WORLD METEOROLOGICAL ORGANIZATION
GLOBAL ATMOSPHERE WATCH

No. 150

UPDATED GUIDELINES for ATMOSPHERIC TRACE GAS DATA MANAGEMENT

APRIL 2003

WORLD METEOROLOGICAL ORGANIZATION
GLOBAL ATMOSPHERE WATCH

No. 150

UPDATED GUIDELINES for ATMOSPHERIC TRACE GAS DATA MANAGEMENT

Ken Masarie and Pieter Tans

WMO TD No. 1149

TABLE OF CONTENTS

1. INTRODUCTION

Making the necessary measurements of atmospheric CO_2, its stable isotopes, and other trace gas species to better our understanding of the global carbon cycle consumes increasing amounts of time and resources. Changing scientific objectives impose demands for greater measurement precision and better temporal and spatial coverage. Advances in technology have made these demands attainable. Instruments have been designed to be more compact, robust, and efficient making it easier to produce reliable in situ measurements from ships, aircraft, towers, and from permanent and temporary remote locations. The quality of sample storage containers used in discrete sampling programs has improved, increasing the long-term stability of sampled air for a greater number of trace gas species. Semi-automatic analytical systems have been designed to use smaller volumes of air enabling multiple-species analyses from a single ambient sample. Finally, the affordability of computers and electronic storage media has greatly reduced a once formidable component of the operational budget. This partial list of advancements combined with an increasing scientific demand for more observations has resulted in the generation and accumulation of greater volumes of data. These data must be organized and maintained so that the effort required to make the measurements results in advances in our understanding of the scientific issues.

Measurement programs maintain records of sample collection details (e.g., collection location, date), raw analysis data (e.g., chromatograms, voltages), processed analysis data (e.g., mixing ratios, isotopic ratios), instrument diagnostic data (e.g., temperature, pressure), and standard gas calibration histories. In many instances, these data are maintained for each trace gas constituent measured. Further, individual laboratories may operate both continuous and discrete (flask and cylinder) measurement strategies from both fixed (land surface, towers, and ice cores) and moving (ship and aircraft) platforms. All of this information must be managed so that measurement data can be readily viewed, re-processed, selected, analyzed, and disseminated. A poor data management strategy can render even the very best measurements almost useless. Thus, it is imperative that measurement programs employ a well-designed data management strategy.

The carbon cycle measurement community, and in particular, the CO_2 community, has spent considerable effort developing instrument and technical guidelines for scientists entering into the atmospheric trace gas measurement field [*WMO*, 1998]. Guidelines which emphasize measurement techniques, calibration methods, and common pitfalls, provide fledgling measurement programs with a “recipe” for obtaining high-precision measurement results more quickly. A similar effort is required for the management and maintenance of data produced by these programs. A first attempt to establish guidelines for atmospheric trace gas database management is presented. The data management strategy described is a compendium of concepts in use among many of the laboratories making trace gas measurements. The National Oceanic and Atmospheric Administration (NOAA) Climate Monitoring and Diagnostic Laboratory (CMDL) Carbon Cycle Greenhouse Gases (CCGG) Group has committed considerable resources towards the management of discrete and continuous data from the cooperative air sampling network, the CMDL baseline observatories, and the tall towers and light aircraft programs. Many of the examples presented are from CCGG programs that address many of the same data management considerations as other carbon cycle measurement laboratories. The majority of guidelines set forth here have been in use by CCGG for many years; however, the CCGG strategy has considerable room for improvement and continues to evolve as the measurement programs expand and new technologies become available. This discussion represents the direction in which the CMDL Carbon Cycle Greenhouse Gases Group is moving towards a robust database management system.

2. GENERAL DATABASE MANAGEMENT CONCEPTS

Database management strategies can vary infinitely, and while there is probably no “right” approach, some approaches are clearly better than others. A working knowledge of general database management concepts and a well-conceived data management strategy are minimum requirements needed to design a robust and enduring database management system (DBMS). The schematic in Figure 1 depicts the relative placement of a DBMS within the computer environment. The DBMS is generally a software package that runs on top of the operating system and manages the access to and security of data according to the design criteria set forth by the database architect(s). While the database management software is dedicated to the maintenance of the data and supporting information, additional software (e.g., S-Plus, IDL, custom applications) usually exists that can access the data directly or indirectly by sending requests to the database management software. Other software applications (e.g., word processors, desktop publishers, spreadsheets) may coexist with the DBMS and will likely compete with the management software for computer memory and processing time.

Database management system software packages are commercially available (e.g., Oracle, Sybase, MySQL, Microsoft's Access, and Borland's Paradox) or can be custom-built to meet laboratory design requirements. Design considerations, investment of time, cost, flexibility, and available expertise will likely determine which package is pursued. While the task of designing a strategy and adopting a DBMS is daunting, it is made easier by the fact that the majority of database management systems employ standard data management terminology and concepts.

A typical "relational" DBMS (RDBMS) will require that data be grouped or classified by function or purpose. For example, site name, site code, and position coordinates are data items that together describe a sampling location. Information pertaining to the collection of a discrete ambient sample (e.g., site code, date/time, container identification, collection method, position, etc.) is classified as sample collection details. Data with common classification are placed in tables (database terminology) or files (data processing terminology) which are managed by the RDBMS (Table 1).

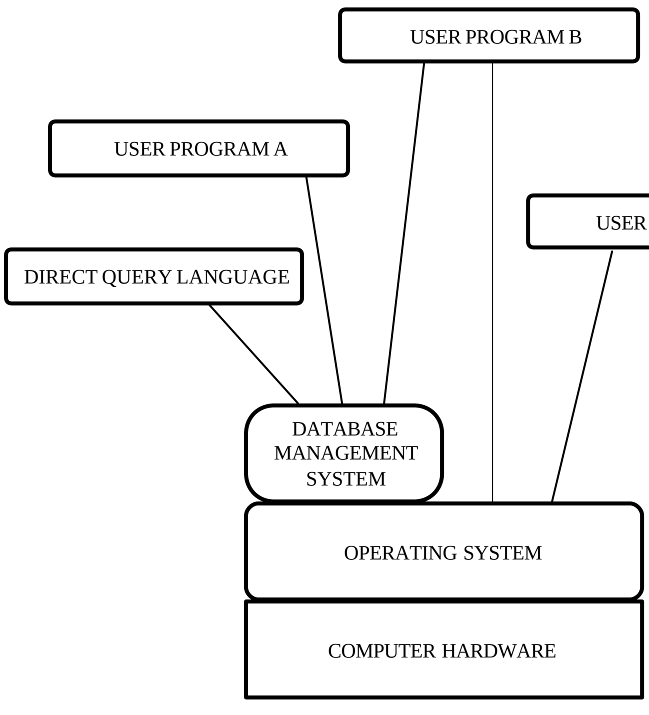

Figure 1: The relative placement of a database management system within the computing environment.

Tables are organized by rows and columns. Each column (field) heading is a data attribute such as site code, sample year, month, etc. Each row (record) in the table is a set of data attributes. Records within a table must be unique. Uniqueness is accomplished by requiring that at least one attribute in a record can distinguish the entire record from all others in the table. This attribute or set of attributes that uniquely identifies a record within a table is called the "primary key." For example, the primary key in a table describing the CMDL cooperative air sampling network (Table 1, top) is the single attribute "site." As a second example, discrete sampling events from the CMDL network are uniquely identified by the set of 8 shaded attributes in Table 1 (bottom). Attributes that do not constitute the primary key are referred to as "non-key" attributes. The use of a primary key maximizes the efficiency with which data can be accessed from a table. For example, the site description for the sampling program at Alert can be unambiguously identified by searching the network description table (Table 1, top) for "ALT."

While a record of data can be uniquely identified using the primary key, it is often valuable to access subsets of a table using non-key attributes. For example, searching the table of sample collection details (Table 1, bottom) and requiring that the site be "ALT" and that wind speed at the time of sample collection be greater than 5 ms^{-1} will return a subset of all samples that satisfy these search conditions. If a unique sampling event is desired then data items in the sample collection details primary key must be specified.

Cooperative Air Sampling Network Description

site	Location Name	Country	Lat.	Long.	masl	Cooperating Agency
ALT	Alert, N.W.T.	Canada	82°27'N	062°31'W	210	Environment Canada/Atmospheric Environment Service
AMS	Amsterdam Island	France	37°57'S	077°32'E	150	Centre des Faibles Radioactivities/TAAF
ASC	Ascension Island	U.K.	07°55'S	014°25'W	54	DOD/U.S.A.F. and Pan American World Airways
ASK	Assekrem	Algeria	23°11'N	005°25'E	2728	Tamanrasset GAW Observatory
⋮						

Sample Collection Details

site	yr	mo	dy	hr	mn	id	m	lat	lon	alt	wd	ws
ALT	1997	09	11	16	48	135-91	P	82.45	-62.52	210	233	10
ALT	1997	09	11	16	48	136-91	P	82.45	-62.52	210	233	10
ASC	1997	10	30	08	15	331-91	P	-7.92	-14.42	54	105	18
ASC	1997	10	30	08	15	332-91	P	-7.92	-14.42	54	105	18
ASK	1997	09	20	14	00	983-91	P	23.18	5.42	2728	110	04
ASK	1997	09	20	14	00	984-91	P	23.18	5.42	2728	110	04
⋮												

Table1: Data with common classification are represented in tables. Tables are organized by rows and columns. Each column heading identifies a data attribute. Each row is a collection of attributes. The intersection of a row and column is a data item. A primary key is a set of attributes that, together, uniquely identify a table row. Top: Portion of a table containing the CMDL cooperative air sampling network description. Bottom: Portion of a table containing discrete sample collection details from the CMDL cooperative air sampling network. The primary key is shown as the set of shaded attributes.

Because data are grouped according to function (i.e., site description, sample collection, sample analysis, etc.), describing the complete "story", from sample collection to final analysis result, for any one air sample will require referencing or "relating" data from several tables. The high-level RDBMS query language SQL (Structured Query Language) was developed by IBM Corporation [*Astraham et al.*, 1976] to manipulate information in and among tables. Important to this general discussion is the basic SQL concept that subsets of information from one or more tables may be obtained using standard Boolean operators. In the above simple example, a subset of Alert sampling events was obtained by satisfying the "If this is true AND that is true" conditional statement. Boolean logic is extremely powerful in creating more complicated subsets involving multiple tables, e.g., select ['a' AND ('b' OR 'c')] from table 1 AND 'd' from table 2 AND NOT 'e' from table 3 where a-e are desired conditions.

This brief discussion on database management system concepts will be helpful in developing an atmospheric trace gas data management strategy. A thorough treatment of database management concepts can be found in *Rumble and Smith*, [1990] and elsewhere.

3. GENERAL ATMOSPHERIC TRACE GAS DATA MANAGEMENT STRATEGY CONSIDERATIONS

A database management strategy will include an assessment of the overall “flow” and organization of information within the measurement program, users’ needs, and future considerations. It is also important to consider all measurement products within the laboratory. For example, CCGG operates many different measurement programs including continuous and discrete measurements from fixed and moving platforms. By identifying common data management requirements and analysis needs, the data management strategy can include the shared needs of these relatively independent programs. The first task is to formulate the data management strategy and then to determine the RDBMS best suited to the strategy. The discussion that follows describes general data management strategy considerations. Each issue must be addressed before a RDBMS package is selected and regardless of which package is ultimately selected.

3.1 Data Flow

A critical starting point in the development of any database management strategy is an assessment of the flow of information within the measurement program. The term “flow” describes the hierarchy or progression of data from its rawest form to its most processed form. The hierarchy of information depends largely on how and by whom the different types of data will be used. While the RDBMS does not control the flow of data, it does structure the data to ensure efficient management and retrieval. The schematic in Figure 2 describes the typical flow of information within a continuous or discrete atmospheric measurement program. A similar diagram describes the flow of data from a laboratory’s standard gas calibration program and its “relation” to the hierarchy of ambient measurement data.

Measurement programs will undoubtedly deviate from this general flow scheme. For example, the “ANALYTICAL SYSTEM” used for the analysis of discrete samples from the CMDL cooperative air sampling network and aircraft programs has two independent components; the first component includes an NDIR CO_2 analyzer and a GC system for the separation and subsequent analysis of CH_4, CO, H_2, N_2O, and SF_6; the second includes an on-line CO_2 extraction system with mass spectrometer for the analysis of the stable isotopes of CO_2 (?^{13}C and ?^{18}O). While the two analytical systems are independent and, in fact, reside in different buildings, the flow of data from these two systems is integrated into the overall flow of information derived from the air sampling network. Similarly, the NDIR CO_2 analyzers in continuous operation at each of the four CMDL baseline observatories are independently operated, but because the flow of data from each continuous program is similar, that data can be readily incorporated into the overall flow scheme of the RDBMS maintained in Boulder.

The assessment of data flow within a measurement program is not complete without a critical evaluation of the data hierarchy required to properly maintain and propagate laboratory reference scales. The hierarchy of information from the calibration of standard gases within CCGG is nearly identical to that described for the discrete and continuous programs (Figure 2). The critical difference being that calibration tables must not only relate to one another but must also be accessible to computer programs that process discrete and continuous analysis details where the determination of the relative amount of trace gas in the ambient sample requires an assigned “best” value to the relevant working reference standard(s). The set of attributes that relates atmospheric measurement tables to calibration data tables provides the mechanism by which changes to calibration scales or values assigned to laboratory standards can be readily propagated throughout the calibration tables and the discrete and continuous data tables when necessary.

