ORNL/CDIAC-112/R2

NARSTO Data Management Handbook

Sigurd W. Christensen, Thomas A. Boden,
Les A. Hook, and Meng-Dawn Cheng

NARSTO Quality Systems Science Center
Oak Ridge National Laboratory, Oak Ridge, Tennessee

Latest revision available online at
http://cdiac.esd.ornl.gov/programs/NARSTO/narsto.html#qsmp

Environmental Sciences Division
Publication No. 4917

Revision Date: August 16, 1999

Prepared for the
Environmental Sciences Division
Office of Biological and Environmental Research
U.S. Department of Energy
Budget Activity Number KP 12 04 01 0

Prepared by the
Carbon Dioxide Information Analysis Center
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831-6335
managed by
LOCKHEED MARTIN ENERGY RESEARCH CORP.
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-96OR22464

TABLE OF CONTENTS

SUMMARY OF CHANGES BETWEEN EDITIONS i
ABBREVIATIONS ii

1.0 INTRODUCTION 1-1
1.1 Purpose/scope 1-2
1.2 Audience 1-2
1.3 How to Use this Handbook 1-4
1.4 Data Management Roles and Responsibilities 1-4
1.5 Organization of this Handbook 1-5

2.0 RECOMMENDED DATA REPORTING CONVENTIONS 2-1
2.1 Expressing the Amount of a Substance 2-1
2.2 List of Common NARSTO Parameters and Recommended Reporting Units and Formats 2-1
2.3 SI Multiples, Prefixes, and Symbols 2-3
2.4 Reporting Data Below Minimum Detectable Limits 2-4
2.5 Time Averaging 2-4
2.6 Date and Time Formats 2-4
2.7 Reporting Missing Data 2-4
2.8 Reporting Calibration Values and Uncertainty Estimates 2-5
2.9 Qualifying Data 2-5

3.0 DATA VALIDATION 3-1
3.1 NARSTO Data and Research Product Validation 3-1
3.2 Data Validation Codes for External Data Sets 3-3

4.0 DOCUMENTING AND TRANSMITTING DATA 4-1
4.1 NARSTO Data and Information Categories 4-1
4.2 Recommended Data Transfer Format 4-2
4.3. Data Exchange Standard File Overview 4-4
4.4 Data Exchange Standard Key Phrases 4-5
4.5 Metadata Tables 4-19
4.6 Measurement Data Table 4-22
4.7 Sample Data Exchange Standard File 4-24
4.8 Steps to Follow to Prepare and Send Files in the Data Exchange Standard . . 4-29

5.0 ARCHIVAL AND DISSEMINATION OF NARSTO DATA AND INFORMATION 5-1
5.1 Data Archival 5-1
5.2 Data and Research Product Access and Distribution 5-1
5.3 Product Access 5-2
5.4 Product Distribution 5-2
5.5 Retrieval of Data from the Permanent Data Archive 5-2

6.0 RECORDS MANAGEMENT 6-1

7.0 SOFTWARE CONFIGURATION CONTROL AND DATA MANAGEMENT SYSTEM ADMINISTRATION GUIDELINES 7-1
7.1 Project-specific Database and Software Requirements 7-1

7.2 Day-to-Day Operation of Data Management Systems 7-3

APPENDIX A: COMPLETE LISTING OF STANDARD VARIABLE NAMES, UNITS, AND FORMATS A-1

APPENDIX B: SI MULTIPLES, PREFIXES, AND SYMBOLS B-1

APPENDIX C: KEY PHRASE SUMMARY TABLES C-1

APPENDIX D: NOTES FOR DATA ORIGINATORS USING SPREADSHEETS D-1

SUMMARY OF CHANGES BETWEEN EDITIONS

This is a “living” document; it will be changed and updated from time to time. Major changes will be issued as new whole revision numbers (e.g., Revision 2 is ORNL/CDIAC-112/R2). More minor changes will be issued as updates to the major revisions. Users should ensure that they have the current edition of this document.

First Edition, April 21, 1998

The original NARSTO Data Management Handbook was released.

Revised Edition 1.1, February 12, 1999

Various clarifications were added, and corrections were incorporated. Section 4 was reorganized.

Revised Edition 2, July 28, 1999

The footer was modified to decouple the version of this handbook (identified by a revision number and date) from the version of the Data Exchange Standard that it describes (identified by a generally different date and a version number). Additional clarifications and details were added, and changes and corrections were incorporated, mostly in Sections 2, Sections 4.4 through 4.7, and Appendices A, C, and D. The changes made requirements more explicit, often by changing “should” to “must.” A few requirements were added, particularly relating to how the station identifier is defined. One key new requirement is that **all** short names must be six or fewer characters, not only those that may be associated with flags. This enables automated file transfer verification statistics to be generated as part of the Quality Assurance process. The example file in Section 4.7 was replaced with one that illustrates a more complicated file structure and conforms with current requirements. Many formats in Appendix A were increased slightly, some new variable names were added, and a few short names for volatile organic compounds were revised. Appendix C was updated, reflecting changes in the QA code’s short names (which do not affect Data Originators), and increases in the allowed lengths of many fields used with the Key Phrases.

Revised Edition 2a, August 16, 1999

Clarifications were added. The recommended data qualification flag codes in Section 2.9 (Qualifying Data) were modified by expanding the missing value codes. The Measurement Data Table instructions now make explicit the requirement for a validation level code with the results.

ABBREVIATIONS

AES	Canadian Atmospheric Environment Service
AQAR	Annual Quality Assessment Report
BIL	Band Interleaved By Line
BSQ	Band Sequential Format
CSV	Comma Separated Values
DAAC	Distributed Active Archive Center
DQOs	Data Quality Objectives
DU	Dobson Units
EOS	NASA Earth Observing System
EOSDIS	NASA Earth Observing System Data and Information System
FGDC	Federal Geographic Data Committee
FTP	File Transfer Protocol
GC	Gas Chromatograph
GIS	Geographic Information System
LaRC	Langley Research Center (Langley, Virginia, USA)
MDL	Minimum Detectable Limits
NASA	U.S. National Aeronautics and Space Administration
NARSTO	(formerly) North American Research Strategy for Tropospheric Ozone
NAtChem	Canadian National Atmospheric Chemistry Database and Analysis Facility (Toronto, Ontario, Canada)
PDA	Permanent Data Archive (LaRC, Langley, Virginia, USA)
QA	Quality Assurance
QC	Quality Control
QSMP	Quality Systems Management Plan

PQMP	Program Quality Management Plans
QIWP	Quality Integrated Work Plan
QSSC	NARSTO Quality Systems Science Center (Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA)
RCS	Revision Control Software
SCOS97	Southern California Ozone Study 1997
SDTS	Spatial Data Transfer Standard
SOPs	Standard Operating Procedures
U.S. EPA	United States Environmental Protection Agency
UTC	Coordinated Universal Time
WWW	World-Wide Web

1.0 INTRODUCTION

The long-term usefulness of results and products stemming from the NARSTO program will only be realized by ensuring that NARSTO measurements and models are adequately described and maintained. This ***NARSTO Data Management Handbook***, prepared by the NARSTO Quality Systems Science Center (QSSC), is intended to offer guidance to NARSTO participants concerning the management of NARSTO data, products, and records and to ultimately help ensure the long-term utility of NARSTO products. Universal adoption of the data management guidance offered here by NARSTO scientists will speed the flow of NARSTO data from collection to archival, reduce data processing and re-formatting requirements by users of NARSTO data, facilitate and ease analysis of NARSTO data, improve understanding of NARSTO data, and enable the NARSTO QSSC and Permanent Data Archive (PDA) to develop tools to automate quality assurance and archival efforts. The data and records management guidance that follows encompasses documentation and reporting, formatting, transmission, archival, and dissemination.

This handbook, the **NARSTO Data Management Handbook**, is one of many NARSTO documents compiled to ensure quality within NARSTO. The NARSTO ***Quality Systems Management Plan (QSMP)*** defines a set of quality systems documentation, starting with the upper level QSMP, that identifies NARSTO program quality assurance and data management requirements and guidelines. Within the flexible QSMP framework, Program/Project Managers will establish a ***Program Quality Management Plan (PQMP)*** as the second level of quality planning. Individual PQMPs should describe only the directly applicable quality activities and expectations for that specific program. The third level of quality planning will occur at the project level, where each Principal Investigator is responsible for developing a ***Quality Integrated Work Plan (QIWP)***. The scope, content, and level of detail of a QIWP should be based on the nature of the project and applicable PQMP guidance. With each successive level, the specificity of requirements and quality management activities increases. This ***NARSTO Data Management Handbook*** is a supplement to the QSMP and provides specific tools to assist principal investigators in developing QIWPs and in delivering their quality assured research products to the client community in a timely and efficient manner. Program/Project Managers and Principal Investigators should evaluate the applicability of the guidance to their project and incorporate appropriate elements into the project documentation. Figure 1, of the QSMP Executive Summary, lays out alternative pathways for project planning and archival of NARSTO data.

The handbook is intended to be a dynamic document that can be modified and expanded as often as needed to meet the needs of the NARSTO program. We strongly encourage readers and users of this handbook to provide comments and suggestions at any time. In order to present this information in as cost-effective and as efficient a manner as possible, the hard copy distribution of this handbook will be kept to a minimum and the handbook will be available in English only. A configuration-controlled web accessible copy of the current handbook will be linked to the NARSTO Program Office and QSSC Home Pages.

1.1 Purpose/scope

The purpose of this handbook is to provide data management guidance for NARSTO programs/projects to use, first, as they develop project planning documentation, and second, as they implement project data and records management activities. The data management guidance will be a valuable reference for Principal Investigators as they prepare QIWPs, while the data processing and records management protocols can be implemented as needed by projects. Using the guidance in this handbook will help ensure adequate and consistent documentation of data products as the NARSTO Program works to facilitate consolidation and archival of appropriate new project data, historical data, and research products for subsequent use by researchers. Benefits to NARSTO scientists derived from adopting these guidelines includes ease of using, analyzing, ingesting, documenting, and transferring NARSTO data; cost and time savings derived from repeated use of data tools developed to handle NARSTO data; and greater ease in preparing documentation that will improve the likelihood that future users will gain sufficient understanding of their data to permit appropriate use.

The NARSTO programmatic goal in providing this guidance is to move towards the implementation of more consistent, comprehensive, and efficient approaches for managing environmental data. The handbook supports a life-cycle approach to environmental information management. While established programs/projects may have processes and systems in place, new projects to be funded in the future will benefit from having available a supported set of planning and implementing resources.

1.2 Audience

The principal readers and users of this guidance document are NARSTO program/project managers, Principal Investigators, and data coordinators. The Principal Investigators have the primary responsibility for preparing a project's QIWP. The data coordinator may refer to this handbook to ensure that the project's data management activities are properly addressed and will be completed in a timely, appropriate, and cost-effective manner. Project staff implementing data management processes and computer systems may refer to the applicable protocols. Other members of the project team may also find this document useful in understanding the issues important in the management of project environmental information. Users of NARSTO data will benefit from this handbook by knowing the formats of NARSTO data well before the data actually reach the PDA at the Langley Research Center Distributed Active Archive Center (DAAC). The present and future atmospheric chemistry community will benefit from having a common data standard shared by many NARSTO groups across North America.

Project Planning and Data Archival Process

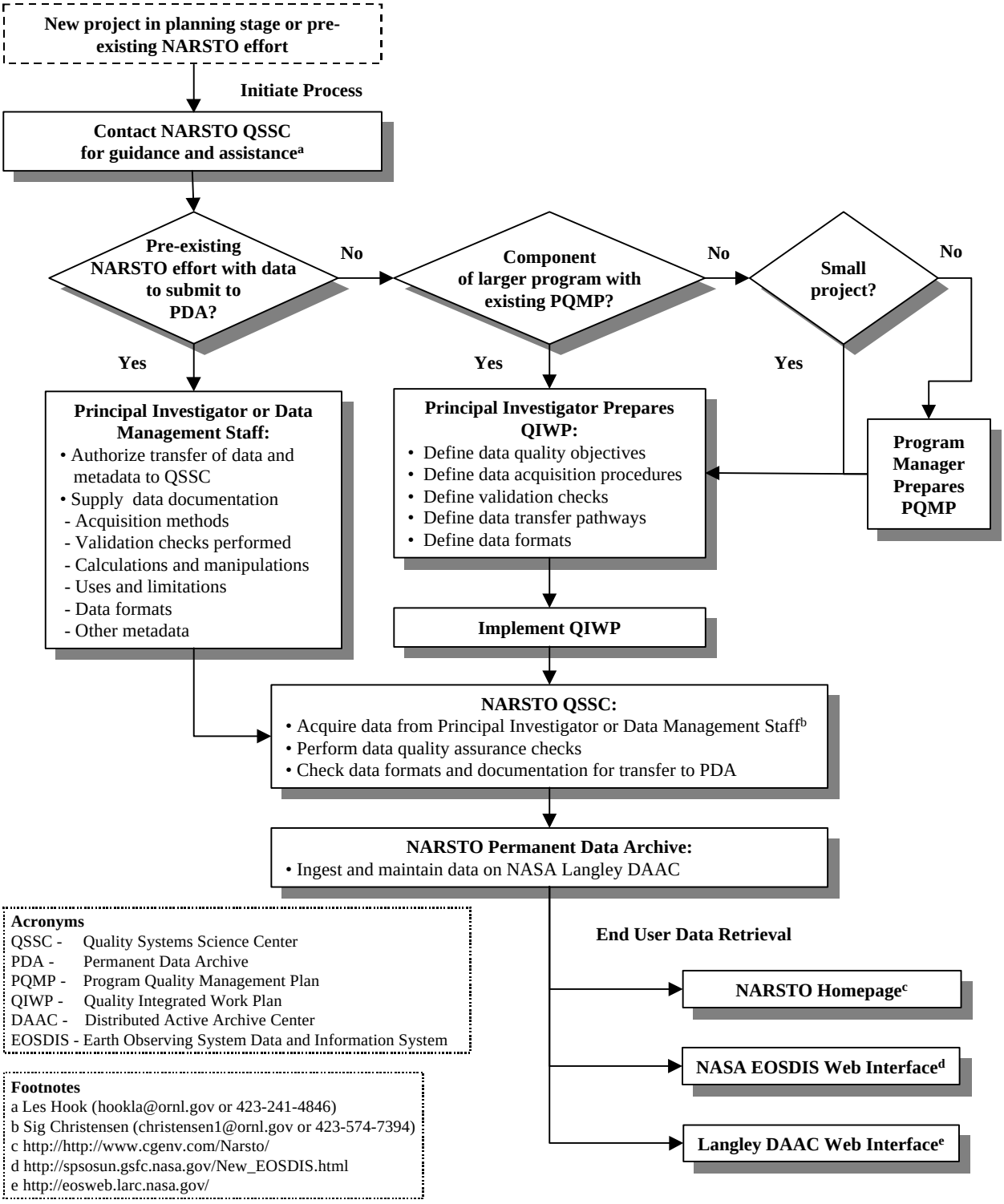

Figure 1. Project Planning and Data Archival Process

1.3 How to Use this Handbook

The QSMP identifies the NARSTO program quality assurance and data management requirements and guidelines for ensuring NARSTO product credibility, reliability, accessibility and quality. This handbook supplements the QSMP by providing more specific guidance and tools for planning and implementing data management-related project tasks.

Established research and modeling programs may already be performing many of the components discussed in the following sections but may find this handbook valuable as a review or check of presently implemented activities. When a new task or revisions of current work plans (QIWPs) is undertaken, this guidance should be evaluated for applicability.

New programs/projects should incorporate as many of the elements as appropriate into their new QIWP, follow up with applicable references, and consult with the QSSC to take advantage of previously implemented processes and systems that may meet their needs.

NARSTO scientists are encouraged to use the International System of Units (SI) for reporting purposes, where appropriate, and to comply with prevailing and future spatial data policy, public access policy, and metadata content standards. NARSTO project scientists and data managers are asked to adopt the data format conventions and data validation protocols discussed below to ensure data compatibility throughout NARSTO.

1.4 Data Management Roles and Responsibilities

QSSC

The QSSC will work closely with Program/Project Managers, Principal Investigators, Project Data Staff, and the Committee on Quality Systems and Data Management to develop data and records management specifications and guidance. This guidance will be compiled, distributed, and maintained in the ***NARSTO Data Management Handbook*** (this document).

The QSSC will coordinate the transfer of quality assured NARSTO data from Principal Investigators and program/project archives to the NARSTO Permanent Data Archive.

Permanent Data Archive

The Permanent Data Archive (PDA) is the facility for the long-term archival and public distribution of NARSTO validated electronic data and data products. The PDA is located at the NASA/Langley Research Center (LaRC) Earth Observing System (EOS) Distributed Active Archive Center (DAAC).

