditions: **CldComp_Sodium_InSitu_None_Bulk_MassAMB**

CldOpt_CoreName_AcquisitionMethod_WL
Example of an in-situ measurement of cloud particle extinction coefficient measured in the blue wavelength: **CldOpt_Extinction_InSitu_Blue**

CldMacro_CoreName_AcquisitionMethod_None
Example of a vertical column measurement of liquid water path:
CldMacro_LWP_VertCol_None

Met_CoreName_AcquisitionMethod_None
Example of an in-situ measurement of static temperature:
Met_StaticTemperature_InSitu_None

Platform_CoreName_AcquisitionMethod_None
Example of an in-situ measurement for aircraft Yaw angle: **Platform_YawAngle_InSitu_None**

MeasurementCategory_CoreName_AcquisitionMethod_MeasurementDirection_SpectralCoverage_Products

Example of photolysis rate coefficient for reaction NO2 + hv → NO + O(3P) derived from total actinic flux measurement:
GasJvalue_jNO2_InSitu_Total_Full_NO2-O3P

Rad_CoreName_InSitu_WLMode
Example of an in-situ measurement of Downwelling Diffuse Broadband Solar Irradiance between 0.2 and 3.6 micron: **Rad_IrradianceDownwellingDiffuse_InSitu_BB**

3 Maintenance

The Practical Guide for Atmospheric Composition Standard Names and Controlled Vocabulary [7] is intended to be a living document that contains detailed instructions on how to construct an atmospheric composition variable standard name and the controlled vocabulary lists for each component of the atmospheric composition variable standard names. To stay relevant to the measurements and user community, CoreNames will be updated and/or modified as part of each major field campaign to represent newly developed measurements or instruments. The vocabulary for descriptive attributes may also be updated, but far less often as the current lists capture all possibilities that are known to exist in the atmospheric composition field study data holdings. The process of creating new CoreNames happens on a case-by-case basis, between the document manager, principal investigator (PI), and other subject matter experts to ensure that the CoreNames follow the general format set forth already in the document (i.e., chemical nomenclature vs literature terminology) and are in line with what is commonly used in the literature and atmospheric composition community. The process for adding new DescriptiveAttributes will be the same process as adding new CoreNames, heavily relying on field experts to decide the best term to use. Due to the ever-growing nature of the controlled lists, the Practical Guide for Atmospheric Composition Variable Standard Names and Controlled Vocabulary document containing these lists is located at the Suborbital Science Data for Atmospheric Composition data repository, https://www-air.larc.nasa.gov/missions/etc/AtmosphericCompositionVariableStandardNames.pdf [7]. This document will be maintained by the ASDC DAAC in conjunction with field campaign data management teams. This team will be responsible for organizing the review process and conducting an independent literature search for verification purposes.

4 References

Normative References

[1] Practical Guide for Atmospheric Composition Variable Standard Names and Controlled Vocabulary, May 2022, https://www-air.larc.nasa.gov/missions/etc/AtmosphericCompositionVariableStandardNames.pdf

Informative References

[2] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G.,

Axton, M., Baak, A., Blomberg, N., Boiten, J-W.,
da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J.,
Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R.,…Goble, C. (2016) The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3. https://doi.org/10.1038/sdata.2016.18

[3] NASA Earth Science Data Systems Program. (2021, May 12). *Common Metadata Repository (CMR).* NASA Earthdata. https://www.earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr

[4] NASA Earthdata. *Earthdata Search.* search.earthdata.nasa.gov/

[5] Northup, E., Chen, G., Aikin, K., Webster, C., & Batchelor, D. (September 2016). *ICARTT Format Enhancements based on ESDSWG Recommendations* (Publication No. ESDS-RFC-029). https://www.earthdata.nasa.gov/s3fs-public/imported/ESDS-RFC-029.pdf

[6] Northup, E., Aikin, K., Webster, C., Chen, G., Aknan, A., Beach, A., Beyersdorf. A., Cantrell, C., DiGangi, J., Drdla, K., Early, A., Evans, K., Follette-Cook, M., Flynn, C., Gangl, M., Liang, Q., Shook, M., Silverman, M. (January 2017). *ICARTT File Format Standards V2.0* (Publication No. ESDS-RFC-029v2). https://cdn.earthdata.nasa.gov/conduit/upload/6158/ESDS-RFC-029v2.pdf

[7] NASA Atmospheric Science Data Center (2021, March 25). SUB-ORBITAL ORDER TOOL (SOOT) POWER USER INTERFACE (UI) RELEASE ANNOUNCEMENT [Press release]. https://asdc.larc.nasa.gov/news/sub-orbital-order-tool-soot-power-user-interface-ui-release-announcement