3.2 Data Classification

A second critical step in the development of any data management strategy is the identification of the types of data that are currently being produced or may be generated in the

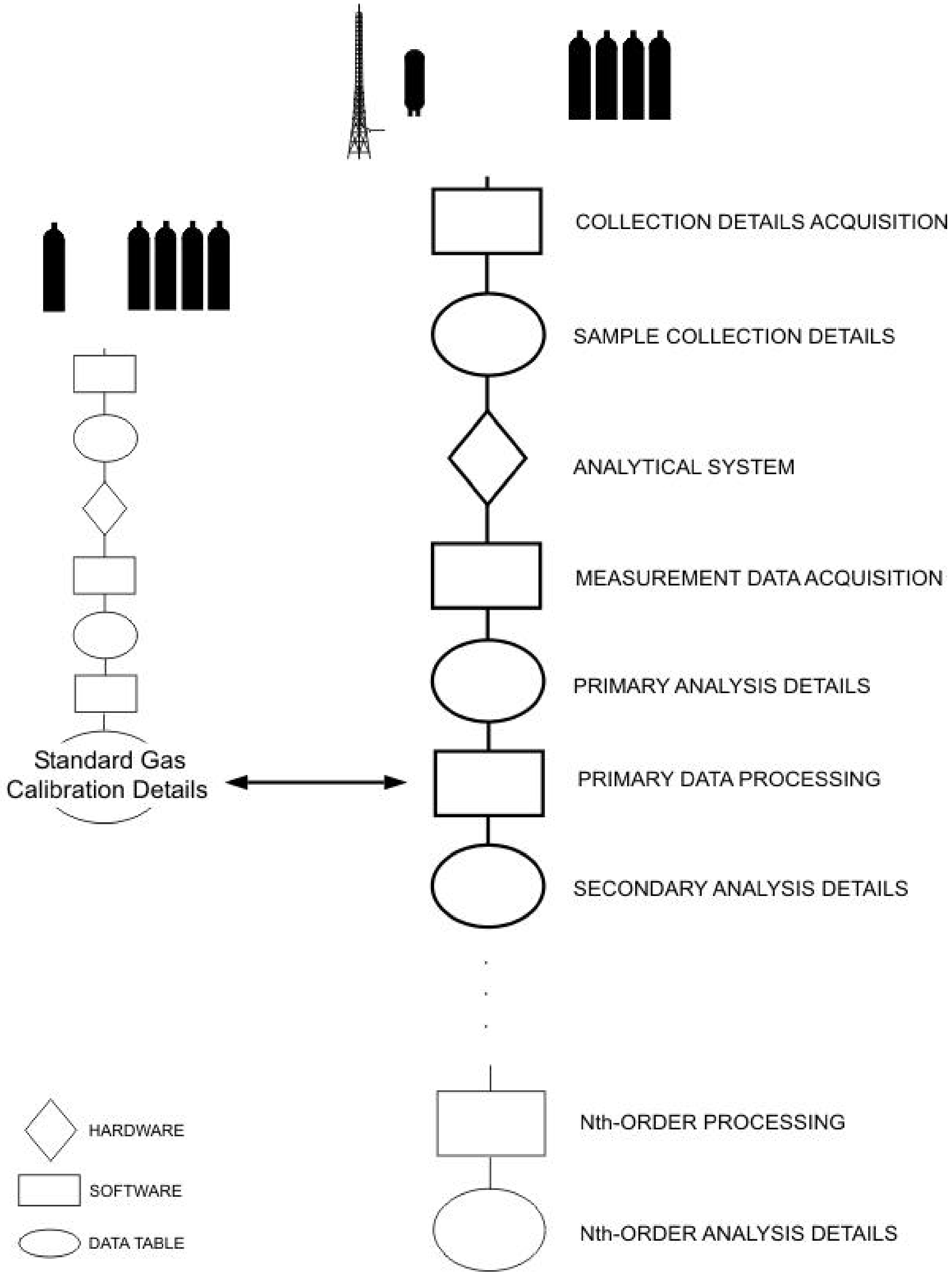

Figure 2: Typical flow of information within an atmospheric trace gas measurement program. The flow of information may be similar for the calibration of laboratory standard gas cylinders. Calibration results must be accessible to measurement programs that determine relative amounts of atmospheric compounds.

future. Data produced by a typical atmospheric trace gas measurement program may consist of voltages, frequencies, currents, or chromatographic parameters, and sample collection details. In addition, there may be a host of supporting information which may include flow rates, temperatures, pressures, and operator comments. Some of these data will be critical to the determination of the amount of trace gas constituent in the sample, other data will uniquely identify the location, time, and details of the sample collection, and still other information will be important in determining the stability of the measurement system and the overall quality of the measurements. Because there are many different types of data, it is important to classify or group data according to their purpose or function. While there may be infinite classifications and sub-classifications of the data, for the purpose of this discussion, a general classification scheme will be described.

An obvious starting point in the discussion of data classification for an atmospheric trace gas data management system is the introduction of a small sample of air either from the atmosphere or from a standard gas cylinder into an analytical system (Figure 2). Information that uniquely identifies this sample makes up the "sample collection key" (the primary key in this discussion) and will typically include the location, date (GMT), collection technique, identification of the sample container (flask or cylinder serial number or port position), and any additional information that is required to distinguish this sample from previous or subsequent samples. There may, of course, be supporting information that might include flow rates, pressures, and meteorological conditions; while this information may be useful in assessing the quality of the collection event, it is extraneous (i.e., non-key attributes) to the identification of a unique sampling event. This supporting information combined with the sample collection key is classified as "sample collection details." Table 2 shows a portion of two tables containing sample collection details from a discrete and a continuous measurement program. Data generated as a result of the analysis of a trace gas constituent in the sample are classified as "primary analysis details" (Table 3) and as the classification suggests are the "rawest" form of the sample analysis data. Primary analysis details will likely include analysis date and time, analytical system identification, sample response voltages, currents, frequencies, and/or chromatographic parameters, as well as supporting information such as temperature, pressure, flow rates, and operator comments.

While primary analysis details are critical to the determination of the quantity of material in the sample and to the assessment of the quality of the measurement, they are generally not suitable for interpretation of the measurement result. Primary analysis details are generally processed further to produce the relative amount of the trace gas constituent in the sample referenced to some standard scale. Mixing ratios and isotope ratios derived from primary analysis details form a higher classification level termed "secondary analysis details" (Table 4). In addition to the quantity of trace gas in the sample, secondary analysis details will often include information describing the methods used to determine the amount of material in the sample, pertinent standard gas and calibration scale information, and operator comments. For example, in Table 4 (top), the system used in the analysis of the discrete samples from Niwot Ridge is specified as a 2-character alphanumeric. This information combined with the analysis date and time (also included in the table) can be used to reference individual secondary analysis records to the corresponding primary analysis details in the appropriate table.

Primary analysis details are critical to the assessment of the measurement system stability and have been used with rule-based expert systems to "objectively" identify samples that are analyzed during periods of system instability [e.g., *Masarie et al*., 1986]. Measurements that are suspect due to system instability can be identified and assigned an "edit" flag to be included with the secondary analysis details. For example, in Table 4 (bottom), two CH_4 measurements of air collected semi-continuously at Barrow, Alaska, and analyzed using gas chromatography with flame ionization detection have been automatically assigned the "%" rejection flag by a rule-based expert system. The expert software has determined that the measurement system was not performing optimally during the period of measurements because variability among bracketing standard gas peak heights (id?24307 in Table 3, bottom) exceeded acceptable levels.

Sample Collection Details [discrete samples]

site	yr	mo	dy	hr	mn	id	m	lat	lon	alt	wd	ws
UUM	1997	08	06	07	00	6577-66	P	44.45	111.10	914	340	2.0
UUM	1997	08	06	07	00	6578-66	P	44.45	111.10	914	340	2.0
NWR	1997	09	30	17	00	4587-91	N	40.05	-105.58	3475	270	9.0
NWR	1997	09	30	17	00	4588-91	N	40.05	-105.58	3475	270	9.0
OPB	1997	08	21	05	11	4231-91	N	-5.00	-165.08	8	135	8.7
OPB	1997	08	21	05	11	4232-91	N	-5.00	-165.08	8	135	8.7

Sample Collection Details [semi-continuous samples] - CH_4

id	yr	mo	dy	hr	mn	port
24307	1997	01	17	19	30	3
BRW	1997	01	17	19	37	5
24307	1997	01	17	19	45	7
BRW	1997	01	17	19	52	1
24307	1997	01	17	20	00	3
BRW	1997	01	17	20	07	5
24307	1997	01	17	20	15	7
BRW	1997	01	17	20	22	1
24307	1997	01	17	20	30	3
BRW	1997	01	17	20	37	5
24307	1997	01	17	20	45	7
BRW	1997	01	17	20	52	1
24307	1997	01	17	21	00	3
BRW	1997	01	17	21	07	5
24307	1997	01	17	21	15	7

Table 2: An excerpt from tables containing sample collection details from the CMDL cooperative air sampling network (top) and from the CMDL continuous CH_4 measurement program at Barrow, Alaska. The primary key is shown as the set of shaded attributes.

Primary Analysis Details [discrete samples] - CO_2

site	yr	mo	dy	hr	mn	id	m	ayr	amo	ady	ahr	amn	voltage
UUM	1997	08	06	07	00	6577-66	P	1997	10	07	11	08	196.116
TK	1995	01	01	00	00	71618	A	1997	10	07	11	17	168.639
UUM	1997	08	06	07	00	6578-66	P	1997	10	07	11	22	195.527
TK	1995	01	01	00	00	1420	A	1997	10	07	11	30	229.557
NWR	1997	09	30	17	00	4587-91	N	1997	10	07	11	35	220.928
TK	1995	01	01	00	00	1406	A	1997	10	07	11	43	302.393
NWR	1997	09	30	17	00	4588-91	N	1997	10	07	11	48	220.788
TK	1995	01	01	00	00	71618	A	1997	10	07	11	57	168.440
OPB	1997	08	21	05	11	4231-91	N	1997	10	07	12	02	233.222
TK	1995	01	01	00	00	1420	A	1997	10	07	12	10	229.511
OPB	1997	08	21	05	11	4232-91	N	1997	10	07	12	15	233.113
TK	1995	01	01	00	00	1406	A	1997	10	07	12	24	302.142

⋮

Primary Analysis Details [semi-continuous samples] - CH_4

id	yr	mo	dy	hr	mn	port	peak height	peak area	rt	bc
24307	1997	01	17	19	30	3	2.64667E+06	2.50889E+07	122.9	BB
BRW	1997	01	17	19	37	5	2.74599E+06	2.60534E+07	122.9	BB
24307	1997	01	17	19	45	7	2.64855E+06	2.51048E+07	122.9	BB
BRW	1997	01	17	19	52	1	2.74651E+06	2.60466E+07	122.9	BB
24307	1997	01	17	20	00	3	2.64959E+06	2.51195E+07	122.9	BB
BRW	1997	01	17	20	07	5	2.74496E+06	2.60305E+07	122.9	BB
24307	1997	01	17	20	15	7	2.73310E+06	2.60615E+07	122.9	BB
BRW	1997	01	17	20	22	1	2.74532E+06	2.60536E+07	122.9	BB
24307	1997	01	17	20	30	3	2.64860E+06	2.51079E+07	122.8	BB
BRW	1997	01	17	20	37	5	2.75071E+06	2.60793E+07	122.9	BB
24307	1997	01	17	20	45	7	2.64902E+06	2.51053E+07	122.9	BB
BRW	1997	01	17	20	52	1	2.74891E+06	2.60504E+07	122.9	BB
24307	1997	01	17	21	00	3	2.64834E+06	2.51123E+07	122.9	BB
BRW	1997	01	17	21	07	5	2.75032E+06	2.60622E+07	122.9	BB
24307	1997	01	17	21	15	7	2.65069E+06	2.51363E+07	122.9	BB

⋮

Table 3: Top: Portion of a table containing CO_2 primary analysis details from the CMDL cooperative air sampling network. Bottom: Portion of a table containing continuous primary analysis details from the CMDL continuous CH_4 measurement program at Barrow, Alaska. The primary key is shown as the set of shaded attributes.

Secondary Analysis Details [discrete samples] - Niwot Ridge, CO_2

site	yr	mo	dy	hr	mn	id	m	value	flg	in	ayr	amo	ady	ahr	amn
NWR	1967	05	18	19	06	28-67	N	324.85	T..	L1	1967	05	19	08	42
NWR	1967	05	18	19	06	95-67	N	-999.99	*..	L1	1967	05	19	08	00
⋮															
NWR	1997	09	30	17	00	4587-91	N	359.68	...	L3	1997	10	07	11	35
NWR	1997	09	30	17	00	4588-91	N	359.65	...	L3	1997	10	07	11	48
⋮															

Secondary Analysis Details [semi-continuous samples] - Barrow, CH_4

site	yr	mo	dy	hr	mn	CH_4 (pk ht)	CH_4 (pk ar)	port	inst	flag
BRW	1986	02	01	00	12	1761.17	1755.20	5	C3	.
BRW	1986	02	01	00	36	1762.73	1758.00	1	C3	.
BRW	1986	02	01	01	00	-999.99	-999.99	5	C3	*
⋮										
BRW	1997	01	17	20	07	1792.29	1787.50	5	H3	%
BRW	1997	01	17	20	22	1792.86	1789.50	1	H3	%
BRW	1997	01	17	20	37	1824.89	1825.37	5	H3	.
⋮										

Table 4: Top: Portion of a table containing CO_2 secondary analysis details from discrete samples collected at Niwot Ridge, Colorado. Bottom: Portion of a table containing secondary analysis details from the CMDL continuous CH_4 measurement program at Barrow, Alaska. The primary key is shown as the set of shaded attributes. Missing concentrations are identified with an unambiguous default value.

Sample collection, and primary and secondary analysis details described here serve as an example of the most general data classifications. The number of classifications and content depends completely on the measurement application. For example, some GC measurement systems incorporate an integrating instrument which determines chromatographic parameters such as peak height and area, retention time, and baseline code during the measurement. For these systems, the chromatographic parameters become part of the primary analysis details. Other GC measurement systems generate a distribution of frequencies or voltages which together with integration software determine the chromatographic parameters. Because the chromatographic peak integration can be done during and/or after sample detection, the distribution of frequencies or voltages may be considered primary analysis details, while the chromatographic parameters determined by the integration software may be grouped as secondary analysis details. In this latter case, a third-order classification will be required. As a second example, consider a continuous measurement program where the volume of data is great. While the classification of primary analysis details is unchanged, secondary analysis details may consist of mixing ratios determined for each ambient sample. Third-order analysis details may contain hourly averages (Table 5, bottom), while subsequent classifications may consist of daily or monthly-averaged values. It is important to note that data classified as secondary analysis details in one measurement program may be considered third or fourth-order analysis details in another. This does not present a problem provided the data classifications are consistent within each measurement program.

Another important data classification is "meta data." Meta data is information about data themselves and includes any information that is relevant to an assessment of the quality of a measurement. Some of these details may already be attributes of data tables but more often this information is manually logged into laboratory notebooks and may include comments regarding replacing or modifying system components, changing traps, changes in system performance, or making procedural changes. These seemingly minor details can often get lost within large measurement operations where several scientists and technicians may be required to maintain and operate several systems critical to each stage of the measurement process. Admittedly, CCGG has experienced this problem. For example, to understand changes in the performance of a CO_2 flask measurement system, a complete reconstruction of the system's modification history was required. The resulting 10-year reconstruction was incomplete because information most likely written into a notebook could not be located. This difficult learning experience emphasizes the importance of organizing, maintaining, and centralizing meta data. Meta data do not necessarily have a flow or hierarchy making it difficult to establish guidelines. Rather, meta data tend to be coupled with components of the measurement program. For example, the collection of a discrete air sample into a storage container is one component in the measurement procedure. Portable sampling units (PSU) and collection methods used by CCGG continue to evolve sometimes in incremental steps and sometimes dramatically with the implementation of a completely new design. Details of the overall development process including test results, known problems, and design modifications should be maintained in a meta data table. Another table may be useful to document PSU changes at each sampling location. For example, a table of PSU details (Table 6) provides a history of units employed at each sampling location. Also from this table, a history of PSU deployment can be constructed by searching the table using specific PSU "type" AND "#" attributes. Each sampling location should also have a meta data table that documents sampling unit modifications and replacements, personnel and procedural changes, and geographical changes, all critical factors in assessing the quality of a collection event and resulting measurements. While it is difficult to set guidelines for the management of meta data, RDBMS architects should appreciate the importance of this information and consider how best to incorporate these data into the RDBMS.