Committee on Quality Systems and Data Management

The Committee on Quality Systems and Data Management assists the QSSC in formulating NARSTO quality assurance and data management policy and guidance. The Committee had significant input regarding the content of this Handbook.

Principal Investigators

The Principal Investigator will ensure that an approved QIWP is in place to guide NARSTO research and modeling activities and that it includes the environmental data and records management tasks to ensure delivery of quality assured research products to the client community in a timely and efficient manner. Principal Investigators of existing NARSTO projects with data to submit to the PDA will prepare and transfer their data and metadata to the QSSC.

Data Management Staff

Working with the Principal Investigators, Data Management Staff will implement the guidance in this handbook and consult with the QSSC to ensure that project data are transferred in a consistent format and are adequately documented for the PDA.

1.5 Organization of this Handbook

This document is divided into seven sections, followed by four appendixes.

Section 1 defines the purpose, scope, organization, and audience for the document and gives general suggestions on how to use this handbook.

Section 2 describes data reporting conventions for NARSTO parameters including data formats, units, and symbols.

Section 3 describes validation activities.

Section 4 provides data documentation and data transfer guidance.

Section 5 discusses data archival and dissemination guidelines.

Section 6 discusses records management guidelines.

Section 7 identifies software configuration control and data management system administration guidelines.

Appendix A: Complete Listing of Standard Variable Names, Units, and Formats

Appendix B: SI Multiples, Prefixes, and Symbols

Appendix C: Key Phrase Summary Tables

Appendix D: Notes for Data Originators Using Spreadsheets

2.0 RECOMMENDED DATA REPORTING CONVENTIONS

APPLICABILITY:	The consistent use of data format, unit, processing, and qualification reporting conventions will enhance data sharing, improve the efficiency of data transfers, and facilitate analyses.
USE:	Guidance for the project principal investigators, measurement analysts, data reviewers, and project data management staff for NARSTO data management planning and process implementation. During the development of a project's QIWP, select applicable reporting conventions and implement appropriately to the project's needs.
CONSIDERATIONS:	NARSTO scientists are encouraged to use SI reporting units in most cases. However, it is recommended some parameters be reported in non-SI units and NARSTO scientists should be cognizant of these exceptions.
SUMMARY:	Universal adoption of the data reporting guidance offered here by NARSTO scientists will speed the flow of NARSTO data from collection to archival, reduce data processing and re-formatting requirements by users of NARSTO data, facilitate and ease analysis of NARSTO data, improve understanding of NARSTO data, and enable the QSSC and PDA to develop tools to automate quality assurance and archival efforts.

2.1 Expressing the Amount of a Substance

Concentration is the relative content of a component. Mixing ratio is similar to concentration but generally refers to dry samples (e.g., dry air sample) since moisture can influence the relative amount of a component in a substance. All chemical measurements should be reported in mixing ratio units of parts per billion by volume (ppbv). Chemical species collected in “mass per volume” units (e.g., nitric acid, organic nitrates, and sulfate) should be converted to mixing ratio units with specific mention of the conversion factor. Concentrations measured in mass units per standard cubic meter (e.g., kg/m^3) should consider STP as 0°C and 760 mm Hg pressure.

2.2 List of Common NARSTO Parameters and Recommended Reporting Units and Formats

It is imperative that measured parameters be fully described and units explicitly stated in any NARSTO data management system or when NARSTO investigators provide data to others. The following table lists parameters commonly measured or required under the NARSTO program and the recommended units and formats for reporting these parameters. Generally speaking, NARSTO recommends the use of SI units; however, it is recognized that some parameters are routinely reported in non-SI units such as Dobson Units (DU) for column ozone density. Users should follow the reporting guidance in the following table; additional parameters and details are given in Appendix A. If questions concerning the appropriate reporting units and format persist, please contact the QSSC.

Chemical species, group, or parameter	Units	Reporting format	Variable name
O_3	ppbv	xxx.x	o3
NO, NO_y, or NO_x	ppbv	xxx.xx	no, noy, or nox
PAN	ppbv	xxx.xx	pan
CO	ppbv	xxx	co
SO_2	ppbv	xxx.xx	so2
Hydrocarbons	ppbv	xxxx.xx	See Appendix A
Carbonyls	ppbv	xxxx.xx	See Appendix A
Wind speed	m/s	xx.x	winspd
Wind direction	decimal degrees	xxx.x	windir
Temperature, ΔTemperature, or Dew point	degrees C	xxx.x	temp, dtemp, and dewpnt
Mixing height	m agl	xxxx	mixhgt
Relative humidity	%	xxx.x	relhum
Solar radiation	Watts/m^2	xxx.x	solrad
UV radiation	Watts/m^2	xx.x	uvrad
Pressure or partial pressure	pascals	xx.x	press
Barometric pressure	mb (adjusted to sea level)	xxxx.x	bpress
Precipitation	mm	xxx.x	precip
Altitude	m above msl	xxxxx.x	alt
Latitude and Longitude	decimal degrees	±xxx.xxxxx	latdec, londec [1]

For species or parameters not listed above nor in Appendix A, we suggest locating a similar species or parameter and using the same units and reporting format[2]. Please coordinate with the QSSC to determine an appropriate and unique variable name.

---

[1] South latitude and west longitude should be expressed as negative values.

[2] Other general NARSTO-related measurements and acceptable reporting units include the following:
Column abundance - mol/m^2 or kg/m^2

Concentrations of airborne particulate matter - number of molecules per volume (m^3) or mass per volume (kg/m^3).

Aqueous phase mixing ratios and concentrations - mole/kg, mole/L, or μmol/dm

Aerosol particle concentration - number of particles per cm^{-3}, total surface area of particles (e.g., m^2), or mass (e.g., μg) per m^3

Transfer rates - mol/s, molecule/s, or μg/s

Flux rates - mol m^{-1} s^{-1}, molecule m^{-1} s^{-1}, or μg m^{-1} s^{-1}

Emission and deposition rates - kg/h or Mg/m^2

If you cannot locate a similar entry in the above table, we recommend the use of the following standard SI units. If you are still unsure of the proper reporting convention, please contact the QSSC for further consultation.

Measure or Calculation	Units
Length/area/volume	meter (m)/m^2/m^3
Time	second (s), minute(m), hour (h), day(da), and year (yr)
Frequency	hertz (Hz)
Speed and acceleration	m/s and m/s^2
Energy	joule (J)
Power	watt (W)
Force	newton (N)
Mass	kilogram (kg)
Density	mass $length^{-3}$ (kg m^{-3} or g cm^{-3})
Velocity	length $time^{-1}$ (cm/s)
Electric current/electric potential/electric resistance	ampere (A), volt(V), and ohm
Electric conductivity	siemans
Luminous intensity	candela (cd)
Plane angle	radian (rad)
Solid angle	steradian (sr)
Spectral irradiance	W m^{-2} nm^{-1}
Broadband irradiance	W m^{-2}

2.3 SI Multiples, Prefixes, and Symbols

The standard SI multiples, sub-multiples, prefixes, and symbols are recommended by NARSTO for use in specifying units for NARSTO variables. Refer to the table in Appendix B for a complete listing.

2.4 Reporting Data Below Minimum Detectable Limits

Measurement results below the minimum detectable limits (MDL) of the instrument should be reported as measured and to the level of precision of the instrument, but flagged accordingly. Data values (e.g., averages) derived from any MDL data should be flagged as specified in Section 2.9 below.

2.5 Time Averaging

Time-averaged data must include times of the beginning and the end of the time-averaging period. Contributors are encouraged to submit one-hour averages and preferably time-averaged data reflecting the finest level of sampling granularity. A valid time-averaged data set must contain validated data points representing at least 75% of the total data points possible for the averaged time period. Therefore, a 60 minute time averaged data set based on 1-minute samples must contain at least 45 validated 1-minute data points. Otherwise, the time averaged value is flagged and reported as missing datum. The sampling frequency and averaging period should be indicated in the header records of each data file.

2.6 Date and Time Formats

Sample dates and times must be reported for all measurements and in two formats: (1) local time (either clock time or standard time for the local time zone), and (2) Coordinated Universal Time (UTC). In addition, the local time coordinates used and the UTC time coordinates are specified for every record. Both the begin time and end time must be reported for every record, and in both time formats. Begin time must be reported as time at the beginning of the averaging period. End time must be reported as the time at the end of the averaging period. The daily time cycle runs from 00:00:00 to 23:59:59 (24:00:00 is not a legitimate value). Sampling times should be reported as hh:mm, or as hh:mm:ss, when practical and possible. Analysis dates should be reported for all measurements where air sample determinations are made on dates other than the sampling date. Reported dates must include the day, month, and year and be formatted as yyyy/mm/dd (e.g., 1997/08/15 or 1997-08-15). Character values may not be used to denote sampling or analysis months and leading zeros should be used for day and month entry values less than ten (i.e., 08 to represent August, not 8 or AUG).

2.7 Reporting Missing Data

All data fields (i.e., non-character data) must have a value present, either the measured or adjusted data value or a missing value representation. **There may not be blank data fields**. It is recommended that all flag fields also have entries. Contributors should report data where possible and use flag codes. All missing values for non-character fields should be numerical values, not character or alphanumeric values, to aid quality-control efforts. Missing values for data parameters should be represented by negative numerical values outside the range of potential values (e.g., -9.9 should not be used for temperature) and should be reported with the same level of precision as all other values for that field (e.g., -999.99 for ozone mixing ratios). Dates, times, and spatial coordinates have special requirements for missing value codes, which are explained in Section 4.4.3. Data flag codes should differentiate between valid values, invalid values, estimated values, interpolated values, MDL values, and missing values (see Section 2.9).

2.8 Reporting Calibration Values and Uncertainty Estimates

Calibration values for all measurements should be reported as a separate (ancillary) file using the same reporting conventions and formats as the actual measurement. Access to calibration values is crucial for many quality-assurance, analytical, and modeling exercises.

Uncertainty estimates should be reported for **each** parameter. These estimates should be provided either in the measurement method information table or in the primary data table as separate data fields. Uncertainty estimates should not be offered in a separate file nor should they be inferred as part of a flag code. The metadata that accompanies the data file should describe the investigator's method of calculating uncertainty for each parameter.

2.9 Qualifying Data

The following table lists the recommended NARSTO data qualification flag codes and definitions. Every data point should have an associated data qualification flag code, in addition to any field or laboratory data qualifiers. Flags begin with the letter "V" for valid values, "M" for missing values, or "H" for historical data unable to be assessed or validated. No invalid non-missing data will be placed in the PDA to avoid possible inadvertent use. Project data systems or program data archives may contain invalid data under appropriate administrative controls.

Data qualification flag codes	Definition
V0	Valid value
V1	Valid value but comprised wholly or partially of below-MDL data
V2	Valid estimated value
V3	Valid interpolated value
V4	Valid value despite failing some statistical outlier tests
V5	Valid value but qualified because of possible contamination (e.g., pollution source, laboratory contamination source)
V6	Valid value but qualified due to non-standard sampling conditions (e.g., instrument malfunction, sample handling)
M1	Missing value because no value is available
M2	Missing value because invalidated by Data Originator
H1	Historical data that have not been assessed or validated

3.0 DATA VALIDATION

APPLICABILITY:	All NARSTO data products must be validated.
USE:	Guidance for NARSTO program/project managers, principal investigators, and project data coordinators in planning and implementing data validation activities and assigning data validation codes.
CONSIDERATIONS:	Per the NARSTO QSMP, additional validation activity descriptions will be in the individual project's QIWP.
SUMMARY:	All NARSTO data products must be validated and all NARSTO data products should include a flag designating the level of validation. The data product validation flags will have values ranging from zero (0) to three (3).

Data validation is the process of determining and denoting the quality of a data set (data having either a common method of collection or data collected by various methods in one location). The validation process consists of evaluating the internal, spatial, temporal and physical consistency of each data set for invalid data and for outliers (data that are physically, spatially, or temporally inconsistent). During validation, physically unrealistic data are invalidated, biases and instrumental drift are noted, and gross errors are identified. The objective of this process is to produce an archive with values that are of known quality.

All NARSTO data should include a flag designating their level of validation with values ranging from zero (0) to three (3). Initially, data will be at the lowest level of validation. In response to further quality checks, the validation flag will be upgraded. Revisions to flag codes will be made at the project level and records of these changes will be maintained. When upgraded supplemental data are received or when additional validation work is performed on existing data, the project will send the QSSC a revised data set reflecting the new validation status or provide the QSSC and PDA with specific instructions to revise flag codes in archived data sets. Details on validation activities and actions may be included with the metadata. Because not all activities are performed concurrently, at any given time the archive may contain data at different validation levels.

3.1 NARSTO Data and Research Product Validation

All NARSTO products must be validated. Recognizing the potential for NARSTO to generate research products other than "data" is important. Other possible products are simulation models, methods, procedures, and reports. Each of these products must be validated. A validation level and status discussion must be included in the metadata record or information

associated with the data or research product. The following general levels of validation will be used in NARSTO; however, Program/Project Managers and Principal Investigators must specify in the PQMP and/or QIWP more specific validation checks that will be performed for any given research product. Table 3-1 summarizes the characteristics and status of data and research products at each level of the validation process.

Level 0 Validation indicates a reasonably complete data set of unspecified quality that consists of research products subjected to minimum processing in the field and/or in the laboratory by project staff. Level 0 designations will be given to raw data and other research products that have not been audited or peer reviewed. Level 0 status will remain in force until all audits or peer reviews associated with Level 1 validation of the product have been completed and the investigator's response recorded.

Level 0 data contain all available measurement data and may also contain data in the form of quality control checks and flags indicating missing or invalid data. Level 0 data consist of instrument outputs expressed in engineering units using nominal calibrations. Missing data from on-site backup loggers or strip charts have been filled in. Level 0 data may include flags indicating QC check data, power failures, excessive rate-of-change, insufficient data for the averaging period, or other logger programmed occurrences.

Level 1 Validation indicates a complete data set of specified quality that consists of research products subjected to quality assurance and quality control checks and data management procedures. As part of the Level 1 process, site documentation is reviewed for completeness and performance compared with other locations. Compliance with documented data quality objectives, standard operating procedures (SOPs), and research protocols is evaluated in the Level 1 process. Audit and peer review reports have been evaluated (and necessary corrective actions taken) for all research products designated Level 1. The comparison and cross-checking activities done under Level 1 may be conducted by the project staff, QSSC, and/or the scientific community.

Level 1 data are generated by project groups. In response to audits, data may have been adjusted. The project group, responsible for submitting the data, will adjust the data for "blank bias" (lab analyses) or "zero drift" (continuous ambient measurements), will determine precision and accuracy, and will perform consistency checks with other data within the same data set. These internal consistency checks might include diurnal analyses to look for expected patterns or time series analyses to detect outliers, extreme values, or time periods with too little or too much variation. Level 1 designation will be assigned after the project group has performed all quality control activities identified in their QIWP and addressed all quality issues stemming from audits and reviews. The appropriate project group, responsible for submitting the data, will fill in missing data and transmit the Level 1 data to the QSSC in the format described later in this handbook within twelve months of data collection.

Level 2 Validation indicates a complete, externally consistent data set of specified quality that consists of research products that have undergone interpretative and diagnostic analysis by the project staff or user community. A validation level and status discussion must be included in the metadata record associated with the research product.

Level 2 data have been closely examined by the data manager and/or data users for external consistency when compared to other related data sets. External checks might include correlation by scattergram, comparison of data with other similar data for the same time period,

and comparison of a measurement made by two different methods (e.g., hydrocarbons by auto-GC and canister).

If comparisons are not within the precision of the measurements, then measurement records and other information will be reviewed. If a check of measurement records uncovers a process error, the value will be corrected or invalidated. If such errors are not found, then an annotation will be entered. If the value is invalidated, it will be deleted from the database and replaced by a missing value and flagged appropriately. A record of changes will be permanently retained. Level 2 designation will be assigned after the project data manager and/or data users have performed comparative tests and addressed the quality issues and the project staff have evaluated the test results and supporting QA documents. Authority for Level 2 designation lies at the project level.

Level 3 Validation consists of data that have received intense scrutiny through analysis or use in modeling. As analysis of the data proceeds, analysts may raise questions about portions of the Level 2 data set. Additional checks and tests will be performed on such data and the Level 3 code will be affixed to data passing these tests. If this scrutiny reveals an inconsistency that appears to be caused by a measurement error, the entire chain of evidence for the measurement will be reviewed. This includes reviewing site logs and quality control test data as well as reviewing performance audit results and any other relevant documents. The data users will recommend a Level 3 designation to project staff on the basis of the reevaluations. Alterations to the data validation codes, if warranted, will be made by the project data manager.