[8] Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Blower, J., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., Pamment, A., Juckes, M., Raspaud, M., Horne, R., Whiteaker, T., Blodgett, D., Zender, C., Lee, D.,…Herlédan, S. (2022, August 31) *NetCDF Climate and Forecast (CF) Metadata Conventions*, https://cfconventions.org/Data/cf-conventions/cf-conventions-1.10/cf-conventions.html

[9] *Scientific Variables Ontology (SVO)*, 2019, Retrieved March 8, 2022, from https://scientificvariablesontology.org/svo/index.html

[10] Peckham, S.D. (2014). The CSDMS Standard Names: Cross-Domain Naming Conventions for Describing Process Models, Data Sets and Their Associated Variables. In D. P. Ames, N. W. T. Quinn, & A. E. Rizzoli (Eds.), *Proceedings of the 7th International Congress on Environmental Modeling and Software.* ISBN: 978-88-9035-744-2

[11] Magagna, B., Moncoiffé, G., Devaraju, A., Stoica, M., Schindler, S., Pamment, A. (2022). InteroperAble Descriptions of Observable Property Terminologies (IADOPT) WG Outputs and Recommendations. Research Data Alliance. https://doi.org/10.15497/RDA00071

[12] *ACTRIS Vocabulary*. (n.d.). ACTRIS. Retrieved August 26, 2022, from https://vocabulary.actris.nilu.no/skosmos/actris_vocab/en/

5 Authors' Addresses

Primary Authors
Morgan Silverman, SSAI/NASA LARC, MS 401B, Hampton, VA, USA Tel: 757-864-3219
email: morgan.l.silverman@nasa.gov
Michael Shook, NASA LARC, michael.a.shook@nasa.gov
Rebecca Hornbrook, NCAR/UCAR, rsh@ucar.edu
Luke Ziemba, NASA LARC, luke.ziemba@nasa.gov
Sam Hall, NCAR/UCAR, halls@ucar.edu
Kirk Ullmann, NCAR/UCAR, ullmannk@ucar.edu
John Crounse, CalTech, crounjd@caltech.edu
Ryan Bennett, NSRC, r.bennett@baeri.org
Megan Buzanowicz, SSAI/NASA LARC, megan.e.buzanowicz@nasa.gov
Gao Chen, NASA LARC, gao.chen@nasa.gov

Contributing Authors
Barbara Barletta, University of California – Irvine, bbarlett@uci.edu
Helen Conover, UAH ITSC/NASA MSFC/GHRC DAAC, helen.conover@uah.edu
Josh DiGangi, NASA LARC, joshua.p.digangi@nasa.gov
Keith Evans, JCET/UMBC, evans@umbc.edu
Arlene Fiore, Columbia University, af2544@columbia.edu
Melanie Follette-Cook, Morgan State University/NASA GSFC, melanie.cook@nasa.gov
Peter Hall, NASA GSFC, peter.f.hall@nasa.gov
Peter Leonard, ADNET Systems/NASA GSFC, peter.j.leonard@nasa.gov
Qing Liang, NASA GSFC, qing.liang@nasa.gov
John Nowak, NASA LARC, john.b.nowak@nasa.gov
Scott Peckham, University of Colorado – Boulder, scott.peckham@colorado.edu
Hampapuram Ramapriyan, SSAI/NASA GSFC, hampapuram.ramapriya@ssaihq.com
Scott Ritz, NASA GSFC, scott.a.ritz@nasa.gov
Martin Schultz, Forschungszentrum Jülich, m.schultz@fz-juelich.de
Maria Stoica, University of Colorado – Boulder, maria.stoica@colorado.edu
Qian Tan, NASA ARC, qian.tan@nasa.gov
Andrew Thorpe, JPL, Andrew.K.Thorpe@jpl.nasa.gov
Jian Zeng, ADNET Systems/NASA GSFC, jian.zeng@nasa.gov

Working Group Members
Aubrey Beach, NASA LARC/ASDC currently at Booz Allen Hamilton
Stephane Beland, University of Colorado – Boulder, sbeland@colorado.edu
Stephen Berrick, NASA ESDIS, Stephen.W.Berrick@nasa.gov
Allan Doyle, International Interfaces, Inc., adoyle@intl-interfaces.com
Amanda Benson Early, NASA LARC/ASDC currently at ActiveCampaign
Scott Gluck, JPL, Scott.Gluck@jpl.nasa.gov
Nathan James, NASA ESDIS, nathan.l.james@nasa.gov