3.3 User Requirements

The assessment of data flow and classification of information build a general framework for the RDBMS strategy. Database architects, administrators, and users must next identify the short and long-term needs and goals of the laboratory. Database architects may find that based on users' needs, data classification and/or flow of information may require modifications. This

Third-Order Analysis Details [discrete samples] - Niwot Ridge, All species

site	yr	mo	dy	hr	mn	id	m	lat	lon	alt	wd	ws	CO_2	flg	?^{13}C(CO_2)	flg	?^{18}O(CO_2)	flg	CH_4	flg	CO	flg	H_2	flg	N_2O	flg	SF_6	flg
NWR	1967	05	18	19	06	28-67	N	40.05	-105.63	3475	999	-9.9	324.85	T..	-999.999	*..	-999.999	*..	-999.99	*..	-999.99	*..	-999.99	*..	-999.99	*..	-999.99	*..
NWR	1967	05	18	19	06	95-67	N	40.05	-105.63	3475	999	-9.9	-999.99	*..	-999.999	*..	-999.999	*..	-999.99	*..	-999.99	*..	-999.99	*..	-999.99	*..	-999.99	*..
⋮																												
NWR	1997	09	30	17	00	4587-91	N	40.05	-105.58	3475	270	9.0	359.68	...	-7.726	...	-0.343	...	1788.49	...	98.75	...	445.84	...	314.34	...	3.93	...
NWR	1997	09	30	17	00	4588-91	N	40.05	-105.58	3475	270	9.0	359.65	...	-7.702	...	-0.351	...	1787.39	...	99.93	...	445.48	...	313.97	...	3.93	...
⋮																												

Third-Order Analysis Details [semi-continuous samples] - Barrow, CH_4

site	yr	mo	dy	hr	CH_4
BRW	1986	02	01	00	1761.95
BRW	1986	02	01	01	1769.02
BRW	1986	02	01	02	1772.72
⋮					
BRW	1997	01	17	20	1824.28
BRW	1997	01	17	21	1827.16
BRW	1997	01	17	22	1839.43
⋮					

Table 5: Top: Portion of a table containing sample collections details from discrete samples collected at Niwot Ridge, Colorado and analysis summaries for all trace gas species currently measured by CMDL CCGG. Bottom: Portion of a table containing hourly average mixing ratios from the CMDL continuous CH_4 measurement program at Barrow, Alaska. In both tables, the primary key is shown as the set of shaded attributes. Missing concentrations are identified with an unambiguous default value.

Portable Sampling Unit Details

	PSU		shipped			begin use			end use			received			
site	type	#	yr	mo	dy	yr	mo	dy	yr	mo	dy	yr	mo	dy	comments
CBA	P3	-99	1981	01	01	1981	11	19	1991	06	13	1991	07	15	Fill pressure set to ~10 psig
CBA	MAKS	22	1991	05	29	1991	06	16	1996	07	07				Fill pressure set to ~10 psig
CBA	AK	04	1996	06	12	1996	07	11							Excludes condenser, f.p. set ~4 psig
CGO	P3	-99	1984	01	01	1984	04	19	1990	08	07	1990	09	15	Fill pressure set to ~10 psig
CGO	MAKS	01	1990	05	25	1990	08	16	1995	02	08	1995	05	12	Fill pressure set to ~10 psig
CGO	MAKS	50	1994	11	30	1995	02	17							Fill pressure set to ~4 psig
CGO	AK	06	1997	05	01	1997	07	15							Fill pressure set to ~4 psig
CMO	P3	-99	1982	01	01	1982	03	10	1991	04	22	1991	05	15	Fill pressure set to ~10 psig
CMO	MAKS	32	1991	04	17	1991	04	22	1993	07	17	1993	09	16	Fill pressure set to ~10 psig
:															

Table 6: An excerpt from a meta data table containing the history of portable sampling units used at CMDL cooperative air sampling locations. The primary key is shown as the set of shaded attributes.

exercise helps to further define the database management strategy and begins to place serious constraints on the system design.

Observational data serve many purposes. Primary analysis details are useful in the assessment of sample collection and analytical system performance and are typically of interest to only a few people. Data processed further are critical to interpretative studies where the focus may be on a single trace gas species or on correlations among several trace gas constituents (Table 5, top). These data are of interest to a wider audience, particularly for comparative studies and in the construction of cooperative data sets for use in carbon cycle modeling studies [e.g., *GLOBALVIEW-CO_2*, 2002; *GLOBALVIEW-CH_4*, 2001]. Sample collection details which describe a sampling event are critical in most applications of the data and are consequently of interest to the greatest number of users.

The carbon cycle measurement community has an additional obligation to make measurement records readily available beyond the internal organization to both national and international scientific communities and to the general public (via WWW and anonymous FTP distribution). Thus an atmospheric trace gas RDBMS must serve internal needs as well as facilitate the timely dissemination of atmospheric measurements.

3.3.1 Accessibility

Data tables must be readily accessible. Users need direct access to the data tables using operating system commands, simple programs written in the C, FORTRAN, and BASIC programming languages, and commercial data visualization and analysis packages such as Mathematica, S-Plus, and IDL. Data made available to the wider community must also be easily accessible.

It is equally important that users access the actual tables and not copies of the tables. Tables within the RDBMS will be maintained by the RDBMS and will reflect the most current information about an air sample including adjustments resulting from standard gas or calibration scale changes. Copies of tables are not generally maintained by the RDBMS and thus do not reflect the dynamically changing contents of the RDBMS. The practice of using copies of data tables for interpretive studies can have serious consequences as these data can quickly become "stale." To discourage this practice, the RDBMS should reside on a database server, a central computer that has been configured to provide easy database access to appropriate users from any location.

While ready access to data tables is critical, so too is the speed at which data are retrieved. Table access speed is both a computer hardware and RDBMS design consideration. Because the volume of data from ongoing systematic atmospheric observations steadily increases, it is prudent to develop a strategy and hardware platform that can accommodate many times more data than currently exists.

3.3.2 Portability

Closely tied to the issue of accessibility is portability. Because of the growing interest in high-precision atmospheric measurements of carbon cycle trace gas species particularly for use with modeling studies, a measurement program becomes increasingly more valuable if its measurement records can be readily used by other research organizations. This requires that the data tables themselves be portable to a variety of common hardware platforms and software applications.

Data tables can be stored on the server as unformatted or formatted files. Unformatted files (also termed binary files) contain the internal binary representation of the data directly from the computer's memory. While saving data to an unformatted file is the most efficient method of storing information, it is also the least portable method and always requires a translator (computer program) to view the file's contents. Formatted files (also termed ASCII or text files) contain data that have been converted from binary representation to ASCII character representation. While saving data to a formatted file is less efficient, formatted files are extremely portable across hardware and software platforms and are easily read using operating system commands and simple text editors.

For example, data sets managed by a commercial RDBMS such as Paradox are generally saved as unformatted files which are directly accessible to only those organizations that also happen to use Paradox or that use another RDBMS that can read Paradox tables. Typically, packages like Paradox can save tables in the more common ASCII file format, but this action has disadvantages. First, translations result in copies of tables that are no longer maintained by the RDBMS. Because table translations require effort, RDBMS administrators may be unduly burdened by data requests that require this extra step. Further, translations do not always guarantee portability. Additional efforts may be required to ensure that data items are easily distinguishable and that missing items are properly identified.

At the moment, a RDBMS that stores tables as formatted files will ensure the greatest level of portability. Database architects should take care to design data tables and select a RDBMS package that will increase the value of their data product by ensuring that the measurement records can be easily used by the many institutions and agencies working on carbon cycle issues.

3.3.3 Readability

Because data tables are required to be readily accessible and portable, their contents must be unambiguous. Tables are required to be appropriately formatted with attributes that are clearly defined and follow universally-accepted protocol where possible. Because carbon cycle data produced by one laboratory may be widely distributed and often integrated with data from other laboratories, it is important that the format of attributes be standardized so that the use and integration of data are not impeded by format incompatibilities and ambiguities. Perhaps the easiest attribute to standardize is sample collection date. In efforts to coordinate and construct cooperative global data sets, CCGG has found dates specified as year, month, day, hour, minute, and second in Greenwich Mean Time (GMT = UTC) to be most common and least ambiguous. For example, the date

```
1997 02 26 23 58 31
```

specifies February 26, 1997 at hour 23 and 58 minutes and 31 seconds in GMT. The format consists of a 4-digit year, followed by a 2-digit month, day, hour, minute, and second with each field separated by at least one space and including leading zeros (0) where required. When the date is specified with hourly resolution, for example, “1997 02 26 23”, the hour “23” consists of information for hour 23 beginning at minute 00, and second 00 and ending at hour 23, minute 59, and second 59. Thus, hours range from 00 to 23, minutes from 00-59, and seconds 00-59. In contrast, day-of-year or Julian dates (6267.998969908 in days since January 1, 1980) and decimal year dates (1997.156161561) are less common and subject to misinterpretation. Julian dates are referenced to an arbitrary starting date defined by the data author. Decimal year dates are convenient when plotting data, but when the decimal year has minute resolution, conversion to a more standard format becomes dependent on the machine-precision of the computer. Dates in either format can be difficult to interpret without first converting to a more common representation which requires knowledge of how the conversion was made including considerations for leap years.

A general characteristic of all attributes that is also easy to standardize is the identification of missing data items within a record. Missing data are handled in a variety of ways which seem to depend mostly on the software package that is used as the RDBMS. Again, in the context of integrating measurement records from many different laboratories, CMDL has found that it is relatively straightforward to make use of a data set when missing values are identified by an unambiguous default value that is clearly defined and consistent throughout the RDBMS. CCGG has adopted “-999.99[9]” as a default value for missing information in a concentration field.

Specifically, this value indicates that the air sample was analyzed but a relative quantity could not be determined. When an air sample is not analyzed, e.g., when a container breaks during analysis or when a measurement system is not functioning, the sample does not have an entry in any table containing those analysis details. Different and more appropriate default values are used in other record fields. Operator comments would be an exception to this requirement (see discussion on meta data, Section 3.2). Tables that use a null or blank character to identify fields with missing data are most difficult due to the way in which these characters are translated between hardware environments and software applications.

Data tables can be saved as either “free” or “explicitly” formatted files. “Free” format protocol has minimal structure and requires little effort during RDBMS development, but sacrifices readability. “Explicitly” formatted files can maximize readability but will require a more detailed description of the tables’ contents. The secondary analysis details from the CMDL discrete samples at Niwot Ridge, Colorado shown in Table 4 (top) are used to illustrate the broad range of formatting protocol.

(example 1)
```
NWR 1967 5 18 19 6 28-67 N 324.85 T.. L1 1967 5 19 8 42
NWR 1967 5 18 19 6 95-67 N *.. L1 1967 5 19 8 0
NWR 1997 9 30 17 0 4587-91 N 359.68 ... L3 1997 10 7 11 35
NWR 1997 9 30 17 0 4588-91 N 359.65 ... L3 1997 10 7 11 48
```

(example 2)
```
NWR,1967,5,18,19,6,28-67,N,324.85,T..,L1,1967,5,19,8,42
NWR,1967,5,18,19,6,95-67,N,,*..,L1,1967,5,19,8,0
NWR,1997,9,30,17,0,4587-91,N,359.68,...,L3,1997,10,7,11,35
NWR,1997,9,30,17,0,4588-91,N,359.65,...,L3,1997,10,7,11,48
```

(example 3)
```
NWR 1967 5 18 19 6 28-67 N 324.85 T.. L1 1967 5 19 8 42
NWR 1967 5 18 19 6 95-67 N -999.99 *.. L1 1967 5 19 8 0
NWR 1997 9 30 17 0 4587-91 N 359.68 ... L3 1997 10 7 11 35
NWR 1997 9 30 17 0 4588-91 N 359.65 ... L3 1997 10 7 11 48
```

(example 4)
```
NWR,1967,5,18,19,6,28-67,N,324.85,T..,L1,1967,5,19,8,42
NWR,1967,5,18,19,6,95-67,N,-999.99,*..,L1,1967,5,19,8,0
NWR,1997,9,30,17,0,4587-91,N,359.68,...,L3,1997,10,7,11,35
NWR,1997,9,30,17,0,4588-91,N,359.65,...,L3,1997,10,7,11,48
```

(example 5)
```
NWR,1967,05,18,19,06,28-67,N,324.85,T..,L1,1967,05,19,08,42
NWR,1967,05,18,19,06,95-67,N,-999.99,*..,L1,1967,05,19,08,00
NWR,1997,09,30,17,00,4587-91,N,359.68,...,L3,1997,10,07,11,35
NWR,1997,09,30,17,00,4588-91,N,359.65,...,L3,1997,10,07,11,48
```

(example 6)
```
NWR 1967 05 18 19 06   28-67 N   324.85 T.. L1 1967 05 19 08 42
NWR 1967 05 18 19 06   95-67 N  -999.99 *.. L1 1967 05 19 08 00
NWR 1997 09 30 17 00 4587-91 N   359.68 ... L3 1997 10 07 11 35
NWR 1997 09 30 17 00 4588-91 N   359.65 ... L3 1997 10 07 11 48
```

It is clear from the above examples that as the level of formatting increases, readability also improves.

Finally, data tables must be complete and able to convey enough information to stand on their own. RDBMS developers must find a balance between including too much detail in a table and not including enough. As an extreme example, viewing a table containing only mixing ratios and quality assurance flags without the associated sample collection details will have limited usefulness as the following example demonstrates:

```
 324.85 T..
-999.99 *..
 359.68 ...
 359.65 ...
```

3.3.4 Data Security

The security of data managed by the RDBMS is the responsibility of the RDBMS administrator(s). Access to data tables should be restricted so that only appropriate individuals, projects, or groups can modify and/or read selected tables. It is critical that the RDBMS that maintains a measurement program's most valuable asset, data, be protected against loss due to hardware failure, user error, or natural disaster. The RDBMS should be easily incorporated into existing routine and automatic mass storage backup procedures. RDBMS backups should be frequent, well documented, and stored in secure locations both on and off site.

3.3.5 Data Integrity

A RDBMS that is secure does not necessarily guarantee that data cannot be compromised. Experience proves that both computer hardware and humans can potentially and inadvertently corrupt data. The RDBMS must accommodate routine and automatic checks of the tables to ensure that the integrity and consistency of data have not been compromised.

3.3.6 Quality Assurance

Every measurement program must have methods in place to assess the quality of the data produced. Problems during sample collection or analysis often result in computed mixing ratios that have been compromised. Occasionally, samples appear to have no apparent analytical or sampling problems yet yield values that deviate substantially from "typical" observations, and thus, appear to be not representative of "background" conditions. Conveying information about the quality of a measurement is critical to those individuals making the measurements (as diagnostic data) as well as those trying to interpret the data (as representative data). This information, generally encoded into a "flag", accompanies a mixing ratio or isotope ratio in tables of secondary or higher-order analysis details (see Tables 4 and 5). RDBMS architects and users together should design a flagging scheme that can easily and adequately convey the reason(s) why the quality of a measurement is being questioned.

The RDBMS should provide a mechanism whereby flags can be assigned manually or automatically and a strategy to ensure that once assigned, flags cannot be inadvertently changed. Assigning flags manually has traditionally been problematic with flagging criteria often applied subjectively and inconsistently. Standards used to assess measurement quality may differ from person-to-person or even from day-to-day by the same person. The need to apply flags manually can be minimized with the use of diagnostic software designed to detect analytical problems (see section 3.2) and selection software designed to identify "outliers" using statistical techniques. However, it is difficult to devise an automatic flagging routine able to detect all possible scenarios in which a sample can be compromised, and thus the need to apply flags manually is most likely unavoidable. A well designed quality assurance strategy can keep the number of different types of manual flags to a minimum and require that each flag type is well defined so that users can easily understand the reason why a manual flag was assigned to a measurement result.