3.2 Data Validation Codes for External Data Sets

Selected sets of external data, such as data from the EPA AIRS and National Weather Service, might be included in the PDA. Data validation codes will be assigned to external data sets, or select fragments of external data sets, processed by the QSSC using the same convention described above. For external data sets not evaluated by the QSSC but available through the PDA, users should pay close attention to the data validation procedures and designations used by those networks before using the data for a particular research application.

Table 3-1. NARSTO Data and Research Product Validation, Documentation, and Access Summary

	Data and Research Product Status at Validation Levels 0, 1, 2, and 3.			
	Level 0	Level 1	Level 2	Level 3
Review Status	Raw Data and Un-reviewed Products	QA'd Data and Peer Reviewed Products	Data Analyses Completed and Products Assessed	Continuous Use in Analysis and Modeling
Processing and Reviews Performed by --	Project	Project, QSSC and Subject Matter Experts	Project and Scientific and/or Client Community	Project and Scientific and/or Client Community
Metadata Records	Incomplete	Project and QSSC Checks Documented and Review Comments Resolved	Assessment Issues Documented and Users Guides Available	Applicable Records Reviewed (e.g., site logs and performance audits)
Access	Project and NARSTO Program	Public Release	Public Release	Public Release
Distribution	Within 12 Months of Generation	When Level 1 Reviews are Complete	When Level 2 Reviews are Complete	When Level 3 Reviews are Complete
Source	Project	LaRC DAAC and QSSC	LaRC DAAC and QSSC	LaRC DAAC and QSSC
Format	Project Specific	LaRC DAAC Format and Product Specific	LaRC DAAC Format and Product Specific	LaRC DAAC Format and Product Specific
Change Control Point of Contact	Project	Project and Product Specific	Project and Product Specific	Project and Product Specific

4.0 DOCUMENTING AND TRANSMITTING DATA

APPLICABILITY:	Applicable to transmitting of data from project-to-project, PI-to-PI, and from project to the QSSC and PDA.
USE:	Guidance for the project data coordinators for selected tabular data transfer file formats.[3] Reference to applicable remote sensing, GIS, and model output transfer formats.
CONSIDERATIONS:	Format descriptors are not presented for all possible data types nor record contents. Consult with the QSSC for additional information.
SUMMARY:	Data transmission guidelines are offered for *in situ* NARSTO observations. These guidelines are conducive to automating data processing activities and adoption of these guidelines will ease and standardize data transmission and facilitate data analysis.

4.1 NARSTO Data and Information Categories

NARSTO will generate and use data and information in four major categories: *in situ* observations, remote sensing observations, model outputs, and geographic information system (GIS) coverages. The data transmittal format may be different for each of these categories, and data may be stored locally in a number of ways to maximize efficiency. It is important, however, that data transmitted to other NARSTO participants and the PDA conform to a common set of unit, syntax, and format conventions. Limiting the number of formats will facilitate data and research product sharing and minimize the conversions necessary to load the data into the permanent archive.

4.1.1 ***In Situ* Observations**

These data may be generated from fixed or airborne platforms and may be measurements and analyses of either continuous or discrete sampling methods. These data should be reported to the level of precision of the instrument using the data reporting guidelines given in Sect. 2. Supporting data regarding sites, platforms, instruments, and qualifications are transmitted in a similar format. These data types are the main emphasis of this section.

[3] Major portions of the data transfer section were adopted from the NARSTO-Northeast Data Management Plan, Version 3.1, (May 1996) and the standards adopted by the Canadian National Atmospheric Chemistry Database and Analysis Facility, Version: 1998/01/02 (1.0)

4.1.2 Remote Sensing Observations

Remote sensing observations may be generated at the surface from stationary or vehicle-mounted platforms, from airborne platforms, or from space platforms. They may include long path or range-resolved measurements, soundings, or images from a variety of digital or analog systems. The key factor in this type of data is that the sensor system is not in contact with the parameter or object measured. Due to the wide variety of potential data, it is impractical to prescribe data formats at this time, but NARSTO will follow guidelines and standards set by the Federal Geographic Data Committee (FGDC) for geospatial data as they are set, and will generally follow developing standard interchange formats set by the NASA Earth Observing System Data and Information System (EOSDIS) to facilitate cooperation with the PDA at NASA Langley Research Center. At present, there is no FGDC standard profile for raster data, although one is under development. Currently, raster image data may be transmitted in either band interleaved by line (BIL) or band sequential (BSQ) format, and either must be available uncompressed or available with free software for decompression, if compressed. Header or trailer files or records should be included indicating the processing performed and geocoding parameters, where applicable.

4.1.3 Model Outputs

Models may produce spatial or non-spatial outputs. Standards for interchange of model results are under development at this time and NARSTO will follow these developments including efforts by FGDC to develop standards for the interchange of model results.

4.1.4 GIS Coverages

GIS coverages may be stored and manipulated in a variety of formats. Raster GIS coverages should conform to the same standards as remote sensing image data. Vector GIS coverages must be available for transmission according to the format specified by the FGDC under the spatial data transfer standard (SDTS)[4].

4.2 Recommended Data Transfer Format

NARSTO recommends the use of the following Data Exchange Standard[5] for documenting and transferring files among NARSTO scientists and for submitting files to the QSSC.

All NARSTO participants and data users will benefit from widespread acceptance and use of the Data Exchange Standard. Adoption and use of the Data Exchange Standard will

- assure that NARSTO data contain sufficient descriptions to permit appropriate use by a wide variety of users;
- standardize data variable names, station identifiers, data records, and flags;
- offer flexibility to accommodate a wide variety of measurement data;
- offer a file format that requires minimal effort by NARSTO scientists and data managers to modify existing files or create new files;
- provide data in a format easily handled by spreadsheets and other analysis software;

---

[4] Check the FGDC web site for latest information. The URL is http://www.fgdc.gov/index.html .

[5] NARSTO Data Exchange Standard adapted from Sukloff, B. 1998. The Canadian National Atmospheric Chemistry Database and Analysis Facility (NAtChem) Data Exchange Standard. Version: 1998/01/02 (1.0).

- permit the QSSC to develop, apply, and refine tools to automatically generate quality assurance reports,
  graphs, and statistical data summaries; and
- enable the PDA to potentially develop tools to generate automated summary tables and data catalogs.

Data Exchange Standard files are composed of Key Phrases, metadata tables (e.g., station information tables, lookup tables), and data tables. Key Phrases and tables are described in the sections that follow. The Data Exchange Standard uses ASCII files. Files can be created with various software tools, such as a simple text editor, a spreadsheet, or a computer program.

The QSSC web site[6] will provide guidance on the Data Exchange Standard, along with templates that contain all of the Key Phrases. The templates are in comma separated value (CSV) ASCII files, that can be used directly in a text editor or by spreadsheet software, and in at least one spreadsheet format. Exporting data from spreadsheets into ASCII files can be easily accomplished by using the “save as” file option and selecting the type “CSV” (comma-delimited). The QSSC will maintain a library of programs that interpret the Key Phrases, and read the data into variables defined within a file. Data users are encouraged to help populate this library by sharing programs they have written that work with the Data Exchange Standard.

NARSTO project scientists are required to document their data and are encouraged to submit documentation (i.e., metadata or data about data) that satisfies the "20-year rule". That is, will someone 20 years from now, not familiar with the data or how they were obtained, be able to use the data solely with the aid of the documentation archived with the data set? The key for ensuring use of NARSTO data by someone 20 years after completion of the project lies in the metadata. Metadata provide a context for the measurement data. The Data Exchange Standard has pre-defined metadata tables for station, instrument, and measurement method information.

One of the real attractions of the Data Exchange Standard is that metadata are provided in the same file as the measurement data. Storing metadata in the same file as the data ensures that it will not be lost, or associated with the wrong set of data. In cases where metadata are desired to be stored in a separate file, the external file name can be recorded in a Key Phrase. External metadata files may be used in situations where there are multiple data files sharing the same metadata information, or for metadata that are not in the Data Exchange Standard format. Metadata that are not in the Data Exchange Standard include word-processing and image files. These files are referred to as ancillary files, and are linked to the Data Exchange Standard file using an optional Key Phrase. Greater details about the Data Exchange Standard, Key Phrases, and Metadata Tables are provided in the following sections.

[6] http://cdiac.esd.ornl.gov/programs/NARSTO/

4.3. Data Exchange Standard File Overview

Data Exchange Standard files are composed of Key Phrases, metadata tables, and data tables. Key Phrases contain an asterisk, a phrase in uppercase letters, and a comma[7]. They are followed by a series of one or more data fields separated by commas. Quotation marks must surround any textual fields that contain commas as part of the field. (Note: some spreadsheet software will automatically insert these quotation marks when saving the spreadsheet as a comma-delimited file.) Unless otherwise stated, data fields may contain blanks. Examples of Key Phrases are:

***PRINCIPAL INVESTIGATOR NAME-LAST/FIRST, Smith,Dr. Jane**
| |
last name first name

***PRINCIPAL INVESTIGATOR NAME-LAST/FIRST, "Smith, Jr.",John**
| |
last name (containing a comma) first name

***TABLE COLUMN UNITS,none,none,none,ppbv,none,ppbv, none,m/s,none**

Key Phrases must appear in the files exactly as shown in this document (for an example, see section 4.7 *Sample Data Exchange Standard File*), including any spaces and brackets or other special characters. Data fields are restricted to the characters shown in the Key Phrase ***STANDARD CHARACTERS,**. Note that both upper and lower case characters are allowed in fields. No line breaks may occur within Key Phrase lines or within data lines (although some are needed in this document, to display the contents of long lines). Spaces are permitted within character values in data fields, and between the comma defining the beginning of the field and the entry in the field.

The Data Exchange Standard contains metadata table structures for station information, measurement methods, and lookup tables (*see section 4.5 Metadata Tables*). NARSTO scientists and data managers may create their own additional tables. The Data Exchange Standard allows user-defined tables to be automatically read by a software program. An example of a table, showing only some of the rows and columns in the station information table, is shown below:

***TABLE NAME,** station information
***TABLE COLUMN LONG NAMES,** station ID,station name,country code,state/province code,etc....
...
***TABLE BEGINS,**
,KNP,KEJIMKUJIK,CA,NS, ...
,SUT,SUTTON,CA,QC, ...
***TABLE ENDS,**

The measurement method information table and the measurement data table are similarly organized (see Section 4.5 for an example). **Please note that only one measurement data**

[7] NARSTO scientists and data managers using spreadsheets to create Data Exchange Standard files do not have to place a comma after the Key Phrase. When the file is saved as "CSV" (comma-separated values), a comma will be automatically inserted provided the first field is in the cell following the Key Phrase.

table may be present in a Data Exchange Standard file, but the file may contain as many metadata tables as required. The measurement data table appears last in the file, after the metadata tables.

4.4 Data Exchange Standard Key Phrases

Key Phrases are one of the critical elements of the Data Exchange Standard. Key Phrases contain an asterisk in the first column, a phrase in uppercase letters, a comma, and one or more information fields.

Three types of Key Phrases are identified and should be presented in this order:

Header Key Phrases describe the source of the data, provide a general data description, give information about the data file, and **are provided by the Data Originator**;

Quality Assurance Center (QAC) Key Phrases define the archive file name, contain information about the results of quality assurance checks, and **are inserted by the QA center performing the quality assurance checks (e.g., the QSSC)** during processing; and

Table Key Phrases provide specific descriptive and format information about the metadata or measurement data that immediately follows and **are provided by the Data Originator**.

The following sections list and describe mandatory and optional Key Phrases and provide examples of each Key Phrase. Mandatory Key Phrases provide information necessary to document, process, and archive NARSTO measurement data files and are required for all measurement data files. Mandatory Key Phrases may be supplemented with Optional Key Phrases. Use of Optional Key Phrases is highly encouraged but the extent of their use is up to the discretion of the Data Originator. A summary table of Key Phrases is provided in Appendix C. This table includes the length limit specifications for the information that Data Originators supply for the Key Phrases.

4.4.1 Header Key Phrases

These Key Phrases provide information necessary to document, process, and archive NARSTO measurement data files. Although the order is flexible except where noted, we request that the logical order below be followed. Mandatory Key Phrases are listed first.

***DATA EXCHANGE STANDARD VERSION,** *<version>* [MANDATORY]

This Key Phrase must be the first line in the file. The Data Exchange Standard is being refined as its usage increases over time. This Key Phrase provides the version of the Data Exchange Standard that the Data Originator followed when he/she created the file. Data Exchange Standard templates offered by the QSSC will state the version explicitly. Also, files prepared using this document can define the version based on information in the footer of each page.

Example: *DATA EXCHANGE STANDARD VERSION,NARSTO 1999/07/23 (1.2)

***QUALITY CONTROL LEVEL,** *<code>* [MANDATORY]

This code is normally assigned by the Data Originator to indicate the overall quality of the data file. This code does not reflect the data validation level for individual parameters or data points. The QSSC will assign code "9" in cases where historical data have no information about the quality level.

1: Final data that have been quality assured by the Data Originator. A “1" here would be assigned to a file all the data in which met at least “Level 1 Validation” criteria defined in Section 3.1.
1: Data submitted with no indication of the quality control level.

Example: *QUALITY CONTROL LEVEL, 1

***LAST REVISION TO DATA-DATE/REVISION NUMBER,** *<date>, <revision number>*
[MANDATORY]

This information allows the Data User to confirm that he/she has the latest version of a data file. If the Data Exchange Standard file is created from a file maintained in a different format, such as in an Oracle database, then this Key Phrase represents the date of the last revision to the data in the Oracle database, and not the date that the file was created. Data Originators should update this date whenever values, flags, or validation levels in the original database have changed. They will re-submit the entire data file. Data Originators should include a short explanation of the changes they made to the file by using *COMMENT, Key Phrases. Data Users should check that they have the latest revision whenever they are going to use a data file previously obtained from the PDA. The revision number will become part of the PDA file name.

The suggested date format is yyyy/mm/dd

Example: *LAST REVISION TO DATA-DATE/REVISION NUMBER, 1997/06/01, 1

***ORGANIZATION ACRONYM,** *<name or acronym>* [MANDATORY]

This Key Phrase is used to identify sponsors or large organizational affiliations that would not be reflected by the Principal Investigator's affiliation. This Key Phrase will aid efforts to automate inventory and cataloging efforts by the QSSC and PDA. The acronym may not be longer than 10 characters with no blank spaces, and is restricted to alphabetic characters, numbers, and the dash character.

Example: *ORGANIZATION ACRONYM, USDOE

***STUDY OR NETWORK ACRONYM,** *<acronym>* [MANDATORY]

The QSSC will maintain a table of network and study acronyms, along with descriptive information and contact names. Using a pre-defined acronym allows end-users to link to this information. This Key Phrase will facilitate cataloging, inventory, and searching efforts. The acronym may not be longer than 15 characters with no blank spaces, and is restricted to alphabetic characters, numbers, and the dash character.

Examples: *STUDY OR NETWORK ACRONYM, NARSTO-CE1996
*STUDY OR NETWORK ACRONYM, SCOS97-NARSTO

***FILE CONTENTS DESCRIPTION-SHORT/LONG,** *<short description>, <long description>*

[MANDATORY]

Data Originators will use this Key Phrase to provide further information about the contributed file. The first field may not be longer than 15 characters with no blank spaces, and is restricted to alphabetic characters, numbers, and the dash character. The only restriction on the second field is that it may not be longer than 80 characters. If multiple organizations sample at the same site, this Key Phrase can also be used to distinguish among the organizations.

Examples:
*FILE CONTENTS DESCRIPTION-SHORT/LONG, summary-stats, summary statistics
*FILE CONTENTS DESCRIPTION-SHORT/LONG, acetone, acetone measurements
*FILE CONTENTS DESCRIPTION-SHORT/LONG, daily-max-ozone, daily maximum ozone

***PRINCIPAL INVESTIGATOR NAME-LAST/FIRST,** *<last name>*, *<first name>* [MANDATORY]

The last and first name of the principal investigator or head scientist responsible for the data contained in this file.

Example: *PRINCIPAL INVESTIGATOR NAME-LAST/FIRST, Smith , Dr. Jane

***PRINCIPAL INVESTIGATOR AFFILIATION,** *<affiliation>* [MANDATORY]

The QSSC will maintain a table of principal investigator affiliations to facilitate cataloging, inventory, and searching efforts.

Example: *PRINCIPAL INVESTIGATOR AFFILIATION, “Department of Chemistry, Princeton University”

***COUNTRY CODE,** *<code>* [MANDATORY]

Files are segregated by country. Country codes follow the ISO3166 standard.

Code	***Description***
CA	Canada
MX	Mexico
US	United States of America

Example: *COUNTRY CODE, CA

***STATE OR PROVINCE CODE,** *<code>* [MANDATORY]

Data Originators must provide state or province codes. The state or province code follows each state or province name in the table below.