James Johnson, NASA GSFC, james.johnson@nasa.gov
Jeanne Laurencelle, University of Alaska – Fairbanks, jclaurencelle@alaska.edu
Dave Meyer, NASA GSFC, david.j.meyer@nasa.gov
George Milly, Columbia University, gpm2109@columbia.edu
David Moroni, JPL PO.DAAC, david.f.moroni@jpl.nasa.gov
Emily Northup, NASA LARC/ASDC currently at Trader Interactive
Deborah Smith, ESSC/NASA MSFC/IMPACT, deborah.smith@uah.edu
Chris Stoner, University of Alaska Satellite Facility, cstoner5@alaska.edu
Richard Strub, SSAI/NASA GSFC, richard.f.strub@nasa.gov
Jeff Walter, NASA LARC, jeff.walter@nasa.gov

Appendix A - Glossary

Acronym	Description
ACTIVATE	Aerosol Cloud Meteorology Interactions over the Western Atlantic Experiment
ARC	Ames Research Center
ASCII	American Standard Code for Information Interchange
ASDC	Atmospheric Science Data Center
CAMP2EX	Cloud, Aerosol and Monsoon Processes – Philippines Experiment
CCN	Cloud Condensation Nuclei
CF	Climate and Forecast (CF) Metadata Conventions
CSDMS	Community Surface Dynamics Modeling System
DAAC	Distributed Active Archive Center
DCOTSS	Dynamics and Chemistry of the Summer Stratosphere
DLH	Diode Laser Hygrometer
ESA	European Space Agency
ESDIS	Earth Science Data Information Systems
ESDS	Earth Science Data Systems
EVDC	Atmospheric Validation Data Centre
FAIR	Findable, Accessible, Interoperable, and Reusable
FIREX-AQ	Fire Influence on Regional to Global Environments and Air Quality
GHRC	Global Hydrometeorology Resource Center
GSFC	Goddard Space Flight Center
GSN	Geoscience Standard Names Ontology
HDF	Hierarchical Data Format

ICARTT	International Consortium for Atmospheric Research on Transport and Transformation
ITSC	Information Technology and Systems Center
JCET	Joint Center for Earth Systems Technology
JPL	Jet Propulsion Laboratory
LARC	Langley Research Center
MSFC	Marshall Space Flight Center
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NetCDF	Network Common Data Form
NSRC	National Suborbital Research Center
PO	Physical Oceanography
SSA	Single Scattering Albedo
SSAI	Science Systems & Applications, Inc
SVO	Scientific Variables Ontology
TAD	Toolsets for Airborne Data
TRACE-A	Transport and Atmospheric Chemistry near the Equator-Atlantic
UAH	University of Alabama - Huntsville
UCAR	University Corporation for Atmospheric Research
UMBC	University of Maryland, Baltimore County
WL	Wavelength

Appendix B – Identifying and Addressing CF Shortcomings

The CF Metadata Convection has three main shortcomings when handling in-situ atmospheric composition variables involved in suborbital field studies: measurand coverage and description; structure and vocabulary; and usability. First, CF standard names only cover a quarter to a third of the atmospheric composition relevant data products reported in suborbital field campaigns. The largest group of missing variables are hydrocarbon measurements. While measurands for aerosol and cloud variables are generally well represented, many attributes (qualifiers) are missing or do not adequately or accurately describe the measurements. These attributes are needed to properly define the measurement to ensure data usability. For example, the CF standard name "atmosphere_optical_thickness_due_to_particulate_organic_matter_ambient_aerosol" does not clearly define the size range of the aerosol particles, which is critical for comparing different in-situ measurements of aerosol particles. There are also inconsistencies within the CF standard names. For example, some aerosol properties have both dry and ambient relative humidity versions of their standard name, while others do not. With CF explicitly defining all allowed standard names, the addition of tens of thousands of extra CF names would be required to include all the additional species and attributes needed to account for various combinations.

Second, much of the specific vocabulary and terms currently used in CF are not suitable for atmospheric composition measurements since they are designed for model and forecast data. For example, “mass_concentration_of_pm1_ambient_aerosol_particles_in_air” specifically defines the measurement of PM1 aerosol as "…particulate compounds with an aerodynamic diameter of less than or equal to 1 micrometer”. This definition is misleading and incorrect if used for a mass measurement of ambient aerosol particles that is based on optical sizing rather than aerodynamic sizing technique. In atmospheric composition measurements, the measurements of aerosol particles can have different measurement techniques and size ranges. These details are important to understand for comparison of measurements with different techniques or comparison with model simulations.

The construction of CF names is governed by a set of guidelines that allows for the addition of qualifiers, such as surface, component, medium, process, and condition, to a base standard name using underscores. The difficulty in this, particularly from an interoperability standpoint, is that many of these qualifiers are optional, making it hard for a system or user to know the base standard name, since there is no set number of underscores or qualifiers. The standard names may also be derived from other standard names following a set of specified rules. In these cases, standard names can begin with words such as “tendency_of_X” or “product_of_X_and Y” where X and Y are the other standard names. The range of possibilities for CF standard names poses a problem for discoverability and interoperability.