3.3.7 Automation

There are many components of a measurement program that can benefit from a certain level of automation. For example, analytical systems should have direct access to the RDBMS so that data tables can be automatically updated as samples are analyzed. Manual data entry should be avoided wherever possible. If data must be manually entered into the database, RDBMS architects should ensure that these details are entered only once and subsequently made centrally available.

Perhaps the most critical component of a data management strategy that requires automation is the re-processing of primary analysis details required when standard gas values have changed and calibration scales are adjusted. The maintenance and propagation of laboratory standard gases is critical to the measurement program and can quickly become an overwhelming task if not properly included into the overall data management strategy. The level of automation should, at the very minimum, be maintained, but preferably enhanced with the implementation of a RDBMS.

3.3.8 Flexibility

In addition to meeting the above design criteria, the selected RDBMS software package must be easily modifiable, expandable and flexible to meet the growing and changing needs of the measurement program. For example, the CMDL cooperative air sampling network continues to evolve with sites being periodically added or replaced. In addition, the development of new laboratory techniques enables CCGG to add additional atmospheric compounds to its current suite of measurements. These and other developments should require only modest modifications to a RDBMS.

3.3.9 Simplicity

Finally, the overall design of the RDBMS must be simple. While database administrators will manage the RDBMS, the RDBMS itself should be conceptually simple so that users know what data are available, can have a working knowledge of the data, and project leaders of developing programs can understand and work within the adopted data management strategy allowing any new data types to be easily incorporated and managed by the RDBMS. It is important that the RDBMS not serve as a “black box” where the inner workings of the system are obscure and understood by, in a worst-case scenario, a single person. A simple yet well-conceived design strategy will likely yield a more robust and enduring product that can potentially grow with the changing needs and personnel of the atmospheric measurement program.

4. RDBMS SELECTION

With a carbon cycle data management strategy firmly established, the next step is selecting the appropriate RDBMS. There are a few options: purchase a commercially available software package, contract the development of the RDBMS with an outside software firm, or develop the package in-house. Major issues in selecting a RDBMS are cost, company reliability and continuity, readily available technical support, and minimal disruption accompanying product upgrades. It is imperative that the adopted package meets or exceeds all basic performance and strategy requirements and minimizes extra features that are either redundant or superfluous. Finally, the selected RDBMS “package” should include training so that proficiency is quickly achieved by the database architects, administrators, and users. If each option meets the design criteria then the strengths and weaknesses of each option should be considered.

4.1 Commercially Available RDBMS

There is certainly an advantage in purchasing a commercially available RDBMS from an established software manufacturer knowing that the product has been developed by a team of expert database software engineers and has improved with time. However, despite company reputation and product longevity, it will still be critical to conduct a full suite of tests using the prospective RDBMS with a volume of information that substantially exceeds current needs. Not all RDBMS packages are suitably designed to manage scientific data. Many packages have been designed for business applications and have only been "adapted" for technical applications. If a RDBMS requires manufacturer modifications to meet all design criteria, be certain that these modifications are made prior to the purchase or that the estimated cost and completion date for the required modifications have been settled prior to the purchase. Finally, "bells-and-whistles" are an added bonus only after the package meets or exceeds the data management design criteria.

4.2 Contracting RDBMS Development

Contracting the development of the RDBMS to a software company also has advantages. Generally, these companies are small enabling one or two database consultants with expertise in working with clients and understanding their needs to oversee the project. This personal attention will likely mean that the RDBMS can be customized to meet the specific needs of the measurement program. But because consultants' time is costly, it will be important to develop a preliminary strategy beforehand (see section 3). Again, the product should be fully tested and debugged, but this cannot occur before a significant financial investment in the project has been made. Under these conditions, it is imperative that technical support extends for quite some time beyond the delivery date as software bugs will likely continue to surface requiring "fixes" and modifications. Ideally, a programmer within the laboratory can be sufficiently trained to maintain and modify the RDBMS package as the measurement program grows and data management needs change. It is prudent to determine ahead of time if the company's policy supports the training of customers and the eventual release of its source code.

4.3 In-House RDBMS Development

First and foremost in the decision to develop a RDBMS in-house is determining if the required level of programming and data management expertise exists. The RDBMS architect will likely be a software engineer with experience in RDBMS and an understanding of the laboratory's data management needs. Excellent communication and project management skills are required to orchestrate a project that will undoubtedly require input from many people. Developing and maintaining an in-house RDBMS package will take considerable time and require a long-term commitment. Costs can be controlled somewhat by trying not to "re-invent the wheel" and avoiding common RDBMS pitfalls. Visits to laboratories that have implemented RDBMS packages that perform similar tasks may be helpful to gain additional perspectives on data management strategy and implementation. It may be worthwhile hiring a consultant to offer guidance during the development phase of the data management strategy to ensure that the RDBMS is indeed based on a sound strategy.

The in-house RDBMS will likely consist of one or more computer programs and many data tables. The RDBMS architect will need to determine the programming language and operating system most suitable to the data management task. Critical issues, already mentioned, include cost, reliability, available technical support, and minimal disruption accompanying software upgrades.

4.4 RDBMS Caveats

Selecting a commercial RDBMS is a difficult task. Even the most thorough evaluation cannot detect unforeseen events. For example, in the early 1990s, after an exhaustive assessment of the RDBMS package Paradox, CSIRO GASLAB members were confident that this package would meet their data management needs. Much of their discrete analysis processing code was then written in the Paradox programming language and their data tables had been managed successfully by this RDBMS for several years. Following the release of a new version of Paradox in 1995, GASLAB found that while their data tables ported successfully to the product upgrade, their programs did not. For several years, GASLAB was "forced" to maintain the older version of Paradox, which was no longer supported, until they were able to commit resources to convert these processing routines to a more stable platform.

Spreadsheets such as Excel, Lotus 1-2-3, and Xess are not considered in the RDBMS selection process as they are not relational database management applications and thus cannot satisfy database design criteria. While spreadsheets have many useful features for manipulating data within worksheets, they simply are not designed as an RDBMS and are not equipped to manage the growing volume of data produced by long-term monitoring programs. The Institute for Arctic and Alpine Research (INSTAAR) Stable Isotope Laboratory at the University of Colorado, which measures the isotopic ratios of CO_2 in all CCGG discrete samples, learned that after 2 years of measuring a small subset of samples from the CMDL network, the spreadsheet program Excel was an inappropriate package to use with a growing data set. Access to their ?^{13}C and ?^{18}O measurements slowed and data manipulation became increasingly difficult. Converting their data tables to a proper RDBMS was a painful but necessary task that took nearly one full year to complete.

The Carbon Cycle Greenhouse Gases Group has developed an in-house RDBMS that has been operational since 1992. The RDBMS consists of several computer programs written in the C and OSF/Motif programming languages residing on a central personal computer (PC) file server running the Red Hat Linux operating system. Since 1992, computers, operating systems, and programming languages have been upgraded several times. On more than one occasion following an upgrade, modifications to the existing RDBMS programs were required. While changes have always been minor, architects developing an in-house RDBMS should consider a programming language and style that is likely to withstand computer system upgrades.

These examples illustrate that despite efforts to select the best database management package for the measurement program, problems can and do still arise. Those faced with the task of developing a data management strategy and selecting a RDBMS may learn from these experiences and move more quickly towards implementing a robust and enduring RDBMS.

5. CMDL CCGG RDBMS

The Carbon Cycle Greenhouse Gases Group operates both discrete and semi-continuous long-term monitoring programs. Discrete samples are collected approximately weekly from 51 sites of the cooperative air sampling network and from 9 sites of the aircraft sampling program. Continuous measurements are made at the 4 CMDL baseline observatories and at the tall tower sites in Wisconsin and Texas. The sampling programs are summarized in Figure 3. An important feature of the discrete sampling programs is that each sample provides a snapshot in time of the atmospheric mixing ratio of several important trace gas constituents. Spatial and temporal patterns in the observations and persistent correlations among these trace gas species provide clues as to the source and strength of emissions and the effect of dilution from atmospheric mixing. Measurements of CO_2 and ?^{13}C (CO_2) from network samples, for example, have been used with transport models in an effort to better understand the partitioning of CO_2 between the ocean and terrestrial biosphere [*Ciais et al.*, 1995].

Nearly 647,000 measurements from over 150,000 flask air samples have been made since the cooperative air sampling network's inception in the late 1970s. An ongoing effort is necessary to ensure that these data are highly organized so that measurements of any species from any sampling location can be readily accessed. Prior to 1992, a proper data management strategy had not yet been formulated. As a result, it was difficult, for example, to match measurements of CO_2 with measurements of CH_4 and CO from the same samples.

In 1991, CCGG decided to develop, in-house, a RDBMS package capable of managing all information associated with the cooperative air sampling network program. The resulting RDBMS package has successfully managed information derived from discrete sampling from fixed sites, moving sites (aircraft and shipboard), and special projects (firn ice). While specifically designed to manage data from a discrete sampling program, the data management strategy is straightforward

Discrete Sampling Programs

?Cooperative Air Sampling Network

CO_2	NDIR
CH_4	GC-FID
CO, H_2	GC-HgO
$?^{13}C$, $?^{18}O$ (CO_2)	On-line extraction-MS
N_2O, SF_6	GC-ECD
$?^{13}C$ (CH_4)	Continuous Flow-MS

?Aircraft Network

CO_2	NDIR
CH_4	GC-FID
CO, H_2	GC-HgO
$?^{13}C$, $?^{18}O$ (CO_2)	On-line extraction-MS
N_2O, SF_6	GC-ECD

Semi-continuous Sampling Programs

?CMDL Baseline Observatories

Barrow, Alaska
Mauna Loa, Hawaii

CO_2	NDIR
CH_4	GC-FID
CO, H_2	GC-HgO

American Samoa
South Pole

CO_2	NDIR

?Tall Towers

CO_2	NDIR

Figure 3: Summary of CMDL CCGG measurement programs. For each program, the suite of trace gas species measured and the method of detection are listed.

and can easily interface with the database management strategies employed for the continuous measurement programs. A description of CCGG's approach to managing information derived from the sampling network may be useful to laboratories in the process of developing a data management strategy and should also serve to illustrate the strengths and weaknesses of this in-house approach.

5.1 Description

The CCGG RDBMS resides on a PC network file server running the Red Hat Linux operating system. At the core of the CCGG RDBMS are two computer programs written in the C programming language. A "point-and-click" graphical user interface written in the OSF/Motif windowing language was included with both programs in an attempt to make them as "friendly" and robust as possible. The first RDBMS program, depicted in Figure 4, has two functions. The first component enables CCGG to track the shipping and receiving of nearly 6000 sample containers with the aim of ensuring that each of the 51 cooperative sampling locations has an adequate flask supply. A flask inventory and network description details are also maintained. The second component provides a "point-of-entry" for the sample collection details recorded by the field operator during sampling. The entry of sampling location, container identification, and position

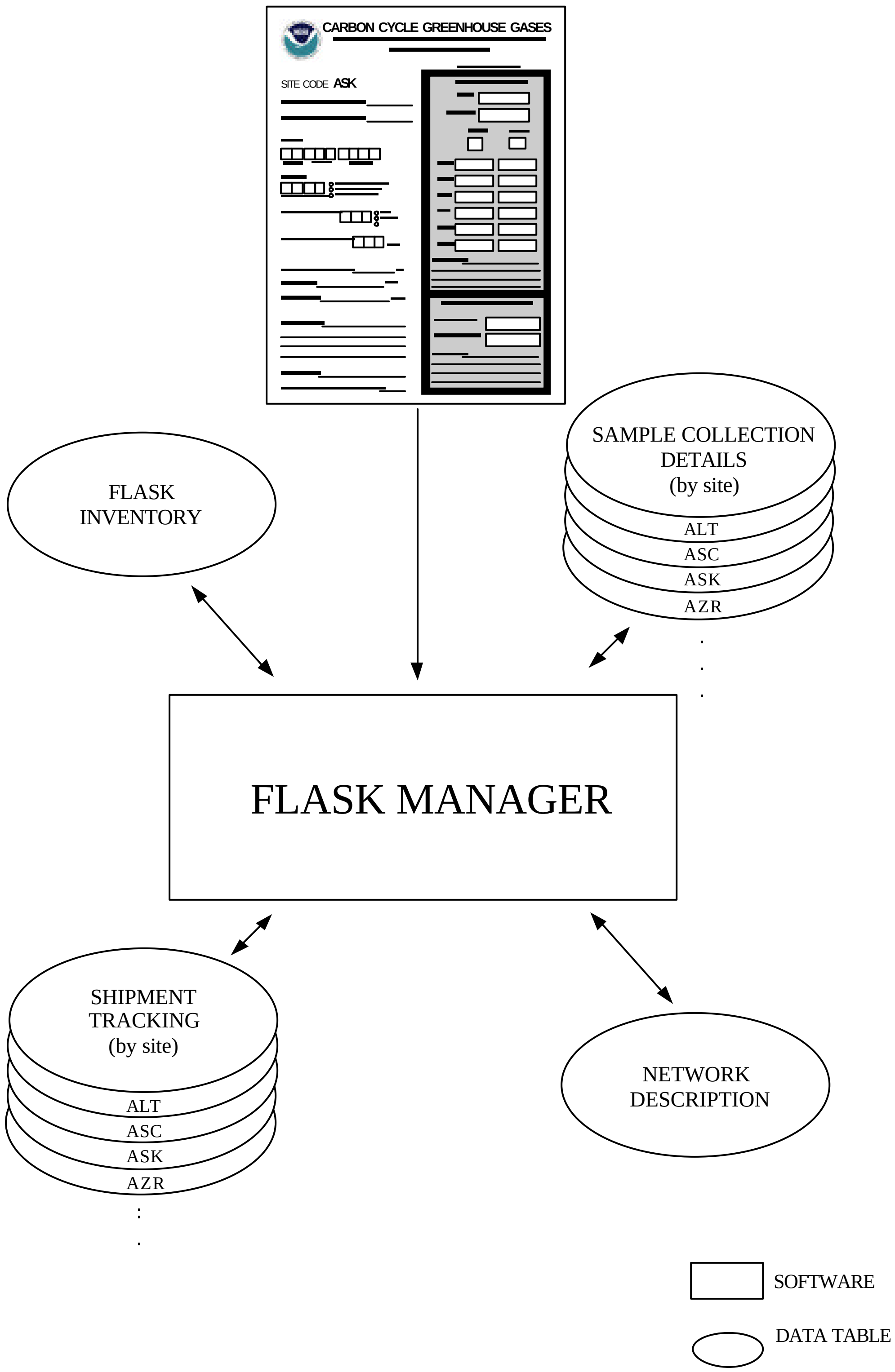

Figure 4: Schematic of the CCGG flask manager program functions. The RDBMS computer program manages the network flask inventory, tracks flask shipments, provides the "point-of-entry" for network sample collection information, and appends tables containing network sample collection details.