US States *(Source: Information Please Almanac, Atlas & Yearbook – 1997 Edition)*

Alabama, AL	Alaska, AK	Arizona, AZ	Arkansas, AR
California, CA	Colorado, CO	Connecticut, CT	Delaware, DE
Dist. Of Columbia, DC	Florida, FL	Georgia, GA	Guam, GU
Hawaii, HI	Idaho, ID	Illinois, IL	Indiana, IN

Iowa, IA	Kansas, KS	Kentucky, KY	Louisiana, LA
Maine, ME	Maryland, MD	Massachusetts, MA	Michigan, MI
Mississippi, MS	Missouri, MO	Montana, MT	Nebraska, NE
Nevada, NV	New Hampshire, NH	New Jersey, NJ	New Mexico, NM
New York, NY	North Carolina, NC	North Dakota, ND	Ohio, OH
Oklahoma, OK	Oregon, OR	Pennsylvania, PA	Puerto Rico, PR
Rhode Island, RI	South Carolina, SC	South Dakota, SD	Tennessee, TN
Texas, TX	Utah, UT	Vermont, VT	Virginia, VA
Virgin Island, VI	Washington, WA	West Virginia, WV	Wisconsin, WI
Wyoming, WY			

Canadian Provinces *(Source: Corpus Almanac & Canadian Sourcebook – 1997)*

Alberta, AB	British Columbia, BC	Manitoba, MB	New Brunswick, NB
Newfoundland, NF	Northwest Territories, NT	Nova Scotia, NS	Ontario, ON
Prince Edward Island, PE	Quebec, PQ	Saskatchewan, SK	Yukon Territory, YT

Mexico *(Source: http://inegi.gob.mx/homepara/geografia/extterri/division.html)*

Aguascalientes, AG	Baja California, BC	Baja California Sur, BS	Campeche, CA
Coahuila, CO	Colima, CL	Chiapas, CH	Chihuahua, CU
Distrito Federal, DF	Durango, DG	Guanajuato, GT	Guerrero, GR
Hidalgo, HG	Jalisco, JA	México, ME	Michoacán, MI
Morelos, MO	Nayarit, NA	Nuevo León, NL	Oaxaca, OA
Puebla, PU	Querétaro, QO	Quintana Roo, QR	San Luis Potosí, SL
Sinaloa, SI	Sonora, SO	Tabasco, TA	Tamaulipas, TP
Tlaxcala, TL	Veracruz, VE	Yucatán, YU	Zacatecas, ZA

Example: *STATE OR PROVINCE CODE, NY

***DATA REPORTING INTERVAL-CODE/DESCRIPTION,** *<#><units> , <description>*

[MANDATORY]

This Key Phrase describes the nominal reporting interval of the data in the data table. Each file must contain data for only one reporting interval. The first field contains a number, immediately followed by the unit (with no space between the number and the unit). The allowable units are as follows:

Units	***Description***
sec	Second
min	Minute
hr	Hour

Units	*Description*
day	Day
mon	Month
yr	Year

The first field of this Key Phrase may be used in the file name. The second field in this Key Phrase is used by the Data Originator to more fully describe the data reporting interval.

Examples:
*DATA REPORTING INTERVAL-CODE/DESCRIPTION, 1hr, hourly average
*DATA REPORTING INTERVAL-CODE/DESCRIPTION, 5min, 5 minute average
*DATA REPORTING INTERVAL-CODE/DESCRIPTION, 10min, grab sample every 10 min
*DATA REPORTING INTERVAL-CODE/DESCRIPTION, 1hr, 20 minute sampling per hour
*DATA REPORTING INTERVAL-CODE/DESCRIPTION, 24hr, 24 hour integrated sample
*DATA REPORTING INTERVAL-CODE/DESCRIPTION, 1yr, annual statistical summaries

Use of Optional Key Phrases is highly encouraged but the extent of their use is up to the discretion of the Data Originator.

***COMMENT,** *<text>* [OPTIONAL]

This optional Key Phrase may be used by Data Originators to add comments to the file. Unlike other header Key Phrases, the *COMMENT, key phrase can also be used in tables (see Section 4.4.3).

Example: *COMMENT, The hourly data in this file were averaged from 5-minute average measurements.

***PRINCIPAL INVESTIGATOR CONTACT INFORMATION,** *<text>* [OPTIONAL]

This Key Phrase may be used to provide the address, e-mail address, web site, and any other information on how to contact the Principal Investigator.

Example: *PRINCIPAL INVESTIGATOR CONTACT INFORMATION, “AES, 4905 Dufferin St., Toronto, Ont. M3H 5T4; e-mail: jane.smith@ec.gc.ca"

***STUDY OR NETWORK NAME,** *<text>* [OPTIONAL]

When the data in the data table are obtained from a formal measurement network or study, the full name of the network or study can be specified here. Note that the related Key Phrase ***STUDY OR NETWORK ACRONYM,** is mandatory.

Example: *STUDY OR NETWORK NAME, NARSTO-Canada East 1996 Field Study

***NAME(LAST-FIRST)/AFFILIATION OF PERSON WHO GENERATED THIS FILE,** <last name>, *<first name>, <affiliation>* [OPTIONAL]

This information can be helpful if at some point in the future the need came up to identify all files that were created by a specific person or if the file was generated by someone other than the contributing principal investigator.

Example: *NAME(LAST-FIRST)/AFFILIATION OF PERSON WHO GENERATED THIS FILE, Smith, Jane Dr., ENVCAN

***DATE THIS FILE GENERATED,** *<date>* [OPTIONAL]

This date will usually occur later than the date of the last revision to the data. This information may be useful if it is found that files created during a specified time were not created correctly, or have other potential problems.

Example: *DATE THIS FILE GENERATED, 1997/10/05

***NAME/VERSION OF SOFTWARE USED TO CREATE THIS FILE,** *<name>,<version>* [OPTIONAL]

This information may be useful if it is found that files created with a specific software package and version were not created correctly, or had other potential problems.

Example: *NAME/VERSION OF SOFTWARE USED TO CREATE THIS FILE, Microsoft Excel, 97

***STANDARD CHARACTERS,** *<text>* [OPTIONAL]

This Key Phrase lists the allowable characters, numbers, and symbols that may be included in Data Exchange Standard files. Characters that may not be converted properly from other coding systems such as EBCDIC are not included in this list. The standard characters are in the example shown below.

Example:
STANDARD CHARACTERS, “!"#$%&'(),-+=./0123456789:;<>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ\^_`abcdefghijklmnopqrstuvwxyz{}~”

***ANCILLARY FILE NAME/FORMAT/VERSION,** *<name>, <format>, <version>* [OPTIONAL]

Ancillary files may be used in situations where multiple data files share the same metadata, and these metadata are not required to be in the data file itself. For example, station information, measurement method information, and flag code definitions are required to be contained within each data file so these cannot be placed in ancillary files. File formats can be in the Data Exchange Standard or other formats such as word processing or image formats. Ancillary file names are formed using the fields documented in the Key Phrase
***QAC NARSTO PERMANENT ARCHIVE FILE NAME,**, but they do not have to contain all of the fields. For example, if the ancillary file contains information that applies to an entire study run by a specific organization, the file name should be comprised of the arguments to the Key Phrases ***ORGANIZATION ACRONYM,** and ***STUDY OR NETWORK ACRONYM,**, followed by a text field that describes the content of the file. Multiple Data Exchange Standard files can then refer to this file by referring to it in this Key Phrase. Ancillary files in the Data Exchange Standard must contain all of the mandatory Key Phrases. If a Key Phrase does not apply to the information in the file, the fields appearing after the Key Phrase may be left blank.

Examples (note that the line breaks used here would not be permitted in the actual file):
*ANCILLARY FILE NAME/FORMAT/VERSION,
envcan_narstoce_description_of_study.doc, word, 6
*ANCILLARY FILE NAME/FORMAT/VERSION, envcan_narstoce_station_table.csv,
NARSTO, narsto 1997/11/05 (1.0)
*ANCILLARY FILE NAME/FORMAT/VERSION,
envcan_aes_narstoce_description_of_calibration_procedures.wp,Word, 6

Current length limits are 67 characters for the file name, 30 characters to describe the format, and 30 characters to identify the version. Multiple ancillary file Key Phrases are allowed in a file. External files are placed in directories that will be created by the QSSC for each unique occurrence of organization acronym and study or network acronym. Data Originators must ensure that file names are not duplicated.

In order to prevent metadata from being lost or associated with the wrong data, Data Originators are encouraged to store metadata in the Data Exchange Standard, and in the same file as the data table.

4.4.2 QA Center Key Phrases (Added to File by the QAC)

The Quality Assurance Center(s) (QACs) performing the QA checks on the data files (in most cases the QSSC) will add some Key Phrases to data files when quality assuring the contents. This section describes the suite of QA Center Key Phrases. **These Key Phrases must not be provided by the Data Originator.** The QAC will add a table with scientifically relevant summary statistics (excluding missing values) for the non-character variables in the measurement data table.

***QAC ORGANIZATION NAME,** *<code>,<description>,<date>*

This Key Phrase identifies the QA center(s) performing the QA checks on the data files (e.g., QSSC, NAtChem).

***QAC COMMENT,** *<text>*

This Key Phrase is used by the QA center to add comments about the file and its contents.

Example: *QA COMMENT, QSSC converted the data format from NARSTO-NE.

***QAC QUALITY ASSURANCE CODE/DESCRIPTION,** *<code>,<description>*

A series of quality assurance checks are performed on files in the Data Exchange Standard format. The result of those checks is indicated by a quality assurance code, as follows:

Code	***Description***
V0	Np problems found in file.
V1	Problems were found in file but not of a severe nature.
I1	Severe problems found in file.

A log file containing the results of the quality assurance tests is produced for the collection of files processed. The log file may be posted on the QSSC web site. Data Originators should review the quality assurance report, and if required, fix any errors identified in the report and resubmit their data.

Example: *QA QUALITY ASSURANCE CODE/DESCRIPTION, V0, No problems found in file.

***QAC DATE/TIME QUALITY ASSURANCE PERFORMED,** *<date>, <time>, <time zone>*

This Key Phrase specifies the date and time that the quality assurance tests were run by the QA center.

Example: *QA DATE/TIME QUALITY ASSURANCE PERFORMED, 1997/12/20,10:05:33 EST

***QAC ARCHIVE FILE NAME,**

The file name is automatically generated by the QA center from information contained in the data file. File names provide information about the file contents so that QA Center Staff and Data Users can quickly identify the file contents. File names may contain Key Phrase information and derived fields. This file name may be longer than permitted by the permanent archive, and therefore may not be the file name used there.

File names may include the following information

* organization acronym,
* study or network acronym,
* principal investigator affiliation acronym,
* country code,
* state / province code,

and derived fields

* station or flight identification, (If there is more than one station or flight in the file, this field is used to specify the number of stations or flights contained in the file [e.g., 25stns])
* measurement category,
* file contents description,
* first date in data table, (The date of the first measurement in the data table is recorded in this field. The date format is yyyy/mm/dd, local time)
* last date in data table, (The date of the last measurement in the data table is recorded in this field. The date format is yyyy/mm/dd, local time)
* date of last revision to data,
* data reporting interval,
* quality control level, and
* file type. (ASCII files have the suffix ".csv". Image files have the appropriate suffix, such as "gif" or "jpg". NARSTO measurement campaigns can generate tremendous volumes of data. Files may be compressed into manageable data packages using PKZIP or the UNIX "tar" or "gtar"

compression utilities. The file name for files compressed using PKZIP or "tarred" should have the same prefix, but end with ".zip" or ".tar").

Examples:
*QAC ARCHIVE FILE NAME, knp_nitr_pan_1996.csv
*QAC ARCHIVE FILE NAME, narstoce_ knp_1hr.zip

***QAC DESIGNATION OF PERMANENT ARCHIVE WEB ADDRESS,** *<URL>*

This Key Phrase provides users with the web address of the long-term Permanent Data Archive (PDA).

Example: *QAC DESIGNATION OF PERMANENT ARCHIVE WEB ADDRESS, http://eosweb.larc.nasa.gov/

***QAC NARSTO PERMANENT ARCHIVE FILE NAME,** *<name>*

This Key Phrase provides the name the file will have in the long-term PDA. This name may be required to be shorter than the full ***QAC ARCHIVE FILE NAME,**.

***QAC DATA ORIGINATOR FILE NAME,** *<name>*

This Key Phrase provides the name that the Data Originator used when the file was submitted to the QA center.

Example:
*QAC DATA ORIGINATOR FILE NAME, aes-ozone.csv

4.4.3 Table Key Phrases

Table Key Phrases are important because they connect appropriate descriptive information to the station and instrument metadata or measurement data that follow. The Header Key Phrase "*COMMENT," may also be used with the Table Key Phrases shown below.

***TABLE NAME,** *<name>* [MANDATORY]

Table names are comprised of any combination of characters and numbers up to 80 characters in length. The automatic data loader uses the first 8 characters in a table to form the data set name. They must be unique across all table names that are defined or referred to in the file.

Example: *TABLE NAME, instrument information

***TABLE FOCUS,** *<focus>* [MANDATORY]

Each table is identified with a code that describes its primary focus. Additional table focus codes may be added as needed. The focus names are as follows:

Focus	Description
Surface	Measurement made at or near surface level, including tower measurements.
Aloft	Atmospheric measurements such as aircraft, tethersonde, balloon, rawinsonde, and radar profiler.
Meta	Metadata such as station information, instrument information, and measurement methods.

Example: *TABLE FOCUS, surface

***TABLE COLUMN LONG NAMES,** *<field>, <field>, <field>, …* [MANDATORY]

Long names may be up to 25 characters in length. Any combination of characters, numbers, and spaces may be used to form a descriptive name for a column.

Example: *TABLE COLUMN LONG NAMES, detection limit, precision, accuracy

***TABLE COLUMN SHORT NAMES,** *<field>, <field>, <field>, …* [MANDATORY]

Short names must not be longer than 6 characters in length except for flag fields, which are two characters longer than the variable they tie to. They provide a unique identifier within a table that can be used as variable names in a program. The short name must start with an alphabetic character (A - Z), followed by up to 5 alphabetic or numerical characters. The underscore is the only non-alphabetic character allowed. Spaces are not allowed in the name. The short names for flag fields must be exactly the same as the associated measurement field short name, except suffixed by the underscore character, followed by the letter "f" (e.g., o3, o3_f).

Example: *TABLE COLUMN SHORT NAMES, stnid, o3, o3_f, pan, pan_f

***TABLE COLUMN TYPES,** *<code>, <code>, <code>, …* [MANDATORY]

The variable type for each column is indicated by a code, as follows:

Code	***Description***
c	character
i	integer
f	floating point (numbers with a decimal place)
d	date
t	time

Example: *TABLE COLUMN TYPES, c, c, i, i, d, t, d, t, f, c, f, c, f,c

The data for the variables must correspond to the specified type. Specifically, character data may not occur in dates, integers, or floating point fields (other than a plus or minus sign).

***TABLE COLUMN FORMATS,** *<format>, <format>, <format>, …* [MANDATORY]

For each column in the table, this Key Phrase specifies the format for the column. The way the format is specified for a column depends on the *TABLE COLUMN TYPE, of that column. The format must be large enough to accommodate any data to which it applies. Suggested formats are provided for most variables and variable types (see Section 2.2 and Apendix A). Formats may be larger than actually represented in the data, but it is desirable for them to match the largest data that occur in the column. In particular, if the suggested reporting format were 7.2 for a variable, but all of the data for the variable have values to only one decimal place, a format of 7.1 (or 6.1 or smaller, as appropriate) is preferable. Note also that to accommodate missing value codes (see Section 2.7 and also the *TABLE COLUMN MISSING CODES, Key Phrase, below), you may need to use formats specifying a wider field width than the recommended reporting formats shown in Section 2.2, or wider than needed to accommodate the nonmissing data values.

For character data, the format specifies the maximum number of characters (e.g. 20 or 150).

For numerical data (floating point, integer, date, or time) the maximum field width is 10.

For floating point numbers (e.g., 734.28; 3.1; -0.3), the format specifies the maximum field width (up to 10) including any sign, and the number of digits to the right of the decimal point. In other words, a format of 7.2 is just large enough to accommodate values of 3273.22 or -999.99; the field widths in both cases are 7, and each has 2 digits after the decimal point.

For integer numbers (e.g., 53), the format specifies the maximum field width (up to 10) including any sign.

For dates, the format specifies the number of characters forming the date (normally 10 for yyyy/mm/dd). Note that some spreadsheets have other default date formats (e.g., Microsoft Excel 97 uses mm/dd/yy); these must be overridden (see Appendix D).
For time, the format specifies the number of digits forming the time (normally 8 for hh:mm:ss).

Example: *TABLE COLUMN FORMATS, 3,10.5, 10.5, 10,10,3,7.2, 2, 6.1, 2, 60

***TABLE COLUMN UNITS,** *<unit>, <unit>, <unit>, <unit>,…* [MANDATORY]

This Key Phrase describes the measurement units for each column in the table. If no units are applicable to a column, specify ‘none’.