Lastly, the usability of CF is difficult for atmospheric composition variables. Heavy (large) nonmethane hydrocarbons can have ambiguous names where one word can represent many variables, while other times many terms are used for the same variable. Therefore, an accurate description, and CAS number in the case of trace gases, is important to clarify what is being referred to.

The atmospheric composition variable standard names intend to address these CF shortcomings in several ways. First, the CoreNames included in the controlled vocabulary of each MeasurementCategory cover all known variables currently measured during suborbital field campaigns. These CoreNames, along with the chosen descriptive attributes for each MeasurementCategory, provide all possible combinations of standard names needed for current measurements. The structure chosen for the atmospheric composition variable standard names allows data providers to mix and match the appropriate values suitable for their measurement, which limits the number of necessary CoreNames. This contrasts with CF, which has taken the approach to explicitly define all allowed standard names as a whole. The structure is also designed to increase interoperability and machine learning by a predefined number of underscores and qualifiers depending on the MeasurementCategory chosen. There are no optional qualifiers; every underscore must be used in that category even if the value chosen is ‘none’. The descriptive attributes chosen for each MeasurementCategory have been chosen to enhance usability for those unfamiliar with the measurements and in research use (e.g., in measurement comparisons) by including important information about each measurement. An important aspect that the atmospheric composition variable standard names addresses is that there are many names to represent one variable but also one name that can represent many variables. Therefore the CAS number and description has also been included in the tables of CoreNames to remove any ambiguity that exists.

Appendix C – Descriptive Attributes

The following is a general overview of the descriptive attributes used within the atmospheric composition variable standard names. For more details, see the Practical Guide for Atmospheric Composition Variable Standard Names and Controlled Vocabulary [7].

Trace Gas Descriptive Attributes:
The "MeasurementSpecificity" attribute specifies whether the CoreName represents a single species, combination of multiple species, or is not applicable (e.g., for gas phase reaction rate(s) or ratio of species).

The "Reporting" attribute describes the way a trace gas is reported. When reporting in standard temperature and pressure (STP), the temperature and pressure conditions under which the measurement is reported must be noted in the header or metadata of the data file, as "standard temperature" varies across the research community.

Aerosol Descriptive Attributes:
Relative humidity (RH) conditions are important for both aerosol particle microphysical and optical measurements because water vapor can condense onto the particle and change its size and optical properties. In-situ aerosol particle measurements can be made or calculated at different RH levels: dry, ambient, or specified. If a specific RH is used, the relative humidity at which the measurement is reported must be documented in the variable description.

Aerosol and Cloud Descriptive Attributes:
"SizingTechnique" is an important descriptive attribute because the measurement of the size of a single particle can vary when using different techniques (based on the properties of the particle, such as its composition, shape, and density). Each technique has inherent assumptions, limitations, and operable ranges that are vital for proper interpretation and comparison of the data. AerComp, AerMP, and CldMicro MeasurementCategories all use this descriptive attribute.

The "SizeRange" delineates the range of particle sizes being measured. This descriptive attribute is necessary for aerosol and cloud MeasurementCategories.

Aerosol particle and cloud optical properties are functions of wavelengths (WL) of light. Therefore, a measurement of aerosol particle optical properties is made at one or more specific wavelength(s).

Aerosol particle and cloud microphysical and chemical composition standard names also have a 'Reporting' DescriptiveAttribute to indicate the reporting method used. For variables that are dimensionless (e.g., fRH, SSA) the reporting attribute is "None". When reporting in standard temperature and pressure (STP), the temperature and pressure conditions under which the measurement is reported must be noted in the header or metadata of the data file, as "standard temperature" varies across the research community. Similarly, if the reporting attribute is

“EnvSp”, the specific environment temperature and pressure must be referenced in the header or variable description.

Photolysis Variable Descriptive Attributes:
MeasurementDirection describes if the photolysis rates are derived from downwelling, upwelling, or total (Downwelling and Upwelling) actinic flux measurements.

SpectralCoverage indicates whether the spectral range of the measurement spans the entire range of photolysis or only a partial range (e.g., UV/Visible range only).

Products is used to list the products from photolysis reactions, separated by a hyphen (“-”). If no specific products are identified in the photochemical kinetic study of the reaction, “Products” has the value of “NoProductsSpecified”.

Radiation Descriptive Attributes:
WLMode refers to the spectral measurement mode. While measurement spectral range is important, fully describing it requires specific wavelength information, which is beyond the scope of the broad ranges and controlled vocabulary of standard names. Specific spectral range information should be given in the variable description, e.g., in the long variable name in the ICARTT format.