(fixed sites) is automatic, but at the moment, the remaining sample collection details [date and time (GMT), collection method, position (moving sites), and meteorological parameters] are manually entered into the RDBMS by CCGG technician. Manual data entry is a weakness in the CMDL approach and has been a concern for many years but because of logistical difficulties and operating costs in managing the cooperative network, it remains the preferred method of entry into the database. Once the sample collection details are entered into the RDBMS, they are readily accessible to any user or computer program. For example, the two independent analytical systems used to make measurements of CO_2, CH_4, CO, H_2, N_2O and SF_6; and ?^{13}C and ?18O of CO_2 directly access the sample collection details and use the primary key from these tables to construct the species-dependent tables containing analysis details. When an entry error does occur in one or more attributes of the collection details, this error is propagated to all tables that include these attributes. Because the error occurs everywhere, the database remains consistent albeit consistently incorrect. For example, if a sample collected "1997 08 11" at Alert, Canada is entered into the RDBMS as being collected "1997 11 08" then the date will be incorrect throughout the RDBMS. There is no opportunity for the date to be correct in some tables and not in others. This approach guarantees that consistency throughout the RDBMS is maintained and greatly simplifies the task of correcting data entry errors, the purpose of the second RDBMS program.

The second RDBMS program (Figure 5) enables a database administrator to identify incorrect attributes in sample collection details and to instruct the program to correct all data tables containing the incorrect attributes. For example, in rectifying the data entry error associated with the Alert sample described above, 17 data tables will be corrected (two for each of the eight measured compounds and the sample collection details). Data entry errors are generally corrected biweekly. Because all data tables reside on a single server, the corrected tables are immediately available to users and computer programs that directly access the data.

Both programs create and edit attributes of the sample collection details and are the only two programs that exist capable of performing these tasks. Because the integrity of the RDBMS depends, in part, on the consistency of information throughout the data tables, these programs are self-documenting, maintaining detailed accounts of operator actions and collection detail corrections.

Several additional programs directly access the sample collection details and have the ability to create, append, and edit tables containing primary and secondary analysis details. The two independent analytical systems used to make measurements of CO_2, CH_4, CO, H_2, N_2O and SF_6; and ?^{13}C and ?^{18}O of CO_2 have already been mentioned. Additional computer programs exist that enable users to manually or systematically apply edit and selection flags to tables of secondary analysis details. For example, users can manually associate a "sample quality" flag with analysis results that are suspect due to sample collection or analytical problems. Or, users can invoke a computer program that systematically applies qualification flags to measurement results within a time series that have been statistically identified as values that are not representative of "background" conditions.

5.2 Data Flow

The flow of information for network samples follows the flow scheme shown in Figure 2 and described in section 3.1. Several data tables are integral to the progression of measurement information from its rawest form to its most processed form. The tables have been described previously and include the cooperative air sampling network details (Table 1, top), sample collection details (Table 2, top), primary analysis details (Table 3, top), secondary analysis details (Table 4, top), and third-order analysis details (Table 5, top).

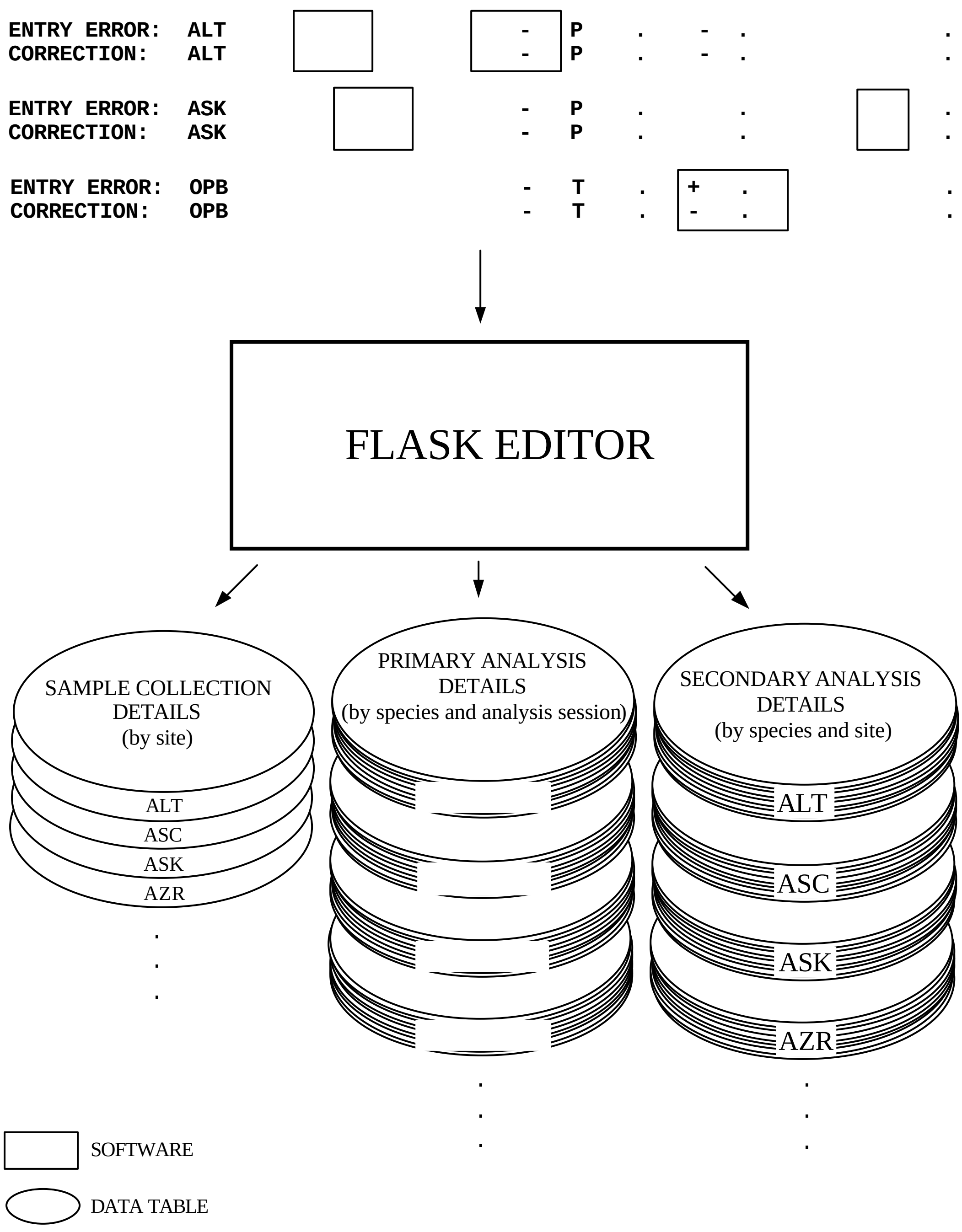

Figure 5: Schematic of the Carbon CCGG flask editor program functions. The RDBMS computer program enables data entry errors in the RDBMS to be corrected. Typical data entry errors are shown. Corrections are made to any data table containing data errors. At the moment, one data entry error may require corrections to as many as 17 data tables.

5.3 Data Classification

The classification of data and table content derived from the cooperative air sampling network follows closely the discussion in section 3.2. However, due to the nature of the cooperative network where samples are collected at many locations and analyzed for many trace gas species, data records and tables are further organized in an effort to maximize data access efficiency. Tables of analysis details managed by the RDBMS are grouped first by the trace gas constituent (CO_2, ?^{13}C and ?^{18}O of CO_2, CH_4, CO, H_2, N_2O, and SF_6), then by the measurement program (continuous or discrete), and finally by sampling location (BRW, MLO, etc.).

Understanding how network samples are analyzed should help clarify the logic behind the RDBMS grouping of data tables. Figure 6 is a schematic describing a typical analysis session from one of the two independent analytical systems. During the session, ~30 air samples from many locations are mounted onto a manifold and subsequently analyzed for CO_2 CH_4, CO, H_2, N_2O, and SF_6. From this single analysis session, six tables are created each containing the primary analysis details for one of the above species. Thus, primary analysis tables are grouped first by species and then by analysis session. During a session, several measurements of laboratory standard gas from high-pressure cylinders are made which help characterize instrument response (which may vary from session to session) and reference computed ambient mixing ratios to the current calibration scale. Primary analysis details from ambient samples are automatically processed further to create secondary analysis details. At this point, records are appended to site-specific tables making long-term site-specific measurement records readily accessible for time series analysis. This grouping is somewhat historic because the measurement of additional trace gas species in network samples traditionally followed the development of a new measurement program within CCGG. However, it does enable individual programs and group members to "exercise" their data with confidence that tables outside their specific program are not affected.

Figure 7 illustrates the grouping of data tables containing secondary analysis details from measurements of samples collected at a network site. Each air sample collected can potentially have a related entry in tables containing secondary analysis details for each trace gas constituent measured. Hence, the number of rows in a table of secondary analysis details can vary among the measured compounds depending on when sampling at the location began and when measurements of a particular compound were added to the suite of measurements. For example, sampling at Cold Bay, Alaska began in 1978 for the measurement of CO_2; the table of CO_2 secondary analysis details at Cold Bay has the greatest number rows, followed by CH_4 (measurements began in 1983), CO and H_2 (1988), ?^{13}C and ?^{18}O of CO_2 (1992), N_2O and SF_6 (1995), and ?^{13}C of CH_4 (2000). Table contents are identical to those shown in Tables 2 and 4. The sample collection key is used to identify all analysis results for a given sampling event. These results are combined into third-order analysis details similar to those shown for NWR in Table 5 (top).

Finally, tables of secondary or higher-order analysis details are grouped by measurement program. The schematic in Figure 8 shows this grouping of data for each CCGG measurement program.

This organization of tables has several advantages. Correlative studies are facilitated by the ready access to the full suite of trace gas measurement data via the sample collection key. Measurement records grouped by sampling location are readily accessible for use in time series analyses. Software applications developed or purchased to investigate one trace gas constituent are readily available for use with all other trace gas measurements because of consistent table design. In addition, proficiency as a result of using an application for one study, improves efficiency in subsequent studies. Because CCGG measurement programs employ standardized attribute formatting protocol, all discrete and continuous measurement results can be readily compared.

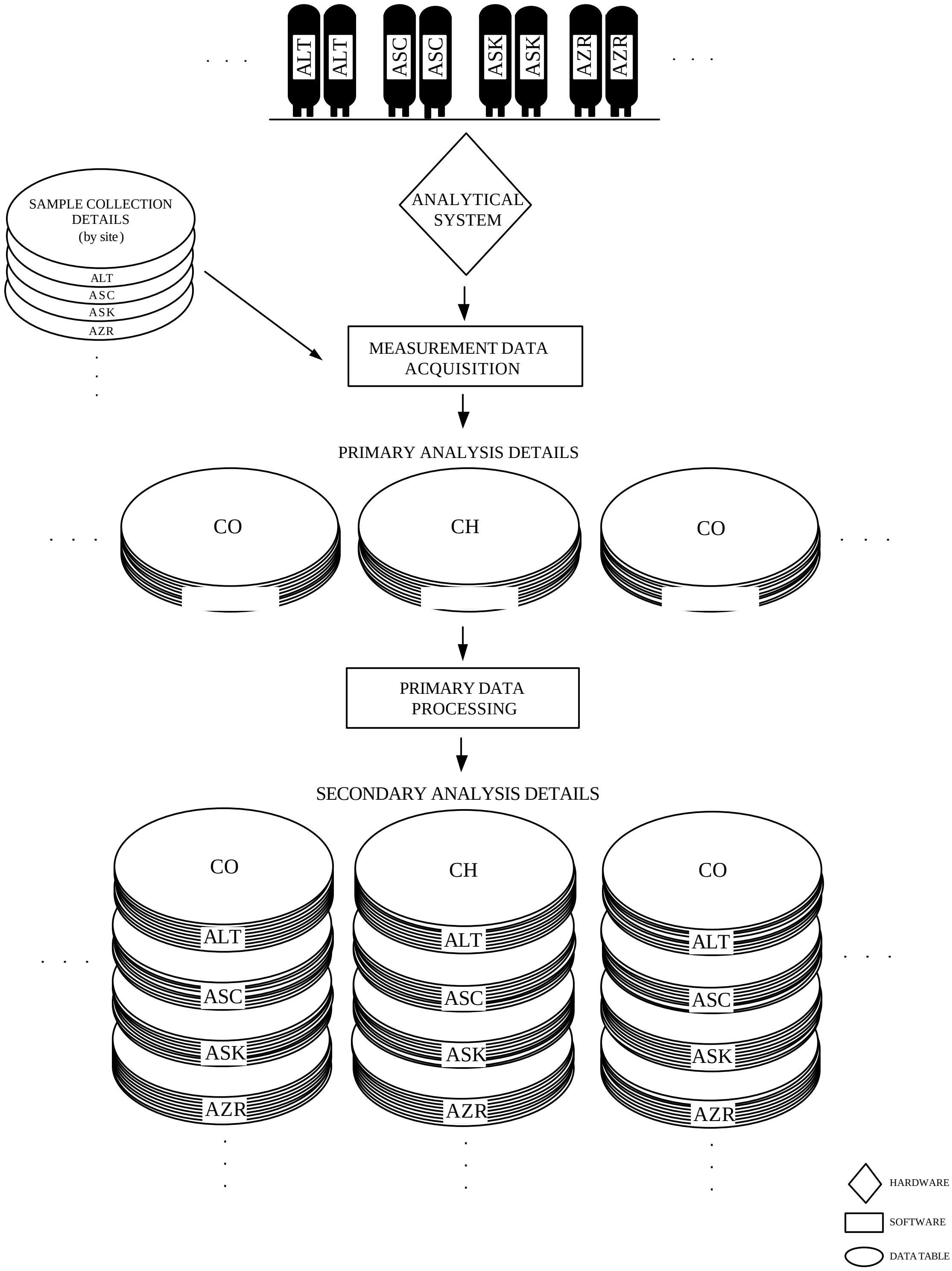

Figure 6: Schematic describing the flow of information from a typical CO_2, CH_4, CO, H_2, N_2O, and SF_6 analysis session. Tables containing H_2, N_2O, and SF_6 details are not shown. See text for a detailed explanation.