Example:
*TABLE COLUMN UNITS,
none,none,yyyy/mm/dd,hh:mm:ss,yyyy/mm/dd,hh:mm:ss,EDT,yyyy/mm/dd,hh:mm:ss,yyyy/mm/dd,hh:mm:ss,UTC,ppbv,none,ppbv,none,pppv,none,none

***TABLE BEGINS,** [MANDATORY]

This Key Phrase indicates that the table data records will appear starting in the next record. No other Key Phrases may appear until the *TABLE ENDS, Key Phrase has occurred. Also, there may be no blank lines, only data lines, between the *TABLE BEGINS, and the *TABLE ENDS, Key Phrases.

Example: *TABLE BEGINS,
<table data>

***TABLE ENDS,** [MANDATORY]

This Key Phrase indicates that the table data records have finished, and it defines the end of the table. The *TABLE ENDS, Key Phrase for the measurements data table must be the last line in the file.

Example:
<table data>
*TABLE ENDS,

***TABLE KEY FIELD NAMES,** *<field>, <field>, <field>, …* [OPTIONAL]

This Key Phrase specifies the column short names that together make a record in the measurement data table unique. Usually, records are uniquely identified by the columns *stnid, sdatel, stimel*. Additional columns may be needed, for example, where two instruments took identical observations at a site. The column name *instid* makes each record unique. This Key Phrase is necessary to allow the QSSC to correctly identify duplicate measurement data records. It is recommended that it be included in the measurement table in all cases.

Example: *TABLE KEY FIELD NAMES, stnid, sdatel, stimel, instid

***TABLE NUMBER OF COLUMNS,** *<#>* [OPTIONAL]

This Key Phrase specifies the number of columns, excluding the column with Key Phrases but including flags, in the data table. This is helpful to Data Users who are writing programs to automatically load the data.

Example: *TABLE NUMBER OF COLUMNS, 10

***TABLE NUMBER OF ROWS,** *<#>* [OPTIONAL]

This Key Phrase specifies the number of rows of data in the data table. This is helpful to Data Users for verifying that all of the data have been read by another application.

Example: *TABLE NUMBER OF ROWS, 1000

***TABLE COLUMN MISSING CODES,** *<code>, <code>, <code>, …* [OPTIONAL]

This Key Phrase specifies the code that is used to represent a missing value. It is mandatory if any missing values are included. The missing code for character variables will normally be specified as ‘none’. (If a character variable is missing, it can be represented in the data as a space between the two surrounding commas, or the two surrounding commas can be next to each other without a space).

Section 2.7 provides additional information and advice about specifying missing values. Integer and floating point data should specify a missing value code that is unambiguous and easily recognized. We recommend using a negative sign followed by a series of 9's in the proper format (with decimal point if appropriate) to fill out the full width of the

format specification. Because small negative values may be valid values for many parameters and measurement techniques, negative missing value codes should be chosen to be large enough to be genuinely unambiguous. For example, -99.9 could be specified as a missing value code for temperature. In this case,to accommodate the minus sign, the format for temperature would be specified as 5.1, even if 4.1 were large enough to accommodate the non-missing (measured) values. If this Key Phrase is used, missing value codes must be defined for all numeric variables, whether or not such missing values actually occur in the data for a particular variable.

Date and time values and spatial coordinates (latitude and longitude specifications) are a special situation. These should never be missing. In the *TABLE COLUMN MISSING CODES, Key Phrase, the missing value code for dates or times must be specified as a dot (i.e., a period). If a date or time value must be reported as missing in a data row, it would be assigned a dot as its value. The missing value codes for latitude/longitude values when expressed in decimal degrees (latdec and londec) should be large enough to be invalid values (e.g., -999.99999). If degrees/minutes/seconds units are used for latitude and longitude, the missing value code for degrees should be -999; for minutes and seconds, it should be -99.

Example: *TABLE COLUMN MISSING CODES, none, none,.,.,.,.,none,.,.,.,.,none, -999.99999, -999.99999, -99.99, none

***TABLE COLUMN MINIMUM INCLUDING MISSING VALUES,** *<code>, <code>, <code>, …*

[OPTIONAL]

This Key Phrase in the measurement data table identifies, for each non-character variable, the minimum value that occurs in the data column. This information is helpful to confirm correct file transfer. Specify ‘none’ for character variables. Any numeric missing value codes must be included in the calculation.

Example: *TABLE COLUMN MINIMUM INCLUDING MISSING VALUES, none, none, none, none, -99.99, none, 0.42, none, -99.99, none, 1.58, none

***TABLE COLUMN MAXIMUM INCLUDING MISSING VALUES,** *<code>, <code>, <code>, …*

[OPTIONAL]

This Key Phrase in the measurement data table identifies, for each non-character variable, the maximum value that occurs in the data column. This information is helpful to confirm correct file transfer. Specify ‘none’ for character variables. Any numeric missing value codes must be included in the calculation; this matter here only if all values are missing.

Example: *TABLE COLUMN MAXIMUM INCLUDING MISSING VALUES, none, none, none, none, 328.25, none, 79.13, none, 261.5, none, 17.82, none

***TABLE COLUMN START/END POSITIONS,** *<start column/end column>,<start column/end column>,<start column/end column>,…* [OPTIONAL]

Data Originators writing computer programs to produce Data Exchange Standard files may want to create fixed column tables. This key field specifies the start and end positions for each of the columns in a table, not including the comma separators. Commas are still required as a separator for each column.

Example: *TABLE COLUMN START/END POSITIONS, 1-5, 7-11, 13-18, 20-26, 28-28, 30-32

***TABLE MAXIMUM RECORD SIZE,** *<#>* [OPTIONAL]

This Key Phrase specifies the maximum number of characters appearing in any record in this table. This information will help Data Users, who cannot read long records (e.g., over 255 characters), to catch truncation errors.

Example: *TABLE MAXIMUM RECORD SIZE, 453

***TABLE COLUMN LOOKUP TABLE NAMES,** *<name>, <name>, <name>, …* [OPTIONAL]

This Key Phrase specifies for most character variable columns in the measurement data table, the name of a table with special columns, known as a *lookup table.* Lookup tables are the same as any other table, except they have two special columns: code and description. The lookup table has an entry for each unique occurrence of a value in the main table. The first 8 characters in a table name are used by the automatic data loader to form the data set name**. They must be unique across all tables names that are defined or referred to in the file.** Lookup tables must begin with the two short variable names *code* and *descript*. (see "Lookup Tables" in Section 4.5.3). Exception: The “station information” table must contain the short character variable name *stnid*, corresponding to the identifier in the measurement data table and need not contain *code* or *descript*. Some lookup tables (e.g., for variable flags) do not need additional columns, while other will need many additional columns.

The measurement data table’s (Section 4.6) *TABLE COLUMN LOOKUP TABLE NAMES, key phrase must point to a lookup table called “station information” for the character variable *stnid*. In addition, the key phrase must point to one or more tables defining the meaning of variable flags. Flag fields that follow the Data Exchange Standard Protocols (see 3.1 *Data and Research Product Validation*) must be identified with the lookup table name "standard flags". Flag variables for which any flag code value has a different definition than the standard must be identified with some other lookup table name. If there is no lookup table for a column, specify ‘none’. We recommend that a table column lookup table be provided for any other character variables in the main data table, except for standard time zone codes and comments, and arbitrary variables such as cannister ID.

Measurement method information may or may not be presented in the measurement data table (see “Measurement Method Information Table, below). If so, it may be referenced to a lookup table called “measurement method information” (which would use the standard lookup table syntax), or alternatively the value ‘none’ may be specified. Which is appropriate will depend on how the measurement method information varies among stations and variables; if the information does not fit into the lookup table format, specify ‘none’. A “measurement method information” table must be provided, as indicated below, whether or not this table is designated as a lookup table.

Example: *TABLE COLUMN LOOKUP TABLE NAMES, station information, measurement method information, none, none, none, none, none, none, none, none, none, none, standard flags, none, nonstandard flags,none

4.5 Metadata Tables

The Data Exchange Standard is designed to be flexible enough to handle all types of metadata. Basic sampling and analysis information can easily be included in the metadata tables described below. Other related information, graphics, or images can be incorporated into referenced ancillary files.

4.5.1 Station Information Table

The station information table is a special lookup table which must be named “station information”. This special lookup table need not contain variables with short name of *code* or *descript*. Some specific requirements exist for this table, because automated codes at the QA centers must extract station IDs and other information. The station information table must contain the following two variables, using these names; the variables must each be defined in the *TABLE COLUMN TYPE, key phrase as being of type c (for character):

Variable	***Notes***
stnid	This is the station identifier which will be provided with each row in the measurement data table (containing the data themselves). A particular stnid must identify the same unique station among all files to be sent to the Archive as a single collection (dataset). (A single dataset would typically comprise, for example, all surface measurement files from a large regional study in a year.) Within a single file, each stnid can occur in only one row in the station information table.
loctyp	Station type with respect to geographic location (latitude-longitude). Values must be one of the following, or contact the QA center if a new value is needed: **fixed**: Station has one set of non-varying latitude-longitude coordinates. These coordinates must be included in the station information file (see below), using the standard coordinate names (e.g., latdec, londec). **mobile**: Station’s latitude-longitude coordinates may vary (e.g., non-tethered sonde; aircraft). Provide the coordinates with each row in the measurement data table, using the standard coordinate names. Do not provide coordinates in the station information file using the standard names. If initial (e.g., release) coordinates are fixed and you want to provide them in the station information table, use different names (e.g., initlat, initlon).

If the station is mobile (not fixed geographically), a comment field should be used to describe the type of platform (sonde, aircraft, etc.).

Location information must be provided for fixed stations. We recommend that this be provided as decimal degrees, in which case the short names must be *latdec* and *londec*. As an alternative, location can be provided as degrees-minutes-seconds. In this case, the short names must be *latdeg*, *latmin*, *latsec*, *londeg*, *lonmin*, and *lonsec*.

Information about the elevation at which samples are collected should be provided in the station information table. The format of this information is up to the data provider, but it should

enable users to determine both height above ground level and altitude above mean sea level. For loctype=fixed stations where samples are collected at a fixed height, we recommend that a variable be defined for sample height above ground level and another for sample altitude above mean sea level. For fixed stations where samples are collected at varying heights (e.g., a tower, or a tethered sonde), we recommend that a variable in the station information table be used to provide the ground altitude above mean sea level, a character variable be used to describe sample height as “varying, see measurement data table”, and a sample height variable for each row in the measurement data table be used to provide the sample height above ground level. (In this case, the *TABLE KEY FIELD NAMES, key phrase would be used to indicate the short names making rows in the measurement data table unique, and these short names would include the variable giving the height of the sample.) For mobile stations, sample elevation normally varies. Treatment of elevation information should be as indicated above for fixed stations where samples are collected at varying heights, except (1) ground elevation above mean sea level should not be included in the station information file (because it varies), and (2) the information with each row in the measurement data table should include, if possible, both the sample elevation above ground level and the sample elevation above mean sea level.

Additional information in the sample information table should include country code, state/province code, and if known, the station start date and end date. If station start and end dates are not known, use dates for the earliest and the latest samples in the file. For mobile platforms, location information is provided in each row of the measurement data table, using the same short names specified above for latitude and longitude.

This example of the station information table is designed for point source measurements (i.e., loctype=fixed).

*TABLE NAME, station information
*TABLE FOCUS, meta
*TABLE NUMBER OF COLUMNS, 12
*TABLE COLUMN LONG NAMES, station ID, station name, station location type, country code, state/province code, latitude, longitude, elevation, height, start date, end date, comment
*TABLE COLUMN SHORT NAMES, stnid, stname, loctyp, cntry, st_pr, latdec, londec, alt, height, stnstr, stnend, cmnt
*TABLE COLUMN UNITS, none, none, none, none, none, decimal degrees, decimal degrees, m asl, m agl, yyyy/mm/dd, yyyy/mm/dd, none
*TABLE COLUMN TYPES, c, c, c, c, c, f, f, f, f, d, d, c
*TABLE COLUMN FORMATS, 6, 25, 6, 2, 2, 11.6, 11.6, 8.1, 7.2, 10, 10, 60
*TABLE BEGINS,
<table data ...>
*TABLE ENDS,

4.5.2 Measurement Method Information Table

This table provides information about measurement methods. The way measurement method information should be presented depends on several factors, including the nature of the stations and whether a single instrument or more than one instruments are used. What is important is that users of the data be able to determine how each value in the measurement data table was obtained, and as many details about the method as are practical to obtain. In the simplest case, a single measurement method/instrument is used at each station in the file to measure all values. The measurement table can then contain a single data row, explaining the method. The first variable, measurement name, should also be included with each row of

the measurement data table. The list below, and the example following it, present some of the fields that may be included in the measurement method information table. As a minimum, we suggest including instrument name, measurement method, detection limits, and standard temperature and pressure conditions (where applicable.)

Metadata table for description of measurement method information - suggested fields:
Measurement name;
Instrument name (e.g., Campbell Scientific Setra 270, Teco Model 49);
Measurement method (e.g., UV photometric, gas filter correlation spectrometer, canister samples);
Calibration / zero frequency (e.g., flow controller cal before and after study, or 5 minutes every hour);
Instrument sample integration time (e.g., continuous - 5 minute means, “integrated 9-13h, 13-17h, 17-21h, 21-9h”);
Reported sample integration time (if different from the instrument sample integration time)
Sample inlet height (e.g., 2 m, 10 m);
Sample line length (e.g., 5 m);
Principal investigator’s name (e.g., Smith, Dr. John);
Conversion factors;
Field blank correction procedure; (e.g., blank corrected or not blank corrected)
Standard Temperature and Pressure (STP) condition;
Detection Limit;
Precision;
Accuracy;
Completeness (proportion of expected samples that were collected and are valid).

Example Table:

*TABLE NAME, measurement method information
*TABLE FOCUS, meta
*TABLE NUMBER OF COLUMNS, 16
*TABLE COLUMN LONG NAMES, measurement name, instrument name, measurement method, calibrations/zeros frequency, instrument sample integration time, reported sample integration time, sample inlet height, sample line length, PI name, conversion factors, field blank corrected, STP condition, detection limit, precision, accuracy, completeness
*TABLE COLUMN SHORT NAMES, mstnm, instnm, mstmth, calfrq, insint, rptint, inlthg, lineln, piname, convfc, fblnkc, stp, detlim, precn, accury, cmplt
*TABLE COLUMN UNITS, none, none, none, none, none, none, m, m, none, none, yes/no, none, none, none, none, none
*TABLE COLUMN TYPES, c,c,c,c,c,c,f,f,c,c,c,c,c,c,c,c
*TABLE COLUMN FORMATS, 15,20,30,30,30,30,6.2,6.2,30,30,30,30,30,30,30,30
*TABLE BEGINS,
<table data>
*TABLE ENDS,

Note that in this example, detection limit, precision, and accuracy are specified as character variables. This is appropriate where units of these could vary with measurement name; the units can be included with the values. In addition, this allows designation of “+-“ (plus or minus). If it always had the same units, detection limit could be specified as type f, with the proper units named in the *TABLE COLUMN UNITS, Key Phrase.

In a more complicated case, a station could be a tower, with different instruments at different heights on the tower. Each row in the measurement data table might key to a station ID and a height. In this case, a measurement name should be provided with each data row. The measurement method information table's data rows could include station, height, measurement name, and the other relevant information.

In an unusually complicated case, various measurement methods might be used for the variables in a single row; and some of the methods could change through time. In this case, to avoid greatly increasing the size of the measurement data table, we recommend that no measurement method information be provided there. Instead, detailed information should be provided in the measurement method information table. Each row would contain a station ID, the method start date, the method end date, the variable, and then the rest of the information indicated above (as available). If data on six variables were presented for three stations, each of which changed methods during the period covered by data in the file (i.e., two methods per station), there would be 6x3x2=36 data rows in the measurement method information table. However, this would eliminate the need for a measurement reference variable to accompany each variable in the measurement data table, making that potentially very wide and long table much more manageable.

4.5.3 Lookup Table

Codes appearing in data tables should be cross-referenced with descriptions in a lookup table. Codes that do not have an entry in a lookup table will be identified as needing additional information in the quality assurance report produced by the QSSC.

*TABLE NAME, <a name entered in a variable definition's "lookup table" parameter in the data information table>
*TABLE FOCUS, Meta
*TABLE COLUMN LONG NAMES, code,description, <optionally other variables as required by the Data Originator>
*TABLE COLUMN SHORT NAMES, code,desc, <optional variables>
*TABLE COLUMN FORMATS, <# of characters in code field>, <# of characters in description field>, ...
*TABLE COLUMN TYPES, c, c, ...
*TABLE BEGINS,
*UNITS, none, none, ...
<table data ...>
*TABLE ENDS,

4.6 Measurement Data Table

The measurement data table must be the last table in the file. Besides presenting the actual data, it must provide information about the when and where of the collection or measurement of each value, define any missing values for non-character variables, and provide references to definitions of standard and/or nonstandard flags as well as any project specific quality codes (see ***TABLE COLUMN LOOKUP TABLE NAMES,**).

Only one measurement data table can occur in a file. The first variable in the measurement data table must be a character-type variable (e.g., station ID, but not a date, height, result, etc.). This is because the QA center's added table with scientific statistics will insert the name of the statistic in this position, so the statistics themselves cannot be calculated for it. We

recommend that the required character variable, *stnid*, be used as the first variable in the measurement data table. As discussed earlier, *stnid* must be associated with the lookup table named ‘station information’.

A character variable, *valcod*, must also be provided. This variable provides a code for each data line (0, 1, 2, or 3; see Section 3.1) representing the minimum validation level that applies for any variable in the line. A lookup table named ‘validation levels’ must be provided, with descriptions of the levels. Valid, nonmissing starting and ending date and time values in the table must be provided in both local and UTC coordinates (see Section 2.6). Specifically, the variables sdatel, stimel, edatel, etimel, tzonel, sdateu, stimeu, edateu, etimeu, and tzoneu must be provided, using these short names, for, respectively, local start date and time, local end date and time, and local time zone, with corresponding UTC start date and time, UTC end date and time, and specification that these are for the UTC time zone (tzoneu will always have the value ‘UTC’).

Any data for laboratory standards, spikes, blanks, duplicates for QC purposes, etc. do not go in the measurement data table. Rather, if these data are to be archived, they would go into an ancillary file. If co-located samples were collected with comparison of results a main purpose, these samples could be provided in the measurement data table either with a different station identifier or with some other appropriate identifier (which would then included in *KEY FIELD NAMES,). If co-located samples were incidental, they should be provided in a metadata table constructed to hold them, including appropriate comments.

Example:
...
***TABLE NUMBER OF COLUMNS,** 13
***TABLE COLUMN MISSING CODES,** none,none,none,none,none, none,none, none, none, none, none,-99.0,none
***TABLE COLUMN TYPES,** c,d,t,d,t,c,d,t,d,t,c,f,c
***TABLE COLUMN FORMATS,** 3,10,8,10,8,3,10,8,10,8,3,6.1,2
***TABLE COLUMN LONG NAMES,** station ID, local start date, local start time, local end date, local end time, local time zone code, UTC start date, UTC start time, UTC end date, UTC end time, UTC code, mean visibility, mean visibility flag
***TABLE COLUMN SHORT NAMES,** stnid, sdatel, stimel, edatel, etimel, tzonel, sdateu, stimeu, edateu, etimeu, tzoneu, mvis, mvsflg
***TABLE COLUMN UNITS,** none, yyyy/mm/dd, hh:mm:ss, yyyy/mm/dd, hh:mm:ss, none, yyyy/mm/dd,hh:mm:ss, yyyy-mm-dd, hh:mm:ss, none, km, none
***TABLE KEY FIELD NAMES,** stnid, sdatel, stimel

4.7 Sample Data Exchange Standard File

This is a fictional example of a Data Exchange Standard file containing surface measurements (i.e., for the SURFACE focus). The file also contains metadata tables for station information, lookup codes for flags, measurement method information, and excerpts from the table for tower meteorology measurements. (Note that an actual file would frequently contain many hundreds or thousands of lines of measurements data.) Key Phrases are bolded, contributed data and information are not. This example shows how data can be presented where several different methods are used to collect data at different heights at each of multiple (two) stations. This example and a blank template are available from the QSSC. Refer to Appendix D for additional useful file creation information. Note that to display all information, some lines here are wrapped; line breaks do not occur within lines in the actual file. Extra trailing commas created by Excel when saving as a .csv file have been removed from this example, but such extra trailing commas are acceptable in an actual Data Exchange Standard file.

***DATA EXCHANGE STANDARD VERSION,**NARSTO 1999/07/23 (1.2)
***COMMENT,**" Fictitious example of a file conforming to the Data Exchange Standard. Organizations, studies, values, etc. are fictitious."
***QUALITY CONTROL LEVEL,**1
***LAST REVISION TO DATA-DATE/REVISION NUMBER,**1998/09/25,1
***ORGANIZATION ACRONYM,**MMMF
***STUDY OR NETWORK ACRONYM,**NARSTO-MENG98
***FILE CONTENTS DESCRIPTION-SHORT/LONG,**tower-met,"Tower meteorology measurements in West Sondeville, MA (fictitious)"
***PRINCIPAL INVESTIGATOR NAME-LAST/FIRST,**"Smith, Jr.",Dr. John
***PRINCIPAL INVESTIGATOR AFFILIATION,**"Meteorology Department, University of West Sondeville"
***COUNTRY CODE,**US
***STATE OR PROVINCE CODE,**MA
***DATA REPORTING INTERVAL-CODE/DESCRIPTION,**1hr,hourly averages
***COMMENT,** NARSTO-MENG98 Web address: http://www.narsto-meng98.org (fictitious)
***PRINCIPAL INVESTIGATOR CONTACT INFORMATION,**"Meteorology Department, University of West Sondeville, 2000 University Drive, West Sondeville, MA, 01006, jgsmith@fictitious.org, http://www.uwsondeville.edu/staff/smithjg"
***STUDY OR NETWORK NAME,**NARSTO-Middle New England 1998 study (fictitious)
***NAME(LAST/FIRST)/AFFILIATION OF PERSON WHO GENERATED THIS FILE,**Christensen,Sigurd W.,"Oak Ridge National Laboratory, Oak Ridge, TN 37831-6036, swc@ornl.gov"
***DATE THIS FILE GENERATED,**1999/08/15
***NAME/VERSION OF SOFTWARE USED TO CREATE THIS FILE,**Microsoft Excel,97

***STANDARD CHARACTERS,**"!#$%&'()*,-./0123456789:;<>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ\^_`abcdefghijklmnopqrstuv wxyz{}~"
***ANCILLARY FILE NAME/FORMAT/VERSION,**meng98_description_of_study.wpd,WordPerfect,8
***ANCILLARY FILE NAME/FORMAT/VERSION,**meng98_data_validation_level_and_status_discussion.wpd,WordPerfect,8

***TABLE NAME,**station information
***TABLE FOCUS,**meta
***TABLE NUMBER OF COLUMNS,**12
***TABLE COLUMN TYPES,**c,c,c,c,c,f,f,c,f,d,d,c
***TABLE COLUMN MISSING CODES,**none,none,none,none,none,-999.99999,-999.99999,none,-99.99,.,.,none
***TABLE COLUMN FORMATS,**6,50,6,2,2,10.5,10.5,60,6.1,10,10,120
***TABLE COLUMN LONG NAMES,**station ID,station name,station location type,country code,state/province code,latitude, longitude, sample height,station ground altitude,start date,end date,comment
***TABLE COLUMN SHORT NAMES,**stnid,stname,loctyp,cntry,st_pr,latdec,londec,height,stnalt,stnstr,stnend,cmnt
***TABLE COLUMN UNITS,**none,none,none,none,none,decimal degrees,decimal degrees,none,m asl,yyyy/mm/dd,yyyy/mm/dd,none
***TABLE BEGINS,**
,NE1031,"Meteorology Bldg., north tower, U. West Sondeville",fixed,US,MA,42.43026,-72.60138,"varying, see tower meteorology measurements",1105.4,1996/02/01,1998/09/25,Station continued beyond stnend date. Latitude and longitude accurate to within about ten feet.
,NE1035,"Denton Airport met tower, West Sondeville",fixed,US,MA,42.44197,-72.62358,"varying, see tower meteorology measurements",1154.1,1998/06/01,1998/06/30,Station may exist outside of date range. Latitude and longitude accurate to within about ten feet.
***TABLE ENDS,**

***TABLE NAME,**standard flags
***TABLE FOCUS,**meta
***TABLE NUMBER OF COLUMNS,**2
***TABLE COLUMN FORMATS,**2,120
***TABLE COLUMN TYPES,**c,c
***TABLE COLUMN LONG NAMES,**Code,Description
***TABLE COLUMN SHORT NAMES,**code,desc
***TABLE COLUMN UNITS,**none,none
***TABLE BEGINS,**
,V0,Valid value
,V1,Valid value but comprised wholly or partially of below-MDL data

,V2,Valid estimated value
,V3,Valid interpolated value
,V4,Valid value despite failing some statistical outlier tests
,V5,"Valid value but qualified because of possible contamination (e.g., pollution source, laboratory contamination source)"
,V6,"Valid value but qualified due to non-standard sampling conditions (e.g., instrument malfunction, sample handling)"
,M1,Missing value because no value is available
,M2,Missing value because invalidated by Data Originator
,H1,Historical data that have not been assessed or validated
***TABLE ENDS,**

***TABLE NAME,**validation levels
***TABLE FOCUS,**meta
***TABLE NUMBER OF COLUMNS,**2
***TABLE COLUMN FORMATS,**1,120
***TABLE COLUMN TYPES,**c,c
***TABLE COLUMN LONG NAMES,**Code,Description
***TABLE COLUMN SHORT NAMES,**code,desc
***TABLE COLUMN UNITS,**none,none
***TABLE BEGINS,**
,0,"Data are of unspecified quality, with minimum processing and incomplete audits/reviews."
,1,Data have been subjected to QA and QC checks and data management procedures.
,2,Data have been subjected to interpretive and diagnostic analyses.
,3,Data have received intense scrutiny through analysis or use in modeling.
***TABLE ENDS,**

***TABLE NAME,**measurement method information
***TABLE FOCUS,**meta
***TABLE NUMBER OF COLUMNS,**10
***TABLE COLUMN TYPES,**c,c,c,c,c,c,c,c,c,c
***TABLE COLUMN MISSING CODES,**none,none,none,none,none,none,none,none,none,none
***TABLE COLUMN FORMATS,**30,30,90,10,15,6,20,60,15,60
***TABLE COLUMN LONG NAMES,**measurement name,measurement method,instrument name,calibrations/zeros frequency,instrument sample integration time,reported sample integration time ,precision,accuracy,completeness,comment
***TABLE COLUMN SHORT NAMES,**mstnm,mstmth,instnm,calfrq,insttm,rptint,precn,accury,complt,cmnt
***TABLE COLUMN UNITS,**none,none,none,none,none,none,none,none,none,none

***TABLE BEGINS,**
,mean temperature,thermistor,Campbell Scientific Model HMP35C Temperature and Relative Humidity Probe,weekly, instantaneous,1 hr,+-0.01 degrees C,+-0.4 degrees C,unknown,values are to be considered fictitious
,mean relative humidity,Vaisala RH,Campbell Scientific Model HMP35C Temperature and Relative Humidity Probe, weekly, instantaneous,1 hr,+-0.1% RH,"+-2.06% RH (0% to 90% RH), +-3.04% RH (90% to 100% RH)",unknown,values are to be considered fictitious
,wind speed,propellor anemometer,RM Young 05103 Wind Monitor,weekly,2 min,1 hr,+-0.01 m/s,+-1% for reported wind speeds from 2.5 to 30.0 m/s,78%,values are to be considered fictitious
,wind direction,wind vane,RM Young 05103 Wind Monitor,weekly,2 min,1 hr,+-0.1 degree,+-5 degrees,77%,values are to be considered fictitious
***TABLE ENDS,**

***TABLE NAME,**tower meteorology measurements
***TABLE FOCUS,**surface
***COMMENT,**This is a fictitious example. Do not rely on any of these values.
***TABLE KEY FIELD NAMES,**stnid,sdatel,stimel,height
***TABLE NUMBER OF COLUMNS,**25
***TABLE COLUMN LOOKUP TABLE NAMES,**station information,validation levels,none,none,none,none,none, none,none,none, none,none,none,none,standard flags,none,standard flags,none,standard flags,none,standard flags,none,standard flags, none, standard flags
***TABLE COLUMN TYPES,**c,c,d,t,d,t,c,d,t,d,t,c,f,f,c,f,c,f,c,f,c,f,c,f,c
***TABLE COLUMN MISSING CODES,**none,none,.,.,.,.,none,.,.,.,.,none,-9.9,-99.9, none,-9.9,none,-9.9,none, -99.9,none,-9.9,none, -9.9,none
***TABLE COLUMN FORMATS,**6,1,10,8,10,8,3,10,8,10,8,3,4.1,5.1,2,4.1,2,4.1,2,5.1,2,4.1,2,4.1,2
***TABLE COLUMN LONG NAMES,**station ID,validation level code,start date (local time),start time (local time),end date (local time),end time (local time),time zone (local),start date (UTC),start time (UTC),end date (UTC),end time (UTC),time zone (UTC),height on tower,temperature,temperature flag,relative humidity,relative humidity flag,wind speed,wind speed flag,wind direction,wind direction flag,maximum wind speed,maximum wind speed flag,minimum wind speed,minimum wind speed flag
***TABLE COLUMN SHORT NAMES,**stnid,valcod,sdatel,stimel,edatel,etimel,tzonel,sdateu,stimeu, edateu,etimeu,tzoneu, height,temp,temp_f, relhum,relhum_f, winspd,winspd_f,windir,windir_f,wnspmx,wnspmx_f,wnspmn,wnspmn_f
***TABLE COLUMN UNITS,**none,none,yyyy/mm/dd,hh:mm:ss,yyyy/mm/dd,hh:mm:ss,none,yyyy/mm/dd, hh:mm:ss,yyyy/mm/dd, hh:mm:ss, none, m agl,degrees c,none,%,none,m/s,none,decimal degrees,none,m/s,none,m/s,none
***TABLE BEGINS,**
,NE1031,1,1998/06/01,00:00:00,1998/06/01,01:00:00,EDT,1998/06/01,04:00:00,1998/06/01,05:00:00,UTC,5.0,14.9,V0,92.5,V0,0.8,V0,222.2,V0,1.6,V0,0.4,V0
,NE1031,1,1998/06/01,01:00:00,1998/06/01,02:00:00,EDT,1998/06/01,05:00:00,1998/06/01,06:00:00,UTC,5.0,14.8,V0,94.4,V0,0.7,V0,223.6,V0,1.4,V0,0.2,V0
,NE1031,1,1998/06/01,02:00:00,1998/06/01,03:00:00,EDT,1998/06/01,06:00:00,1998/06/01,07:00:00,UTC,5.0,14.7,V0,95.3,V0,0.6,V0,217.9,V0,1.2,V0,0.2,V0