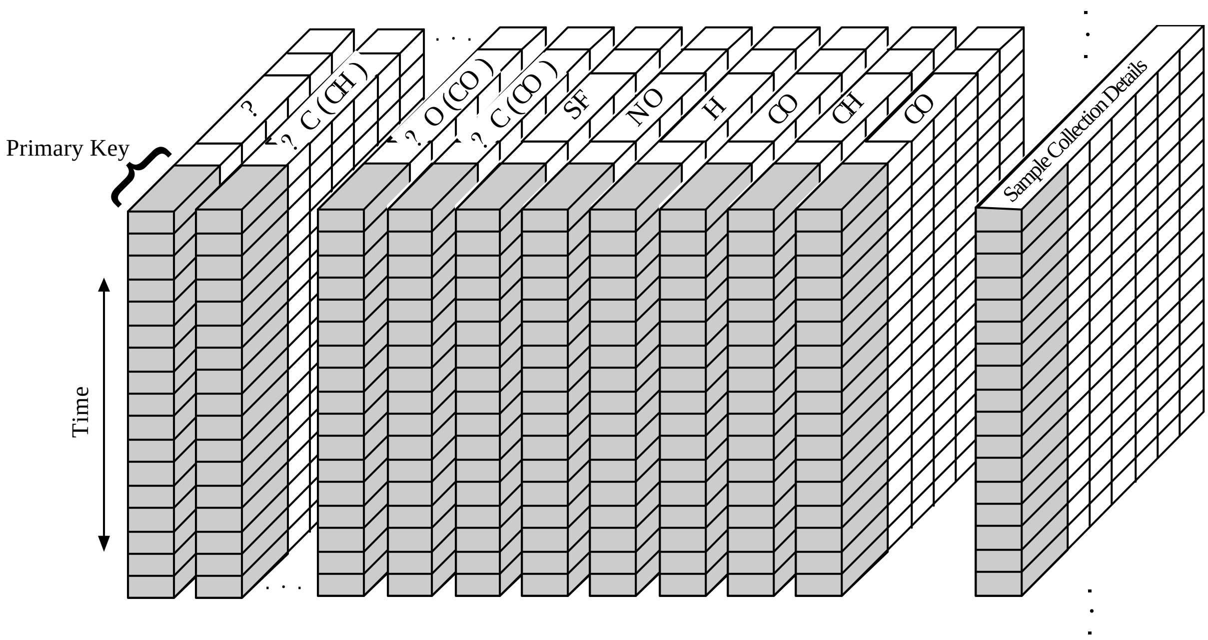

Figure 7: Schematic showing the grouping of tables containing secondary analysis details from measurements of samples collected at a network. The primary key is depicted as the shaded portion in each row.

CCGG has made a first attempt at managing meta data by ensuring that the RDBMS, itself, is self-documenting. This, however, represents only a fraction of the important details pertaining to the maintenance and continuing development of the cooperative air sampling measurement program that should be included in the RDBMS. Work in progress focuses on how meta data will be input into the RDBMS and what "key" attributes will be used to relate meta data tables with data tables.

5.4 User Requirements

Of primary concern in developing and implementing the CCGG RDBMS is to ensure that the final product meets all users' needs and will likely meet their needs for quite some time. The following discussion describes how the design of the CCGG RDBMS attempts to satisfy present and future user requirements as discussed in section 3.1. The RDBMS excels in some areas and falls short of expectations in others. Experience, user feedback, and database performance will likely highlight the RDBMS weaknesses and dictate the direction of future improvements.

5.4.1 Accessibility

The CCGG RDBMS resides on a central PC file server running the Red Hat Linux operating system. The primary task of this computer is to provide file access and storage to the approximately 20 user accounts. The RDBMS software is a nominal burden to this server, requiring very little computer memory and processing time. The current storage requirement for data derived from the cooperative air sampling network is a modest 650 Mb (6.5 ? 10^8 bytes). All tables reside on this server so that data can be readily accessed by any user. Because this approach ensures that group members have immediate access to data without delay, individual measurement projects are motivated to examine their measurement data routinely and when necessary, make timely calibration adjustments, data corrections and selections. Off-site access is available by establishing direct or indirect secured shell sessions. Subsets of the data tables are available to the wider community from NOAA CMDL (http://www.cmdl.noaa.gov/ccgg/index.html), CDIAC (http://cdiac.esd.ornl.gov), and the WMO WDCGG (http://gas.kishou.go.jp/wdcgg.html).

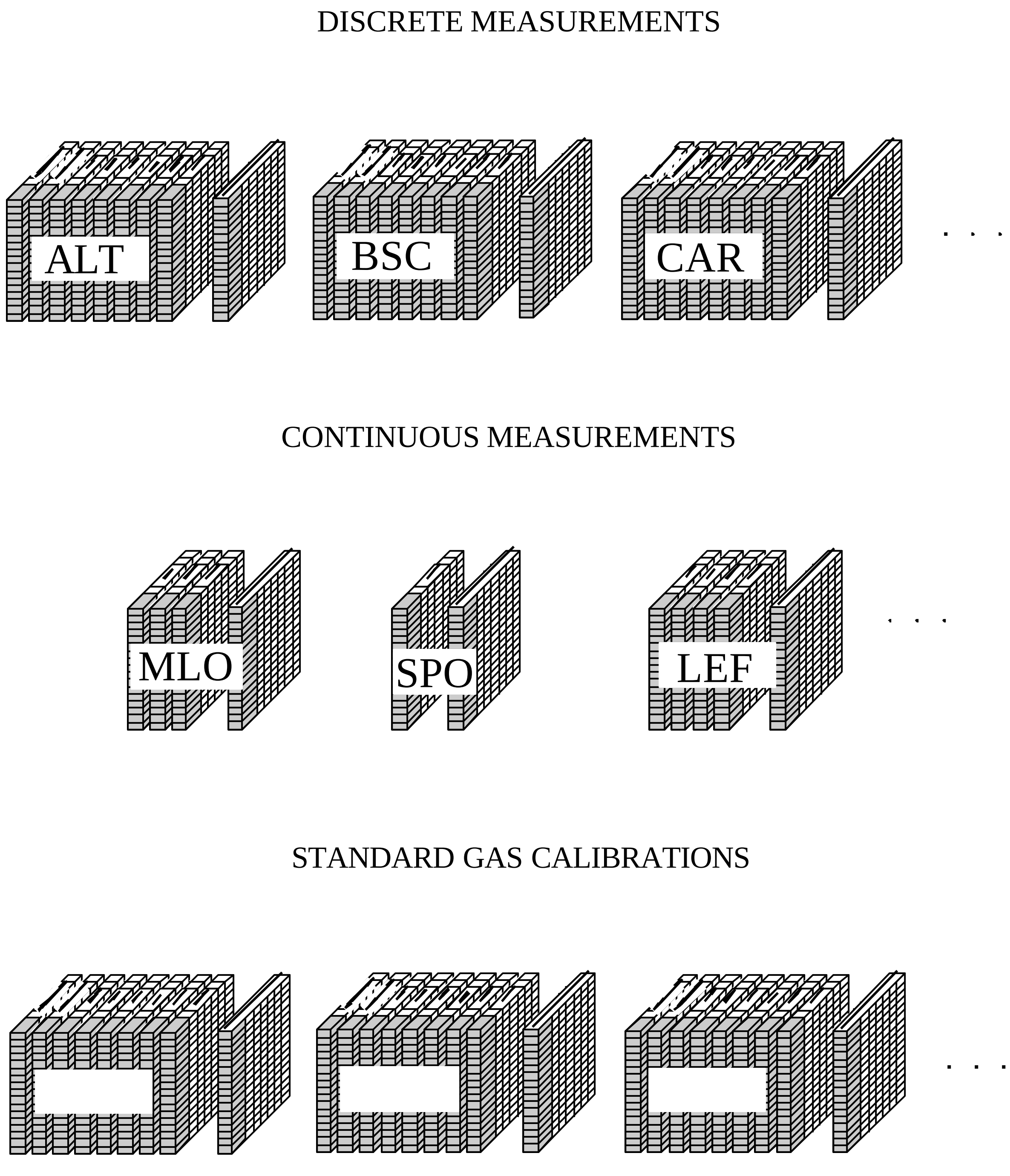

Figure 8: Data tables containing secondary or higher-order analysis details are grouped by each CCGG measurement program. Primary keys are depicted as the shaded portion in each row.

Commercial and custom software packages, several programming languages, and Linux operating system commands are centrally available and access the data tables directly. Users can make copies of complete or partial data tables for their own use, but in practice, copies are seldom made because access is direct, easy and preferable.

5.4.2 Portability

Data tables exist as text files which makes them easily portable to most computing environments and software applications. The Tables 2-5 (top) accurately describe formats used to manage data derived from the CMDL cooperative air sampling network and aircraft measurement programs.

5.4.3 Readability

Tables are designed to employ universally-accepted formatting protocol (Section 3.3.3) to the extent possible (Tables 2-5, top). Missing data are assigned unambiguous default values to ensure that all data items have an entry. Tables are explicitly formatted to maximize readability from a broad range of computing environments. It is even possible to list tables directly to a printing device without the need for further formatting. In some instances, additional details are provided at the beginning of tables to improve understanding of the tables' contents. Tables are designed to be easily read by both humans and software applications. The CCGG RDBMS attempts to use consistent formatting for both continuous and discrete measurement programs. This is particularly helpful when developing software applications that access tables directly. Once developed, these computer programs can easily read tables associated with either discrete or continuous measurements from fixed or moving sampling locations. New measurement programs within or associated with CCGG are strongly encouraged to conform to the RDBMS formatting protocol. This effort ensures straightforward integration of new data streams into the RDBMS and enables new measurement programs to make use of existing data analysis and visualization software applications.

5.4.4 Data Security

Security of the CCGG RDBMS is maintained by restricting access to tables using standard Linux file access permission commands that limit access by "owner" and "group." The ability to append, modify, or create tables is restricted to a few "select" computer programs. Since writing to a table is performed by software, any event that modifies a table is electronically logged providing a dynamic account of database operations.

Data security methods applied to the CCGG RDBMS have been adequate, but could be improved. When the RDBMS was first implemented, problems would arise that occasionally required manual repair to data tables by project members. As a result, several members of CCGG have had full (read and write) access to the data tables. This practice naturally compromised the security of the RDBMS. With time, the number of these occurrences has diminished to the point where only the database administrator(s) are now required to have full access to the RDBMS.

Automatic backups of the entire RDBMS are made monthly to one of several LTO (Linear Tape Open) digital data storage cartridges. A different tape cartridge is employed each month so that the monthly "snapshot" of the RDBMS is preserved for several months before the cartridge is re-used and the saved data overwritten. Incremental backups are made to hard disk daily.

5.4.5 Data Integrity

Computer programs run weekly and automatically to verify consistency among tables of the RDBMS. Because the analysis details for all trace constituents reside in many different tables and are related to each other and to the sample collection details by the sample collection key, it is imperative that the contents of these tables are not compromised. For example, a weekly consistency check will verify that every record in a table of primary analysis details has a "related" record in the appropriate tables of secondary analysis details. Similarly, the check also verifies that every record in a table of secondary analysis details has a “related” record in the appropriate table of primary analysis details. All records in all tables for all species and all sites must be traceable. Summaries from these consistency checks are saved to a log file as well as electronically mailed to the database administrators.

5.4.6 Quality Assurance

CCGG employs a flagging scheme to convey information about the quality of each discrete measurement. Flags can be assigned automatically during analysis or by computer programs that employ various selection schemes. A single computer program exists that enables appropriate users to manually assign flags to measurement values.

The flag is a 3-character alphanumeric attribute that accompanies each measurement value, i.e., in all tables containing secondary or higher-order analysis details (see Tables 4 and 5). To maximize readability in the formatted data tables, the RDBMS requires that all measurement results have a flag. The (automatic) default flag, “…”, implies that the measurement result has no qualifications, i.e., the measured quantity is representative of large well-mixed air typical for the sampling location. Information is conveyed by the type of flag assigned and its position within the 3-character attribute. For example, an alphanumeric other than a period “.”, in the first (left-most) character position implies that the measurement result is questionable due to an obvious sample collection or analytical problem, (e.g., N.., C..) and should not be used when interpreting the time series. An alphanumeric other than a period in the second (middle) position implies that the measurement is valid but is not representative of typical background conditions. For example, the (automatic) flag, “.X.”, indicates that the measurement has been identified as a non-background value determined using a computer program that contains a statistical selection scheme; the (manual) flag, “.D.”, also means that the measurement result is not representative, but in this case, the determination was made from the sample collector’s written remarks noting that the sample was collected during persistent winds from a known contaminated wind sector. An alphanumeric other than a period in the third (right-most) position qualifies the measurement further. For example, the (automatic) flag, “..S”, indicates that the representative measurement is from a single flask (CCGG generally samples in pairs). The (automatic) flag, “*..”, indicates that a sample was analyzed, but for some reason (e.g., insufficient sample or analytical system problem) a mixing ratio or isotope ratio could not be determined. This “rejection” flag always accompanies a default value (-999.99[9]) in the concentration field.

While principal investigators, responsible for assessing the quality of the measurements, must conform to the general flagging scheme described above, they can define flag types and combinations of flag types as required to adequately encode information pertinent to the measurement result. This flexibility can lead to a large number of flag types and combination of flag types that may be specific to each measurement program. For example, while the INSTAAR Stable Isotope Laboratory adheres to the basic flagging scheme, they have many more and different flag types than the flag types used by CCGG. This does not present a problem for those interested in interpreting the measurement records who only need to be concerned about the position in which alphanumeric characters other than the period, “.”, appear in the flag field.

5.4.7 Automation

Like most measurement programs, CCGG employs varying degrees of automation in order to maximize the efficiency of its operation and maintain the high quality of its measurements. For example, once the collection details from a network air sample are entered into the RDBMS, all subsequent data tables are updated automatically. Specifically, once a discrete air sample is analyzed, the appropriate primary, secondary, and third-order analysis tables are updated with no user intervention. Further, when standard gas calibration details change such that primary analysis data must be re-processed, existing computer software interacts with the RDBMS automatically, re-computing mixing ratios and updating the appropriate data tables. Occasionally, post-processing corrections are required to data because of known analytical or sampling problems. Correction functions are documented in tables (Table 7) and automatically applied to the appropriate discrete samples whenever primary analysis data are re-processed. Automating this procedure ensures that these corrections are applied as necessary without the need of human intervention.

CCGG maintains several analytical systems operating in Boulder and at the CMDL observatories. Each of these systems is "networked" to the central server providing access to the RDBMS. This access ensures that results from standard gas calibrations for all trace gas species are readily available to individual measurement programs, providing the "best" standard gas values for use in computing relative amounts of trace gas compounds.

5.4.8 Flexibility

Issues of flexibility and expandability were considered throughout the development of the CCGG data management strategy and RDBMS implementation. The data classification scheme, table organization, and use of formatted text files are components in the design that allow the RDBMS to grow and improve with time. Data from new and developing projects are easily incorporated into the RDBMS provided that minimal guidelines are satisfied. Since its introduction in 1992, the RDBMS has had to adapt to the growing and changing needs of CCGG. Several sampling locations have been added to and removed from the network. Measurements of N_2O, and SF_6 in network samples began in 1995 and recently, measurements of ?^{13}C in CH_4 have been added to the suite of measured trace gas species. And finally, data from the aircraft measurement program which began in 1992 have been integrated into the RDBMS. These changes have required only minor modifications to the CCGG RDBMS.

5.4.9 Simplicity

RDBMS architects may argue that the CMDL CCGG RDBMS is too simple. It is indeed simple and in certain aspects primitive. These characteristics, however, are its strengths. RDBMS architects, administrators, operators, and users alike can easily grasp the flow of information, the grouping of data and tables, and the mechanics of the RDBMS that provide the necessary services. Because the overall organization is simple, users can quickly develop an understanding of the data, its quality, and how it might be used in interpretive studies. "Simplicity" is certainly one reason for its success and longevity.