,NE1031,1,1998/06/01,03:00:00,1998/06/01,04:00:00,EDT,1998/06/01,07:00:00,1998/06/01,08:00:00,UTC,5.0,14.8,V0,95.8,V0,0.7,V0,215.5,V0,1.6,V0,0.2,V0
,NE1031,1,1998/06/30,22:00:00,1998/06/30,23:00:00,EDT,1998/07/01,02:00:00,1998/07/01,03:00:00,UTC,5.0,15.2,V0,88.3,V0,0.9,V0,222.4,V0,1.7,V0,0.3,V0
,NE1031,1,1998/06/30,23:00:00,1998/07/01,00:00:00,EDT,1998/07/01,03:00:00,1998/07/01,04:00:00,UTC,5.0,15.2,V0,91,V0,0.8,V0,227.6,V0,1.6,V0,0.2,V0
,NE1031,1,1998/06/01,00:00:00,1998/06/01,01:00:00,EDT,1998/06/01,04:00:00,1998/06/01,05:00:00,UTC,15.0,-99.9,M1,-9.9,M1,3.4,V0,248.6,V0,5.2,V0,1.8,V0
,NE1031,1,1998/06/01,01:00:00,1998/06/01,02:00:00,EDT,1998/06/01,05:00:00,1998/06/01,06:00:00,UTC,15.0,-99.9,M1,-9.9,M1,3,V0,252.6,V0,4.8,V0,1.4,V0
,NE1031,1,1998/06/30,22:00:00,1998/06/30,23:00:00,EDT,1998/07/01,02:00:00,1998/07/01,03:00:00,UTC,15.0,14.8,V0,96.4,V0,2.8,V0,322.4,V0,6.2,V0,0.4,V0
,NE1031,1,1998/06/30,23:00:00,1998/07/01,00:00:00,EDT,1998/07/01,03:00:00,1998/07/01,04:00:00,UTC,15.0,14.7,V0,96.5,V0,2.3,V0,316.2,V0,4.6,V0,0.5,V0
,NE1035,1,1998/06/01,00:00:00,1998/06/01,01:00:00,EDT,1998/06/01,04:00:00,1998/06/01,05:00:00,UTC,4.8,12.8,V0,97.9,V0,2.2,V0,54.1,V0,3.7,V0,0.8,V0
,NE1035,1,1998/06/01,01:00:00,1998/06/01,02:00:00,EDT,1998/06/01,05:00:00,1998/06/01,06:00:00,UTC,4.8,12.4,V0,98.1,V0,2.1,V0,60.1,V0,3.7,V0,0.9,V0
,NE1035,1,1998/06/01,02:00:00,1998/06/01,03:00:00,EDT,1998/06/01,06:00:00,1998/06/01,07:00:00,UTC,4.8,12.4,V0,97.6,V0,1.1,V0,43.3,V0,2,V0,0.5,V0
,NE1035,1,1998/06/01,03:00:00,1998/06/01,04:00:00,EDT,1998/06/01,07:00:00,1998/06/01,08:00:00,UTC,4.8,12.4,V0,97.3,V0,1.2,V0,66.1,V0,1.9,V0,0.5,V0
,NE1035,1,1998/06/30,22:00:00,1998/06/30,23:00:00,EDT,1998/07/01,02:00:00,1998/07/01,03:00:00,UTC,4.8,15.3,V0,91.1,V0,2,V0,35.5,V0,3.2,V0,1,V0
,NE1035,1,1998/06/30,23:00:00,1998/07/01,00:00:00,EDT,1998/07/01,03:00:00,1998/07/01,04:00:00,UTC,4.8,14.9,V0,92,V0,2.2,V0,39.8,V0,3.5,V0,1.1,V0
,NE1035,1,1998/06/01,00:00:00,1998/06/01,01:00:00,EDT,1998/06/01,04:00:00,1998/06/01,05:00:00,UTC,16.6,12.3,V0,97,V0,2.2,V0,64,V0,3.2,V0,1.1,V0
,NE1035,1,1998/06/01,01:00:00,1998/06/01,02:00:00,EDT,1998/06/01,05:00:00,1998/06/01,06:00:00,UTC,16.6,12,V0,98,V0,2.6,V0,67.7,V0,4.1,V0,1,V0
,NE1035,1,1998/06/30,22:00:00,1998/06/30,23:00:00,EDT,1998/07/01,02:00:00,1998/07/01,03:00:00,UTC,16.6,15.1,V0,89.1,V0,2.2,V0,81.6,V0,4.2,V0,0.9,V0
,NE1035,1,1998/06/30,23:00:00,1998/07/01,00:00:00,EDT,1998/07/01,03:00:00,1998/07/01,04:00:00,UTC,16.6,14.9,V0,92.4,V0,2.3,V0,77.7,V0,3.9,V0,0.9,V0
*TABLE ENDS,

4.8 Steps to Follow to Prepare and Send Files in the Data Exchange Standard

1. Gather the network or study documentation files. The documentation should address the following topics:

   - description of network or study;
   - validation level (Section 3);
   - measurement methods (Section "4.5.2 **Measurement Method Information Table**" for details);
   - calibration procedures;
   - audits;
   - station descriptions;
   - instrument descriptions.

2. Access the QSSC's web site to obtain the latest version of the Data Exchange Standard documentation and available Data Exchange Standard file template.[8] Refer to Appendix D for useful file creation notes.

3. Create Data Standard Exchange metadata tables for the following information:

   - station information (obtain standardized station identification codes from the QSSC);
   - measurement method information;
   - other metadata tables as required.

4. Create measurement data tables, taking into account the following items:

   - time zone conversion to local time and UTC;
   - unit conversion to standard units and formats (describe conversion formulas in the measurement method information table in ***COMMENT,** Key Phrases or in a character variable column defined for that purpose);
   - variable naming using the standardized variable names;
   - concentrations measured in mass units per standard cubic meter (e.g., kg/m^3) should consider STP as 0°C and 760 mm Hg pressure.

5. FTP the Data Exchange Standard files and ancillary files to the QSSC (into separate directories) using the following procedures:

   - use the "binary" option for ZIP, image, and word processing files;
   - use the "ASCII" option for CSV and TXT files.

6. Review the quality assurance report that the QSSC automatically generates for each Data Exchange Standard file. Fix any errors, and resubmit the file to the QSSC. Repeat this process until there are no more errors in the file.

---

[8] Please note that the Data Exchange Standard file template is a MS Excel worksheet (.xls). Excel 97 has a file length limit of 65,000 records. This is a practical/reasonable file size for the spatial and temporal aggregation of most data types for archiving (i.e., Langley DAAC data granule).

5.0 ARCHIVAL AND DISSEMINATION OF NARSTO DATA AND INFORMATION

APPLICABILITY: Applicable to anyone wanting to access NARSTO data products through the PDA or obtain data from a NARSTO project.

USE: Guidance for NARSTO Principal Investigators and Project Data Managers and users of NARSTO data and research products.

CONSIDERATIONS: Not all NARSTO data and information will be archived and disseminated by the PDA.

SUMMARY: This section offers a general discussion of NARSTO data archival, access, and distribution; and research product access and distribution.

5.1 Data Archival

The types of electronic data envisioned in the NARSTO PDA are: observational data (surface and aloft), image data, and modeling data. In addition, the PDA will accept Validation Level 0 (i.e., raw) data to ensure its long-term retention and availability. Hardcopy documentation or data will not be archived by the PDA.

NARSTO Principal Investigators will quality assure, validate, and store the data they produce. Each of these processes shall be carried-out in accordance with the recommendations and guidelines provided in the QSMP. NARSTO data may be stored by the Principal Investigator in one or more project data management systems or regional or program data archive. However, all NARSTO project data should be copied to the NARSTO PDA one year after collection/generation. If this is an unreasonable burden on the project, a program/project may specify an alternative data archival schedule in their approved PQMP/QIWP. Justification for an alternative schedule might include, for example, predefined project data delivery dates or extensive data validation and analysis activities requiring more time. The data archiving schedule should in no way limit or restrict the sharing of data within the program/project.

5.2 Data and Research Product Access and Distribution

The research products generated by NARSTO are intended for use by participants, policy makers, and the general public. These products will be used to advance the state of scientific knowledge of ozone and ozone precursor formation and fate in the troposphere over North America.[9]

The following policy is designed to recognize all existing agreements between sponsor and investigator, foster free exchange of information among participants, avoid debilitating controversy within NARSTO, and provide all participants the opportunity to review and comment on relevant materials. NARSTO research products include data and metadata, data models, methods, procedures, papers for publication, promotional materials, press releases, intellectual

[9] Major portions of the policy presented below were adopted from the Southern Oxidant Study Information Policy, January 1, 1992.

property, and all internal and external program reports and public outreach program communications. Guidelines for publication of research results are contained in the ***NARSTO Quality Planning Handbook***.

5.3 Product Access

Level 0 research products generally consist of initial results that have not been officially checked and validated, but may be useful in providing preliminary consolidated results for interpretative and diagnostic analysis by participants. *Normally, these results are not approved for public release*. Levels 1, 2, and 3 results are approved for public release and will be available via the Internet.

5.4 Product Distribution

Level 0 research may be distributed only within participant and co-investigator groups in response to request or as part of an automatic distribution approved by a project oversight committee. All data/products will be made available to participants within 12 months of their generation/production. Recipients of these data/products are expected to proactively participate in the quality control of the product. They may use these products in their work. However, before reporting any findings of fact, or opinion, based upon these products, the recipient must submit such reports to the provider of the research product for review and comment.

Levels 1, 2, and 3 research products contained in the official NARSTO data archive have been approved for public release. Distribution of other research products will normally be affected through the responsible Program/Project Managers or the QSSC.

5.5 Retrieval of Data from the Permanent Data Archive

The LaRC will provide connectivity, search, browse, order and delivery systems to all parties interested in using the research products generated under the NARSTO umbrella. It will serve as the primary repository for all quality assured NARSTO data and final data products. In addition, the Permanent Data Archive will accept Validation Level 0 (i.e., raw) data to ensure its long-term retention and availability. The QSSC Data Management Coordinator will function as the link between the various programs/projects and the NARSTO Permanent Data Archive located at the LaRC.

NARSTO data can be retrieved over the Internet using the NARSTO Home Page. Links are available from the NARSTO Home Page to the PDA at the LaRC DAAC. Additional links may be available to participating NARSTO organizations. Through these links, interested parties will have access to NARSTO databases and other related data sets. The NARSTO home page address is as follows: **http://www.cgenv.com/Narsto/**.

6.0 RECORDS MANAGEMENT

NARSTO requires that sufficient records shall be specified, prepared, reviewed, and maintained to reflect the achievement of the required quality for the work completed. To reiterate, records are completed documents that provide objective evidence of the quality of an item or process. The PQMP and/or QIWP should state requirements and responsibilities for record transmittal, distribution, retention, protection, preservation, traceability, disposition, and retrievability. The PQMP and/or QIWP should also identify how the disposition of records, in accordance with regulatory requirements, schedules, or directives from senior management, will be accomplished.

Paper records will **not** be sent to the NARSTO PDA for long-term archival. It is the responsibility of the projects to establish a project filing and index system, and storage location, for project records. Projects should determine the best filing structure and indexing mechanism to meet project needs. Projects are encouraged to consolidate data records storage, determine an appropriate tool for maintaining an index, and give careful thought to the records storage area (e.g., limited access, environmental conditions suitable for short-term records storage, administrative controls). Development of data forms and logbooks should be a controlled process.

7.0 SOFTWARE CONFIGURATION CONTROL AND DATA MANAGEMENT SYSTEM ADMINISTRATION GUIDELINES

APPLICABILITY: Software and computer system implementation guidance applicable to projects using project specific software and an electronic database.

USE: Guidance for the project data managers responsible for the day to day operation of data management systems.

CONSIDERATIONS: Applicable data management protocols should have checklists or worksheets that facilitate completing project documentation.

SUMMARY: This section defines documentation, quality assurance, and configuration control guidelines for project software and data management systems.

7.1 Project-specific Database and Software Requirements

The need for project-specific databases and software will vary depending on the scope of the project. This section presents the minimum documentation, QA, and configuration control guidelines for project specific databases and software.

7.1.1 Project Database Documentation

Project specific databases include spreadsheets, data sets, and databases (e.g., FoxPro, ORACLE) defined by the project data management group to manage the project data. The project specific databases should be documented in the project file. The database documentation should identify the commercial database product used, the database name, structure, and locations.

The minimum database documentation will consist of

* name and version of commercial software used;
* names of project databases created;
* database structure definitions, including field names and descriptions; and storage location and media.

7.1.2 Project Software Documentation

Project specific software includes programs written by the project data management group for data management activities and programs written by the project team for the production of data products. Data products are defined as any extraction, summary, or analysis of tabular data that result in an electronic data extraction summary or a hard copy product such as tables, graphs, statistics, or maps.

Software documentation should include the software program name, description, special requirements, author, revision, completion date, and documentation of the QA review. Data products documentation should also include all information to uniquely describe how the data product was produced, including the sources used, the manipulations made, and the tools used

to produce the data product. Software documentation can be maintained in hard copy or it may be included as comment blocks embedded within the project software program.

The minimum software documentation should consist of

* name and version of the commercial software used;
* name and version of the software program written by the project;
* author;
* date;
* revision;
* system requirements; and
* storage location.

7.1.3 Project Software Quality Assurance

The project should define the QA requirements for project specific software. At a minimum, software programs to load data, calculate statistics reported in project deliverables, and produce data products should be reviewed to ensure they meet the desired objectives. The reviewer should be someone other than the person who wrote the software program.

7.1.4 Project Specific Software Configuration Control

Project specific software should be protected from unauthorized modification or deletion. This can be accomplished by administrative controls or file security options provided by many computer operating systems. Changes to project software should be documented and a history of revisions which impact the results or data products should be included in the project file. Commercial products are available to maintain a record of software revisions [e.g., Revision Control Software (RCS)]. Another way to do this is to keep the initial or baseline software in a storage area separate from the working software. Then, when the software changes, the new software can be moved to this separate area also, there maintaining copies of all revisions. Project specific configuration and revision control should be documented in the project QIWP.

The project software configuration control documentation should include

* commercial software used;
* program names;
* revisions (including dates of revision); and
* storage locations.

Before the development of software application programs, a requirements analysis shall be conducted. Developed software application programs shall be tested and validated to ensure compliance with all user requirements and to provide confidence that the software will perform satisfactorily in service. The technical adequacy of results generated by these applications shall also be reviewed by another person, tested and validated (e.g., dose calculations, sample management, risk assessment). Configuration management of the developed software application programs shall be conducted.

7.1.5 Software Configuration Control

Changes and enhancements to software, both during the development process and after formal acceptance, are inevitable. Problems detected during testing must be corrected, existing system requirements often change or are reallocated to improve processing, and new requirements are often added. For any software development effort, this change process is controlled to ensure that only authorized changes or enhancements are incorporated and that system integrity is maintained.

Configuration control is applied to project baselines. A Software Configuration Management process must be identified and implemented. The flow of configuration control is managed by a reporting documentation system geared to track software changes.

Before any changes or enhancements are made to software applications, appropriate approval must be granted and documented. All changes or enhancements made to software applications shall be documented.

7.1.6 Database Structure Configuration Control

Database configuration control includes many of the same requirements as software configuration control. Database configuration control ensures that the physical database design and implementation are properly protected from unauthorized changes or destruction and that authorized changes are identified and tracked.

Database structures are controlled in the same way as software (Section 7.2.3). Before any changes or enhancements are made to a project's database structure, appropriate approval must be granted and documented. Access control requirements shall be adhered to, and all changes or enhancements made to database structures shall be documented.

7.2 Day-to-Day Operation of Data Management Systems

This section addresses the day-to-day operations of a data management system, including backups, access, security, data entry, data control, and data archival.

7.2.1 System Backups

Project data should be protected from loss through preventative database backup and recovery mechanisms in accordance with the data management planning details. Database backups should be performed on a periodic basis at a frequency to be defined by each project. This frequency shall be selected to minimize the extent of consequences of data loss and time required for data recovery. Recovery procedures will be developed and documented in preparation for the event of hardware or software failure.

7.2.2 Data System and Database Access

Projects shall protect systems and data from unauthorized access by implementation of administrative and procedural controls. Access controls shall be managed based upon specific data user roles that will be defined by the types of data and functionality required (e.g., a data management specialist may have the capability of creating and updating data from field logs; a program manager may require read-only access to perform on-line queries). The mechanism for implementing access control should be documented in project data management plans.

7.2.3 Data Entry

Data entry, transfer, and transformation activities should be verified to ensure that data integrity is maintained. This includes all movement/copying of data from one storage medium to another and transformation from one format to another. All data, including analytical data produced and reported by a laboratory, should conform to this requirement.

This verification encompasses all data recording media, such as handwritten or hard copy produced via electronic means, as well as electronically stored, such as in a database. It also includes all data collection methods (e.g., electronic collection through real-time monitoring instrumentation, bar coding equipment, and handwritten log entries).

If a data transformation or transfer activity has occurred before receipt of the data by project personnel (i.e., between creation and final reporting), the verification may be performed by the reporting party but only if sufficient evidence to support the validity of the process can be provided by the reporting party. For example, if a laboratory technician captures data from a laboratory instrument and records it in a logbook, enters the data from the logbook into an electronic data deliverable format, and then transfers the data to the project, the verification process may be performed by the laboratory. The mechanism for a project's data entry, transfer, and transformation verification processes should be documented in the project QIWP.

7.2.4 Database Content Configuration Control

A project should establish configuration control requirements for the contents of the project database. The requirements should ensure traceability of field and laboratory data from the original reported values through authorized data changes to current values stored in the database. The configuration control should define the approval process required for making changes to the database and the documentation required for each database change.

The minimum information maintained for each database change should include

* a description of the change;

* the reason for the change;

* the name of the individual making the change;

* the date of the change; and

* a copy of the data before the change took place.

7.2.5 Data Archival

Project data that have been quality assured should be transferred to the QSSC for transfer to the PDA for archival using formats specified in Section 4. The schedule and mechanism for data archival should be specified in the QIWP.

APPENDIXES

APPENDIX A: COMPLETE LISTING OF STANDARD VARIABLE NAMES, UNITS, AND FORMATS

Concentrations measured in mass units per standard cubic meter should consider STP as 0°C and 760 mm Hg pressure. If measurements are made in units of mass / volume at ambient temperature and pressure, ensure this is clearly documented in the "STP Condition" field of the measurement method information table. **The conversion factor from mass units to ppbv should also be reported.** The number of decimal places shown in the *Format* field is to be used only as a guideline. Formats specified in files should match the actual formats of the data in the file. These formats have been designed to be large enough to allow unambiguously large negative missing value codes to be defined (see Sections 2.7 and 4.4.3).