6. CONCLUSION

The Carbon Cycle Greenhouse Gases Group RDBMS has been in operation since 1992. While the overall strategy and in-house RDBMS package continue to satisfy the growing needs of the CMDL program, weaknesses in the strategy have become apparent during the 10 years of operation. These problem areas are being addressed and, as a result, the CCGG RDBMS continues to evolve. Our experience with commercially-available RDBMS and spreadsheet applications strengthens our commitment to the continued development of an in-house RDBMS operating within the Linux environment. Laboratories that already employ a RDBMS package that meets current data management needs and can easily adapt to future changes will likely continue using a proven package. Developing programs as well as existing laboratories that do not have an acceptable data management strategy should appreciate that managing atmospheric data is equally important to making the measurements. Initial costs to develop and implement a data management strategy are high but necessary. Adopting the RDBMS described here, either in part or as a whole, can reduce these initial costs and can provide a solution that has proven to successfully manage the many different data types of a large measurement program.

```
#  Corrections to Carbon Cycle Group CO2 data
#
#  Correction information fields
#
#  site          site (* = all sites)
#  start         start date for correction
#  stop          stop date for correction
#  by            date (sdate = sample date; adate = analysis date)
#  id            flask identification (*-* = all flasks)
#  m             collection method (* = all methods)
#  in            instruments code (** = all instruments)
#  type          type of correction (must match a defined type, e.g., adate, sdate, value)
#  fnct          correction function
#  tzero         date at to = 0
#  np            number of parameters
#  parameters    parameter list
#
#  All discrete samples with SAMPLE date equal to or after the start date and SAMPLE date equal to or before the end date will be corrected.
#
```

	start			stop													
site	yr	mo	dy	yr	mo	dy	by	id	m	in	type	fnct	tzero	np	parameters		

```
#
#  Corrections to discrete samples analyzed on Komhyr Semi-automatic flask apparatus
#
```

*	1979	01	01	1983	07	31	adate	*-*	*	**	adate	poly	1900.00	1	0.240000		
*	1985	07	12	1987	09	21	adate	*-*	*	**	adate	poly	1900.00	3	699.948840	-16.302490	0.094940
*	1987	09	22	1988	12	07	adate	*-*	*	**	adate	poly	1900.00	1	0.140000		
PSA	1979	01	01	1979	12	31	sdate	*-*	*	**	sdate	poly	1900.00	1	-3.200000		
AMS	1979	01	01	1979	12	31	sdate	*-*	*	**	sdate	poly	1900.00	1	-3.200000		

```
#
#  Corrections to discrete samples analyzed on Siemens flask apparatus
#
```

*	1988	12	01	1993	11	17	adate	*-*	*	S1	adate	poly	1900.00	1	0.100000		
*	1993	11	18	1999	12	31	adate	*-*	*	S1	adate	poly	1900.00	1	0.240000		

Table 7: Table documenting CMDL CCGG post-processing corrections to discrete CO_2 measurements.

Regardless of the RDBMS selected, a working knowledge of database management concepts and a well-conceived data management strategy are required. Users' needs, the flow of information, and the classification of data should be identified early in the development process. The strategy should be flexible to meet the growing and changing needs of the measurement program. Data produced by ongoing measurement programs steadily accumulate and a framework should exist to ensure that data are easily secured and always readily accessible for viewing, re-processing, interpreting, and disseminating.

This discussion is a first attempt at establishing carbon cycle database management system guidelines. The NOAA CMDL CCGG RDBMS has been described as one approach that has successfully met many of the atmospheric measurement data management design criteria common to most programs. It is our hope that this document can serve as a starting point from which a more comprehensive set of guidelines can be established with input from the carbon cycle measurement community.

ACKNOWLEDGMENTS

We thank P. Lang, T. Conway (NOAA CMDL), and Ray Langenfelds (CSIRO DAR) for their helpful discussions while preparing this manuscript.

References

Astraham M.M., et al., System R. Relational Approach to Database Management, *ACM Transactions on Database Systems 1, 2*. June 1976.

Ciais, P., P.P. Tans, M. Trolier, J.W.C. White, and R.J. Francey. A Large Northern Hemisphere Terrestrial CO_2 Sink Indicated by the $^{13}C/^{12}C$ Ratio of Atmospheric CO_2, *Science*, *269*, 1098-1102, 1995.

GLOBALVIEW-CH_4. Cooperative Atmospheric Data Integration Project - Methane. CD-ROM, NOAA/CMDL, Boulder, Colorado. [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: CCGG/ch4/GLOBALVIEW], 2001.

GLOBALVIEW-CO_2. Cooperative Atmospheric Data Integration Project - Carbon Dioxide. CD-ROM, NOAA/CMDL, Boulder, Colorado. [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: CCGG/co2/GLOBALVIEW], 2002.

Masarie K.A., L.P. Steele, and P.M. Lang. A Rule-Based Expert System for evaluating the Quality of Long-Term In Situ Gas Chromatographic Measurements of Atmospheric Methane, *NOAA Technical Memorandum ERL CMDL-3*, 37 pp. NOAA Environmental Research Laboratories, Boulder, Colorado, 1991.

Rumble Jr., J.R. and F.J. Smith. *Database Systems in Science and Engineering*. Published under the Adam Hilger imprint of IOP Publishing Ltd., 1990.

World Meteorological Organization (WMO), Guide on Sampling and Analysis Techniques for Chemical Constituents and Physical Properties in Air and Precipitation as Applied at Stations of the Global Atmosphere Watch, Rep. 134, 40 pp., Global Atmos. Watch, Geneva, 1998.

GLOBAL ATMOSPHERE WATCH REPORT SERIES

1. Final Report of the Expert Meeting on the Operation of Integrated Monitoring Programmes, Geneva, 2-5 September 1980

2. Report of the Third Session of the GESAMP Working Group on the Interchange of Pollutants Between the Atmosphere and the Oceans (INTERPOLL-III), Miami, USA, 27-31 October 1980

3. Report of the Expert Meeting on the Assessment of the Meteorological Aspects of the First Phase of EMEP, Shinfield Park, U.K., 30 March - 2 April 1981

4. Summary Report on the Status of the WMO Background Air Pollution Monitoring Network as at April 1981

5. Report of the WMO/UNEP/ICSU Meeting on Instruments, Standardization and Measurements Techniques for Atmospheric CO_2, Geneva, 8-11; September 1981

6. Report of the Meeting of Experts on BAPMoN Station Operation, Geneva, 23-26 November, 1981

7. Fourth Analysis on Reference Precipitation Samples by the Participating World Meteorological Organization Laboratories by Robert L. Lampe and John C. Puzak, December 1981*

8. Review of the Chemical Composition of Precipitation as Measured by the WMO BAPMoN by Prof. Dr. Hans-Walter Georgii, February 1982

9. An Assessment of BAPMoN Data Currently Available on the Concentration of CO_2 in the Atmosphere by M.R. Manning, February 1982

10. Report of the Meeting of Experts on Meteorological Aspects of Long-range Transport of Pollutants, Toronto, Canada, 30 November - 4 December 1981

11. Summary Report on the Status of the WMO Background Air Pollution Monitoring Network as at May 1982

12. Report on the Mount Kenya Baseline Station Feasibility Study edited by Dr. Russell C. Schnell

13. Report of the Executive Committee Panel of Experts on Environmental Pollution, Fourth Session, Geneva, 27 September - 1 October 1982

14. Effects of Sulphur Compounds and Other Pollutants on Visibility by Dr. R.F. Pueschel, April 1983

15. Provisional Daily Atmospheric Carbon Dioxide Concentrations as Measured at BAPMoN Sites for the Year 1981, May 1983

16. Report of the Expert Meeting on Quality Assurance in BAPMoN, Research Triangle Park, North Carolina, USA, 17-21 January 1983

17. General Consideration and Examples of Data Evaluation and Quality Assurance Procedures Applicable to BAPMoN Precipitation Chemistry Observations by Dr. Charles Hakkarinen, July 1983

18. Summary Report on the Status of the WMO Background Air Pollution Monitoring Network as at May 1983

19. Forecasting of Air Pollution with Emphasis on Research in the USSR by M.E. Berlyand, August 1983

20. Extended Abstracts of Papers to be Presented at the WMO Technical Conference on Observation and Measurement of Atmospheric Contaminants (TECOMAC), Vienna, 17-21 October 1983

21. Fifth Analysis on Reference Precipitation Samples by the Participating World Meteorological Organization Laboratories by Robert L. Lampe and William J. Mitchell, November 1983

22. Report of the Fifth Session of the WMO Executive Council Panel of Experts on Environmental Pollution, Garmisch-Partenkirchen, Federal Republic of Germany, 30 April - 4 May 1984 (TD No. 10)

23. Provisional Daily Atmospheric Carbon Dioxide Concentrations as Measured at BAPMoN Sites for the Year 1982. November 1984 (TD No. 12)

24. Final Report of the Expert Meeting on the Assessment of the Meteorological Aspects of the Second Phase of EMEP, Friedrichshafen, Federal Republic of Germany, 7-10 December 1983. October 1984 (TD No. 11)

25. Summary Report on the Status of the WMO Background Air Pollution Monitoring Network as at May 1984. November 1984 (TD No. 13)

26. Sulphur and Nitrogen in Precipitation: An Attempt to Use BAPMoN and Other Data to Show Regional and Global Distribution by Dr. C.C. Wallén. April 1986 (TD No. 103)

27. Report on a Study of the Transport of Sahelian Particulate Matter Using Sunphotometer Observations by Dr. Guillaume A. d'Almeida. July 1985 (TD No. 45)

28. Report of the Meeting of Experts on the Eastern Atlantic and Mediterranean Transport Experiment ("EAMTEX"), Madrid and Salamanca, Spain, 6-8 November 1984

29. Recommendations on Sunphotometer Measurements in BAPMoN Based on the Experience of a Dust Transport Study in Africa by Dr. Guillaume A. d'Almeida. September 1985 (TD No. 67)

30. Report of the Ad-hoc Consultation on Quality Assurance Procedures for Inclusion in the BAPMoN Manual, Geneva, 29-31 May 1985

31. Implications of Visibility Reduction by Man-Made Aerosols (Annex to No. 14) by R.M. Hoff and L.A. Barrie. October 1985 (TD No. 59)

32. Manual for BAPMoN Station Operators by E. Meszaros and D.M. Whelpdale. October 1985 (TD No. 66)

33. Man and the Composition of the Atmosphere: BAPMoN - An international programme of national needs, responsibility and benefits by R.F. Pueschel. 1986

34. Practical Guide for Estimating Atmospheric Pollution Potential by Dr. L.E. Niemeyer. August 1986 (TD No. 134)

35. Provisional Daily Atmospheric CO_2 Concentrations as Measured at BAPMoN Sites for the Year 1983. December 1985 (TD No. 77)

36. Global Atmospheric Background Monitoring for Selected Environmental Parameters. BAPMoN Data for 1984. Volume I: Atmospheric Aerosol Optical Depth. October 1985 (TD No. 96)

37. Air-Sea Interchange of Pollutants by R.A. Duce. September 1986 (TD No. 126)

38. Summary Report on the Status of the WMO Background Air Pollution Monitoring Network as at 31 December 1985. September 1986 (TD No. 136)

39. Report of the Third WMO Expert Meeting on Atmospheric Carbon Dioxide Measurement Techniques, Lake Arrowhead, California, USA, 4-8 November 1985. October 1986

40. Report of the Fourth Session of the CAS Working Group on Atmospheric Chemistry and Air Pollution, Helsinki, Finland, 18-22 November 1985. January 1987

41. Global Atmospheric Background Monitoring for Selected Environmental Parameters. BAPMoN Data for 1982, Volume II: Precipitation chemistry, continuous atmospheric carbon dioxide and suspended particulate matter. June 1986 (TD No. 116)

42. Scripps reference gas calibration system for carbon dioxide-in-air standards: revision of 1985 by C.D. Keeling, P.R. Guenther and D.J. Moss. September 1986 (TD No. 125)

43. Recent progress in sunphotometry (determination of the aerosol optical depth). November 1986

44. Report of the Sixth Session of the WMO Executive Council Panel of Experts on Environmental Pollution, Geneva, 5-9 May 1986. March 1987

45. Proceedings of the International Symposium on Integrated Global Monitoring of the State of the Biosphere (Volumes I-IV), Tashkent, USSR, 14-19 October 1985. December 1986 (TD No. 151)

46. Provisional Daily Atmospheric Carbon Dioxide Concentrations as Measured at BAPMoN Sites for the Year 1984. December 1986 (TD No. 158)

47. Procedures and Methods for Integrated Global Background Monitoring of Environmental Pollution by F.Ya. Rovinsky, USSR and G.B. Wiersma, USA. August 1987 (TD No. 178)

48. Meeting on the Assessment of the Meteorological Aspects of the Third Phase of EMEP IIASA, Laxenburg, Austria, 30 March - 2 April 1987. February 1988

49. Proceedings of the WMO Conference on Air Pollution Modelling and its Application (Volumes I-III), Leningrad, USSR, 19-24 May 1986. November 1987 (TD No. 187)

50. Provisional Daily Atmospheric Carbon Dioxide Concentrations as Measured at BAPMoN Sites for the Year 1985. December 1987 (TD No. 198)

51. Report of the NBS/WMO Expert Meeting on Atmospheric CO_2 Measurement Techniques, Gaithersburg, USA, 15-17 June 1987. December 1987

52. Global Atmospheric Background Monitoring for Selected Environmental Parameters. BAPMoN Data for 1985. Volume I: Atmospheric Aerosol Optical Depth. September 1987

53. WMO Meeting of Experts on Strategy for the Monitoring of Suspended Particulate Matter in BAPMoN - Reports and papers presented at the meeting, Xiamen, China, 13-17 October 1986. October 1988

54. Global Atmospheric Background Monitoring for Selected Environmental Parameters. BAPMoN Data for 1983, Volume II: Precipitation chemistry, continuous atmospheric carbon dioxide and suspended particulate matter (TD No. 283)

55. Summary Report on the Status of the WMO Background Air Pollution Monitoring Network as at 31 December 1987 (TD No. 284)

56. Report of the First Session of the Executive Council Panel of Experts/CAS Working Group on Environmental Pollution and Atmospheric Chemistry, Hilo, Hawaii, 27-31 March 1988. June 1988

57. Global Atmospheric Background Monitoring for Selected Environmental Parameters. BAPMoN Data for 1986, Volume I: Atmospheric Aerosol Optical Depth. July 1988

58. Provisional Daily Atmospheric Carbon Dioxide Concentrations as measured at BAPMoN sites for the years 1986 and 1987 (TD No. 306)

59. Extended Abstracts of Papers Presented at the Third International Conference on Analysis and Evaluation of Atmospheric CO_2 Data - Present and Past, Hinterzarten, Federal Republic of Germany, 16-20 October 1989 (TD No. 340)

60. Global Atmospheric Background Monitoring for Selected Environmental Parameters. BAPMoN Data for 1984 and 1985, Volume II: Precipitation chemistry, continuous atmospheric carbon dioxide and suspended particulate matter.

61. Global Atmospheric Background Monitoring for Selected Environmental Parameters. BAPMoN Data for 1987 and 1988, Volume I: Atmospheric Aerosol Optical Depth.

62. Provisional Daily Atmospheric Carbon Dioxide Concentrations as measured at BAPMoN sites for the year 1988 (TD No. 355)

63. Report of the Informal Session of the Executive Council Panel of Experts/CAS Working Group on Environmental Pollution and Atmospheric Chemistry, Sofia, Bulgaria, 26 and 28 October 1989

64. Report of the consultation to consider desirable locations and observational practices for BAPMoN stations of global importance, Bermuda Research Station, 27-30 November 1989

65. Report of the Meeting on the Assessment of the Meteorological Aspects of the Fourth Phase of EMEP, Sofia, Bulgaria, 27 and 31 October 1989

66. Summary Report on the Status of the WMO Global Atmosphere Watch Stations as at 31 December 1990 (TD No. 419)

67. Report of the Meeting of Experts on Modelling of Continental, Hemispheric and Global Range Transport, Transformation and Exchange Processes, Geneva, 5-7 November 1990

68. Global Atmospheric Background Monitoring for Selected Environmental Parameters. BAPMoN Data For 1989, Volume I: Atmospheric Aerosol Optical Depth

69. Provisional Daily Atmospheric Carbon Dioxide Concentrations as measured at Global Atmosphere Watch (GAW)-BAPMoN sites for the year 1989 (TD No. 400)

70. Report of the Second Session of EC Panel of Experts/CAS Working Group on Environmental Pollution and Atmospheric Chemistry, Santiago, Chile, 9-15 January 1991 (TD No. 633)

71. Report of the Consultation of Experts to Consider Desirable Observational Practices and Distribution of GAW Regional Stations, Halkidiki, Greece, 9-13 April 1991 (TD No. 433)

72. Integrated Background Monitoring of Environmental Pollution in Mid-Latitude Eurasia by Yu.A. Izrael and F.Ya. Rovinsky, USSR (TD No. 434)

73. Report of the Experts Meeting on Global Aerosol Data System (GADS), Hampton, Virginia, 11 to 12 September 1990 (TD No. 438)

74. Report of the Experts Meeting on Aerosol Physics and Chemistry, Hampton, Virginia, 30 to 31 May 1991 (TD No. 439)

75. Provisional Daily Atmospheric Carbon Dioxide Concentrations as measured at Global Atmosphere Watch (GAW)-BAPMoN sites for the year 1990 (TD No. 447)

76. The International Global Aerosol Programme (IGAP) Plan: Overview (TD No. 445)

77. Report of the WMO Meeting of Experts on Carbon Dioxide Concentration and Isotopic Measurement Techniques, Lake Arrowhead, California, 14-19 October 1990

78. Global Atmospheric Background Monitoring for Selected Environmental Parameters BAPMoN Data for 1990, Volume I: Atmospheric Aerosol Optical Depth (TD No. 446)

79. Report of the Meeting of Experts to Consider the Aerosol Component of GAW, Boulder, 16 to 19 December 1991 (TD No. 485)

80. Report of the WMO Meeting of Experts on the Quality Assurance Plan for the GAW, Garmisch-Partenkirchen, Germany, 26-30 March 1992 (TD No. 513)

81. Report of the Second Meeting of Experts to Assess the Response to and Atmospheric Effects of the Kuwait Oil Fires, Geneva, Switzerland, 25-29 May 1992 (TD No. 512)

82. Global Atmospheric Background Monitoring for Selected Environmental Parameters BAPMoN Data for 1991, Volume I: Atmospheric Aerosol Optical Depth (TD No. 518)

83. Report on the Global Precipitation Chemistry Programme of BAPMoN (TD No. 526)

84. Provisional Daily Atmospheric Carbon Dioxide Concentrations as measured at GAW-BAPMoN sites for the year 1991 (TD No. 543)

85. Chemical Analysis of Precipitation for GAW: Laboratory Analytical Methods and Sample Collection Standards by Dr Jaroslav Santroch (TD No. 550)

86. The Global Atmosphere Watch Guide, 1993 (TD No. 553)

87. Report of the Third Session of EC Panel/CAS Working Group on Environmental Pollution and Atmospheric Chemistry, Geneva, 8-11 March 1993 (TD No. 555)

88. Report of the Seventh WMO Meeting of Experts on Carbon Dioxide Concentration and Isotopic Measurement Techniques, Rome, Italy, 7 - 10 September 1993, (edited by Graeme I. Pearman and James T. Peterson) (TD No. 669)

89. 4th International Conference on CO_2 (Carqueiranne, France, 13-17 September 1993) (TD No. 61)

90. Global Atmospheric Background Monitoring for Selected Environmental Parameters GAW Data for 1992, Volume I: Atmospheric Aerosol Optical Depth (TD No. 562)

91. Extended Abstracts of Papers Presented at the WMO Region VI Conference on the Measurement and Modelling of Atmospheric Composition Changes Including Pollution Transport, Sofia, 4 to 8 October 1993 (TD No. 563)

92. Report of the Second WMO Meeting of Experts on the Quality Assurance/Science Activity Centres of the Global Atmosphere Watch, Garmisch-Partenkirchen, 7-11 December 1992 (TD No. 580)

93. Report of the Third WMO Meeting of Experts on the Quality Assurance/Science Activity Centres of the Global Atmosphere Watch, Garmisch-Partenkirchen, 5-9 July 1993 (TD No. 581)

94. Report on the Measurements of Atmospheric Turbidity in BAPMoN (TD No. 603)

95. Report of the WMO Meeting of Experts on UV-B Measurements, Data Quality and Standardization of UV Indices, Les Diablerets, Switzerland, 25-28 July 1994 (TD No. 625)

96. Global Atmospheric Background Monitoring for Selected Environmental Parameters WMO GAW Data for 1993, Volume I: Atmospheric Aerosol Optical Depth

97. Quality Assurance Project Plan (QAPjP) for Continuous Ground Based Ozone Measurements (TD No. 634)

98. Report of the WMO Meeting of Experts on Global Carbon Monoxide Measurements, Boulder, USA, 7-11 February 1994 (TD No. 645)

99. Status of the WMO Global Atmosphere Watch Programme as at 31 December 1993 (TD No. 636)

100. Report of the Workshop on UV-B for the Americas, Buenos Aires, Argentina, 22-26 August 1994

101. Report of the WMO Workshop on the Measurement of Atmospheric Optical Depth and Turbidity, Silver Spring, USA, 6-10 December 1993, (edited by Bruce Hicks) (TD No. 659)

102. Report of the Workshop on Precipitation Chemistry Laboratory Techniques, Hradec Kralove, Czech Republic, 17-21 October 1994 (TD No. 658)

103. Report of the Meeting of Experts on the WMO World Data Centres, Toronto, Canada, 17-18 February 1995, (prepared by Edward Hare) (TD No. 679)

104. Report of the Fourth WMO Meeting of Experts on the Quality Assurance/Science Activity Centres (QA/SACs) of the Global Atmosphere Watch, jointly held with the First Meeting of the Coordinating Committees of IGAC-GLONET and IGAC-ACE, Garmisch-Partenkirchen, Germany, 13 to 17 March 1995 (TD No. 689)

105. Report of the Fourth Session of the EC Panel of Experts/CAS Working Group on Environmental Pollution and Atmospheric Chemistry (Garmisch, Germany, 6-11 March 1995) (TD No. 718)

106. Report of the Global Acid Deposition Assessment (edited by D.M. Whelpdale and M-S. Kaiser) (TD No. 777)

107. Extended Abstracts of Papers Presented at the WMO-IGAC Conference on the Measurement and Assessment of Atmospheric Composition Change (Beijing, China, 9-14 October 1995) (TD No. 710)

108. Report of the Tenth WMO International Comparison of Dobson Spectrophotometers (Arosa, Switzerland, 24 July - 4 August 1995)

109. Report of an Expert Consultation on 85Kr and 222Rn: Measurements, Effects and Applications (Freiburg, Germany, 28-31 March 1995) (TD No. 733)

110. Report of the WMO-NOAA Expert Meeting on GAW Data Acquisition and Archiving (Asheville, NC, USA, 4-8 November 1995) (TD No. 755)

111. Report of the WMO-BMBF Workshop on VOC Establishment of a "World Calibration/Instrument Intercomparison Facility for VOC" to Serve the WMO Global Atmosphere Watch (GAW) Programme (Garmisch-Partenkirchen, Germany, 17-21 December 1995) (TD No. 756)

112. Report of the WMO/STUK Intercomparison of Erythemally-Weighted Solar UV Radiometers, Spring/Summer 1995, Helsinki, Finland (TD No. 781)

113. The Strategic Plan of the Global Atmosphere Watch (GAW) (TD No. 802)

114. Report of the Fifth WMO Meeting of Experts on the Quality Assurance/Science Activity Centres (QA/SACs) of the Global Atmosphere Watch, jointly held with the Second Meeting of the Coordinating Committees of IGAC-GLONET and IGAC-ACE^{Ed}, Garmisch-Partenkirchen, Germany, 15-19 July 1996 (TD No. 787)

115. Report of the Meeting of Experts on Atmospheric Urban Pollution and the Role of NMSs (Geneva, 7-11 October 1996) (TD No. 801)

116. Expert Meeting on Chemistry of Aerosols, Clouds and Atmospheric Precipitation in the Former USSR (Sankt Peterburg, Russian Federation, 13-15 November 1995)

117. Report and Proceedings of the Workshop on the Assessment of EMEP Activities Concerning Heavy Metals and Persistent Organic Pollutants and their Further Development (Moscow, Russian Federation, 24-26 September 1996) (Volumes I and II) (TD No. 806)

118. Report of the International Workshops on Ozone Observation in Asia and the Pacific Region (IWOAP, IWOAP-II), (IWOAP, 27 February-26 March 1996 and IWOAP-II, 20 August-18 September 1996) (TD No. 827)

119. Report on BoM/NOAA/WMO International Comparison of the Dobson Spectrophotometers (Perth Airport, Perth, Australia, 3-14 February 1997), (prepared by Robert Evans and James Easson) (TD No. 828)

120. WMO-UMAP Workshop on Broad-Band UV Radiometers (Garmisch-Partenkirchen, Gemany, 22 to 23 April 1996) (TD No. 894)

121. Report of the Eighth WMO Meeting of Experts on Carbon Dioxide Concentration and Isotopic Measurement Techniques (prepared by Thomas Conway) (Boulder, CO, 6-11 July 1995) (TD No. 821)

122 Report of Passive Samplers for Atmospheric Chemistry Measurements and their Role in GAW (prepared by Greg Carmichael) (TD No. 829)

123 Report of WMO Meeting of Experts on GAW Regional Network in RA VI, Budapest, Hungary, 5 to 9 May 1997

124 Fifth Session of the EC Panel of Experts/CAS Working Group on Environmental Pollution and Atmospheric Chemistry, (Geneva, Switzerland, 7-10 April 1997) (TD No. 898)

125. Instruments to Measure Solar Ultraviolet Radiation, Part 1: Spectral Instruments (lead author G. Seckmeyer) (TD No. 1066)

126. Guidelines for Site Quality Control of UV Monitoring (lead author A.R. Webb) (TD No. 884)

127. Report of the WMO-WHO Meeting of Experts on Standardization of UV Indices and their Dissemination to the Public (Les Diablerets, Switzerland, 21-25 July 1997) (TD No. 921)

128. The Fourth Biennial WMO Consultation on Brewer Ozone and UV Spectrophotometer Operation, Calibration and Data Reporting, (Rome, Italy, 22-25 September 1996) (TD No. 918)

129. Guidelines for Atmospheric Trace Gas Data Management (Ken Masarie and Pieter Tans), 1998 (TD No. 907)

130. Jülich Ozone Sonde Intercomparison Experiment (JOSIE, 5 February to 8 March 1996), (H.G.J. Smit and D. Kley) (TD No. 926)

131. WMO Workshop on Regional Transboundary Smoke and Haze in Southeast Asia (Singapore, 2 to 5 June 1998) (Gregory R. Carmichael). Two volumes

132. Report of the Ninth WMO Meeting of Experts on Carbon Dioxide Concentration and Related Tracer Measurement Techniques (Edited by Roger Francey), (Aspendale, Vic., Australia)

133. Workshop on Advanced Statistical Methods and their Application to Air Quality Data Sets (Helsinki, 14-18 September 1998) (TD No.956)

134. Guide on Sampling and Analysis Techniques for Chemical Constituents and Physical Properties in Air and Precipitation as Applied at Stations of the Global Atmosphere Watch. Carbon Dioxide

135. Sixth Session of the EC Panel of Experts/CAS Working Group on Environmental Pollution and Atmospheric Chemistry (Zurich, Switzerland, 8-11 March 1999) (WMO TD No.1002)

136. WMO/EMEP/UNEP Workshop on Modelling of Atmospheric Transport and Deposition of Persistent Organic Pollutants and Heavy Metals (Geneva, Switzerland, 16-19 November 1999) (Volumes I and II) (TD No. 1008)

137. Report and Proceedings of the WMO RA II/RA V GAW Workshop on Urban Environment (Beijing, China, 1-4 November 1999) (WMO-TD. 1014) (Prepared by Greg Carmichael)

138. Reports on WMO International Comparisons of Dobson Spectrophotometers, Parts I – Arosa, Switzerland, 19-31 July 1999, Part II – Buenos Aires, Argentina (29 Nov. – 12 Dec. 1999 and Part III – Pretoria, South Africa (18 March – 10 April 2000)

139. The Fifth Biennial WMO Consultation on Brewer Ozone and UV Spectrophotometer Operation, Calibration and Data Reporting (Halkidiki, Greece, September 1998)(WMO TD No. 1019)

140. WMO/CEOS Report on a Strategy for Integrating Satellite and Ground-based Observations of Ozone (WMO TD No. 1046)

141. Report of the LAP/COST/WMO Intercomparison of Erythemal Radiometers (Thessaloniki, Greece, 13-23 September 1999) (WMO TD No. 1051)

142. Strategy for the Implementation of the Global Atmosphere Watch Programme (2001-2007), A Contribution to the Implementation of the Long-Term Plan (WMO TD No.1077)

143. Global Atmosphere Watch Measurements Guide (WMO TD No. 1073)

144. Report of the Seventh Session of the EC Panel of Experts/CAS Working Group on Environmental Pollution and Atmospheric Chemistry and the GAW 2001 Workshop (Geneva, Switzerland, 2 to 5 April 2001)

145. WMO GAW International Comparisons of Dobson Spectrophotometers at the Meteorological Observatory Hohenpeissenberg, Germany (21 May – 10 June 2000, MOHp2000-1), 23 July – 5 August 2000, MOHp2000-2), (10 – 23 June 2001, MOHp2001-1) and (8 to 21 July 2001, MOHp2001-2). Prepared by Ulf Köhler (WMO TD No. 1114)

146. Quality Assurance in monitoring solar ultraviolet radiation: the state of the art

147. Workshop on GAW in RA VI (Europe), Riga, Latvia, 27-30 May 2002

148. Report of the Eleventh WMO/IAEA Meeting of Experts on Carbon Dioxide Concentration and Related Tracer Measurement Techniques (Tokyo, Japan, 25-28 September 2001) (WMO TD No 1138)

149. Comparison of Total Ozone Measurements of Dobson and Brewer Spectrophotometers and Recommended Transfer Functions (prepared by J. Staehelin, J. Kerr, R. Evans and K. Vanicek)(WMO TD No. 1147)