Oxidants

Long name	***Units***	***Format***	***Short Name***
o3	ppbv	5.1	o3
o3 - maximum in specified period	ppbv	5.1	o3max

Nitrogen Species

Long name	***Units***	***Format***	***Short Name***
amonia	ppbv	6.3	nh3
amonium	ppbv	6.3	nh4
nitrate	ppbv	6.3	no3
no	ppbv	6.3	no
no2	ppbv	6.3	no2
nox	ppbv	6.3	nox
noy	ppbv	6.3	noy
pan	ppbv	6.3	pan

Volatile Organic Compounds

Long name	***Units***	***Format***	***Short Name***
1-butene/isobutene	ppbv	7.2	butis1
1-hexene	ppbv	7.2	hexen1
1-pentene	ppbv	7.2	pente1
111-trichloroethane	ppbv	7.2	tri111
123-trimethylbenzene	ppbv	7.2	tri123

124-trimethylbenzene	ppbv	7.2	tri124
13-butadiene	ppbv	7.2	buta13
13-diethylbenzene	ppbv	7.2	diet13
135-trimethylbenzene	ppbv	7.2	tri135
14-dichlorobenzene	ppbv	7.2	dich14
14-diethylbenzene	ppbv	7.2	diet14
2-ethyltoluene	ppbv	7.2	ethyl2
2-methyl-1-butene	ppbv	7.2	but1m2
2-methyl-2-butene	ppbv	7.2	but2m2
2-methylheptane	ppbv	7.2	methp2
2-methylhexane	ppbv	7.2	methx2
2-methylpentane	ppbv	7.2	metpn2
22-dimethylbutane	ppbv	7.2	dime22
22-dimethylhexane	ppbv	7.2	dimh22
22-dimethylpentane	ppbv	7.2	dimp22
224-trimethylpentane	ppbv	7.2	tti224
225-trimethylhexane	ppbv	7.2	tri225
23-dimethylbutane	ppbv	7.2	dimb23
23-dimethylpentane	ppbv	7.2	dimp23
234-trimethylpentane	ppbv	7.2	tri234
24-dimethylhexne	ppbv	7.2	dimh24
24-dimethylpentane	ppbv	7.2	dime24
25-dimethylbenzaldehyde	ppbv	7.2	dime25
25-dimethylhexane	ppbv	7.2	dimh25
3-ethyltoluene	ppbv	7.2	ethto3
3-methyl-1-pentene	ppbv	7.2	mehtl3
3-methylheptane	ppbv	7.2	methp3
3-methylhexane	ppbv	7.2	methx3
3-methylpentane	ppbv	7.2	metpn3
4-ethyltoluene	ppbv	7.2	ethyl4
4-methyl-1-pentene	ppbv	7.2	mthyl4

4-methylheptane	ppbv	7.2	methp4
6.2trans-12-dimethylcyclohexane	ppbv	7.2	tran62
acetaldehyde	ppbv	7.2	acetal
acetone	ppbv	7.2	aceton
acetylene	ppbv	7.2	acetyl
acrolein	ppbv	7.2	acrole
benzaldehyde	ppbv	7.2	benzal
benzene	ppbv	7.2	benzen
butane	ppbv	7.2	butane
carbontetrachloride	ppbv	7.2	cartch
chloroform	ppbv	7.2	chlorf
chloromethane	ppbv	7.2	chlorm
cis-12-dimethylcyclohexane	ppbv	7.2	dimc12
cis-13-dimethylcyclohexane	ppbv	7.2	dimc13
cis-14/t-13-dimethylcyclohexane	ppbv	7.2	dimc14
cis-2-butene	ppbv	7.2	butci2
cis-2-hexene	ppbv	7.2	hexec2
cis-2-pentene	ppbv	7.2	pentc2
cis-3-methyl-2-pentene	ppbv	7.2	pentc3
cis-4-methyl-2-pentene	ppbv	7.2	metcis
co	ppbv	7.2	co
crotonaldehyde	ppbv	7.2	croton
cyclohexane	ppbv	7.2	cycloh
cyclopentane	ppbv	7.2	cylopa
cyclopentene	ppbv	7.2	cyclop
decane	ppbv	7.2	decane
dichloromethane	ppbv	7.2	dichlo
dodecane	ppbv	7.2	dodecn
ethane	ppbv	7.2	ethane
ethylbenzene	ppbv	7.2	ethylb
ethylene	ppbv	7.2	ethyle

formaldehyde	ppbv	7.2	form
freon11	ppbv	7.2	freo11
freon22	ppbv	7.2	freo22
heptane	ppbv	7.2	heptan
hexanal	ppbv	7.2	hexana
hexane	ppbv	7.2	hexane
indane	ppbv	7.2	indane
iso-propylbenzene	ppbv	7.2	isopro
isobutane	ppbv	7.2	isobut
isopentane	ppbv	7.2	isopen
isoprene	ppbv	7.2	isopre
isovaleraldehyde	ppbv	7.2	isoval
m and p-xylene	ppbv	7.2	xylemp
m-tolualdehyde	ppbv	7.2	mtolua
mek	ppbv	7.2	mek
methylcyclohexane	ppbv	7.2	methex
methylcyclopentane	ppbv	7.2	metpen
mibk	ppbv	7.2	mibk
n-butylbenzene	ppbv	7.2	butyln
n-propylbenzene	ppbv	7.2	npropy
naphthalene	ppbv	7.2	naphth
nonane	ppbv	7.2	nonan
o-tolualdehyde	ppbv	7.2	otolua
o-xylene	ppbv	7.2	xyleno
octane	ppbv	7.2	octane
p-cymene	ppbv	7.2	cymenp
p-tolualdehyde	ppbv	7.2	ptolua
pentane	ppbv	7.2	pentan
propane	ppbv	7.2	propan
propionaldehyde	ppbv	7.2	propio
propylene	ppbv	7.2	propyl

reactive hydrocarbon	ppbv	7.2	rox
styrene	ppbv	7.2	styren
toluene	ppbv	7.2	toluen
trans-12-dimethylcyclohexane	ppbv	7.2	dimt12
trans-14-dimethylcyclohexane	ppbv	7.2	dimt14
trans-2-butene	ppbv	7.2	but2tr
trans-2-hexene	ppbv	7.2	hexet2
trans-2-pentene	ppbv	7.2	pentt2
trans-3-methyl-2-pentene	ppbv	7.2	pentt3
trichloroethylene	ppbv	7.2	trichh
undecane	ppbv	7.2	undeca
valeraldehyde	ppbv	7.2	valera

Gas-phase Aerosol Precursors

Long name	***Units***	***Format***	***Short Name***
so2	ppbv	7.2	so2

Meteorology

Long name	***Units***	***Format***	***Short Name***
mean visibility	m	9.2	mvis
barometric pressure	mb [in a comment, specify the temperature condition]	7.2	bpress
dew point	degrees c	7.2	dewpnt
mixing height	m agl	7.2	mixhgt
precipitation	mm	7.2	precip
pressure or partial pressure	pascals	9.2	press
relative humidity	%	7.2	relhum
solar radiation	watts/m2	7.2	solrad
temperature	degrees c	7.2	temp
tower height	m	7.1	towhgt

uv radiation	watts/m2	7.2	uvrad
wind direction	decimal degrees	7.1	windir
wind speed	m/s	7.2	winspd
wind speed (maximum in sampling	m/s	7.2	wnspmx
wind speed (minimum in sampling	m/s	7.2	wnspmn

Metadata

Long name	***Units***	***Format***	***Short Name***
Canister id	none	6	canist
End Date (UTC)	none	10	edateu
End Date (local time)	none	10	edatel
End time (UTC)	none	8	etimeu
End time (local time)	none	8	etimel
Local Time Zone	none	3	lzone
Local Time Zone	none	3	tzonel
Start Date (UTC)	none	10	sdateu
Start Date (local time)	none	10	sdatel
Start Time (UTC)	none	8	stimeu
Start Time (local time)	none	8	stimel
Station ID	none	5	stnid
UTC Code	none	3	utc
UTC Code	none	3	tzoneu
altitude	m asl	8.2	alt
altitude from Global Positioning	m asl	8.2	gpsa
comments	none	60	cmnt
flight number	none	6	flight
height	m agl	8.2	height
latitude (decimal degrees)	decimal degrees	10.5	latdec
latitude (degrees)	degrees	4	latdeg
latitude (minutes)	minutes	3	latmin
latitude (seconds)	seconds	3	latsec

longitude (decimal degrees)	decimal degrees	10.5	londec
longitude (degrees)	degrees	4	londeg
longitude (minutes)	minutes	3	lonmin
longitude (seconds)	seconds	3	lonsec
validation level code	none	1	valcod

APPENDIX B: SI MULTIPLES, PREFIXES, AND SYMBOLS

Multiples and sub-multiples	Prefixes	Symbols
1 000 000 000 000 000 000 000 000	10^{24} yotta	Y
1 000 000 000 000 000 000 000	10^{21} zetta	Z
1 000 000 000 000 000 000	10^{18} exa	E
1 000 000 000 000 000	10^{15} peta	P
1 000 000 000 000	10^{12} tera	T
1 000 000 000	10^{9} giga	G
1 000 000	10^{6} mega	M
1 000	10^{3} kilo	k
100	10^{2} hecto	h
10	10^{1} deca	da
1	10^{0}	
0.1	10^{-1} deci	d
0.01	10^{-2} centi	c
0.001	10^{-3} milli	m
0.000 001	10-6 micro	μ
0.000 000 001	10-9 nano	n
0.000 000 000 001	10-12 pico	p
0.000 000 000 000 001	10-15 femto	f
0.000 000 000 000 000 001	10-18 atto	a
0.000 000 000 000 000 000 001	10-21 zepto	z
0.000 000 000 000 000 000 000 001	10-24 yocto	y

APPENDIX C: KEY PHRASE SUMMARY TABLES

The following tables summarize the information presented in Section 4.4 and are intended to serve as a quick reference for the Key Phrase elements of the Data Exchange Standard.

The Key Phrase Summary Tables that follow list all of the Data Exchange Standard Key Phrases (except the QA center key phrases, which are discussed). Data Originators can use these tables to determine whether a Key Phrase is mandatory or optional, the number of fields that are expected to follow the Key Phrase, and the maximum allowable formats for these fields. The summary tables contain the following information:

Column	Description
Key Phrase	The name of the Key Phrase.
Mandatory / Optional	m = Data Exchange Standard files must contain this Key Phrase. o = Data Exchange Standard files may optionally contain this Key Phrase.
Total # of Fields	The number of parameters in this Key Phrase. Some table Key Phrases have multiple parameters, depending on the number of variables in the table. In these cases, the upper and lower range is shown.
Field #	Indicates the field number that is being described. Key phrases with more than one field have an entry for each field in the table. The Key Phrase and total number of fields are not repeated after the first description.
Short Name	The name used by software programs that process the Data Exchange Standard files. The variable short names cannot be longer than 8 characters.
Type	i = integer f = floating point c = character d = date t = time
Format	For character fields, the maximum number of characters. For integers, the maximum number of digits. For dates or times, the maximum number of characters (inculding both digits and the slash, dash, or colon separators). Note that although the current Key Phrases do not use floating point fields, the data rows usually do. Floating point formats give the total width of the field (including any sign, all digits, and the decimal point) and the number of digits after the decimal place.

Header Key Phrases

Key Phrase	Mand./ Optional	Total # of Fields	Field #	Short Name	Type	Format
*DATA EXCHANGE STANDARD VERSION,	m	1	1	dtexver	c	60
*QUALITY CONTROL LEVEL,	m	1	1	qclvl	i	3
*LAST REVISION TO DATA-DATE/REVISION NUMBER,	m	2	1	revdt	d	10
	m		2	revnum	c	10
*ORGANIZATION ACRONYM,	m	1	1	orgacrn	c	10
*STUDY OR NETWORK ACRONYM,	m	1	1	stnwacrn	c	15
*FILE CONTENTS DESCRIPTION-SHORT/LONG,	m	2	1	flcnts	c	15
	m		2	flcntl	c	150
*PRINCIPAL INVESTIGATOR NAME-LAST/FIRST,	m	2	1	pilname	c	60
	m		2	pifname	c	60
*PRINCIPAL INVESTIGATOR AFFILIATION,	m	1	1	piafltn	c	60
*COUNTRY CODE,	m	1	1	cntrcd	c	2
*STATE OR PROVINCE CODE,	m	1	1	stprcd	c	2
*DATA REPORTING INTERVAL-CODE/DESCRIPTION,	m	2	1	dtrpcd	c	5
	m		2	dtrpdsc	c	150
*COMMENT,	o	1	1	cmnt	c	150
*PRINCIPAL INVESTIGATOR CONTACT INFORMATION,	o	1	1	picntct	c	200
*STUDY OR NETWORK NAME,	o	1	1	stnwname	c	150
*NAME (LAST-FIRST)/AFFILIATION OF PERSON WHO GENERATED THIS FILE,	o	3	1	gnflln	c	60
	o		2	gnflfn	c	60
	o		3	gnflaf	c	150
*DATE THIS FILE GENERATED,	o	1	1	flgendt	d	10
*NAME/VERSION OF SOFTWARE USED TO CREATE THIS FILE,	o	2	1	gnflswn	c	60
	o		2	gnflswv	c	60
*STANDARD CHARACTERS,	o	1	1	stndchr	c	150

Key Phrase	Mand./ Optional	Total # of Fields	Field #	Short Name	Type	Format
*ANCILLARY FILE NAME/FORMAT/VERSION,	o	3	1	ancflnm	c	67
	o		2	ancflfm	c	30
	o		3	ancflvr	c	30

QA Center Key Phrases

These are Key Phrases that the QA center adds to the file. They are described in the text. A table is not provided here in this draft; it may be provided in a later draft. These Key Phrases must not be provided by the Data Originator.

Table Key Phrases

Key Phrase	Mand. / Optional	Total # of Fields	Field #	Short Name	Type	Format
*TABLE NAME,	m	1	1	tbnm	c	30
*TABLE FOCUS,	m	1	1	tbfc	c	10
*TABLE COLUMN SHORT NAMES,	m	1-999	1-999	tbsn001 - tbsn999	c	6
*TABLE COLUMN LONG NAMES,	m	1-999	1-999	tbln001- tbln999	c	60
*TABLE COLUMN TYPES,	m	1-999	1-999	tbtp001- tbtp999	c	10
*TABLE COLUMN FORMATS,	m	1-999	1-999	tbfm001- tbfm999	c	10
*TABLE COLUMN UNITS,	m	1-999	1-999	tbun001 -tbun999	c	20
*TABLE BEGINS,	m	1	1	tbbeg	-	-
*TABLE ENDS,	m	1	1	tbend	-	-
*TABLE NUMBER OF COLUMNS,	o	1	1	tbnc	i	4
*TABLE NUMBER OF ROWS,	o	1	1	tbnr	i	8
*TABLE MAXIMUM RECORD	o	1	1	tbmaxsz	i	5

Key Phrase	Mand. / Optional	Total # of Fields	Field #	Short Name	Type	Format
SIZE,						
*TABLE KEY FIELD NAMES,	o	1-999	1-999	tbkf001-tbkf999	c	8
*TABLE COLUMN LOOKUP TABLE NAMES,	o	1-999	1-999	tblkp001-tblkp999	c	30
*TABLE COLUMN MISSING CODES,	o	1-999	1-999	tbms001 -tbms999	c	11
*TABLE COLUMN START/END POSITIONS,	o	1-999	1-999	tbse001-tbse999	c	9

APPENDIX D: NOTES FOR DATA ORIGINATORS USING SPREADSHEETS

1. Key Phrases should be stored in the first column, with each parameter stored in separate cells in the subsequent columns. When a spreadsheet is saved as “CSV” (comma-separated values), a comma is automatically inserted between each column. There is no need to manually place a comma after the Key Phrase.

2. The first column in the spreadsheet should be used for Key Phrases. In Tables, the first value should appear in the second column, in order to align it with the column headings.

3. Attention must be paid to controlling the format in which dates are output to and input from “CSV” files. The default date format in Excel 97 is mm-dd-yy. Date fields must be changed to the format yyyy/mm/dd before exporting the data to a "CSV" file. To do this, highlight the date cells, and select Format, Cells, Custom, and type in "yyyy/mm/dd" in the "Type" window. In Windows 95, a custom style can be defined for the short date style in the Regional Settings under Control Panel. This may help preserve the correct date format when importing “CSV” files into Excel.

4. To display all of the characters in each cell of an Excel 97 spreadsheet, highlight the spreadsheet, and select Format, Column, AutoFit Selection.

5. When exporting a spreadsheet file to a "CSV" file, extra commas will be inserted at the end of rows, up to the number of columns in the longest row. The extra commas can be left in the file, and will be ignored in the quality assurance procedures.

6. Please note that the current spreadsheet version of the Data Exchange Standard file template is a MS Excel 97 worksheet (.xls). Excel 97 has a file length limit of 65,536 records. This is a practical/reasonable file size for the spatial and temporal aggregation of most data types for archiving (i.e., Langley DAAC data granule).