

Data Quality Management and Monitoring Networks

by

Jim Augustyn
Augustyn + Company

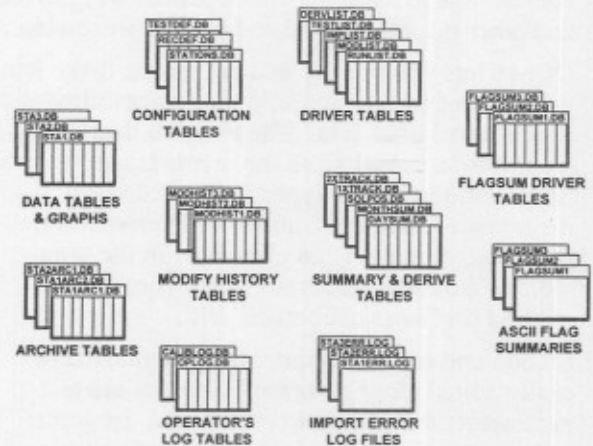
4th NUMUG Workshop
San Francisco, California
April 24-26, 1996

DQMS™ - The Data Quality Management System™

DQMS manages time sequence data in monitoring networks, providing detailed data quality tracking, and the flexibility needed for evolving networks. Its design and context-sensitive help system make it easy to learn, and its depth of features make it far superior to ad-hoc programming for even the simplest monitoring applications.

Key Features:

- ▶ Handles any number of stations and record types
- ▶ Easy to add or change station, record or test definitions
- ▶ Makes customizable data plots
- ▶ Provides multiple data quality tests
- ▶ Provides multiple data repair or modification methods
- ▶ Data modifications are tracked and are fully recoverable
- ▶ Assigns a two part data quality flag to each data value
- ▶ Provides detailed summaries of flagged data
- ▶ Ties instrument calibrations to data values
- ▶ Automatic or manual control of DOS and DQMS processes
- ▶ Creates statistical summaries
- ▶ Exports to several formats
- ▶ Provides access to an Operator's Event Log
- ▶ Context sensitive on-line help
- ▶ Built on a relational database foundation



DQMS Elements

DQMS is written within Paradox for Windows, a flexible and powerful relational database environment. DQMS is available in a "Runtime" version, so it is not necessary to also buy a copy of Paradox for Windows. However, Paradox does provide a wide array of features you can use to compliment DQMS. You have access to the database files DQMS creates through other database and spreadsheet programs. DQMS can export to popular formats including ACSII, dBase, and Microsoft Excel, and through Paradox, links can be made to SQL, Foxpro, Access, and other database formats.

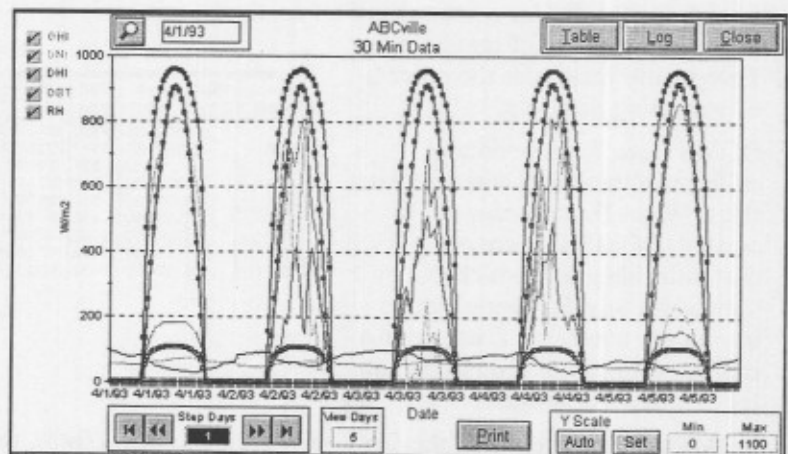
DQMS can integrate data collection and processing through timed, automatic execution of DOS and DQMS processes.

It can handle many different time, date, and data logger record ID protocols, can archive and retrieve data sets to manage disk space, and can run most operations in batch mode.

A key feature of DQMS is a two-part flag assigned to each data value indicating test results, and the nature of any modifications. These flags identify bad or modified values, and control how data files and summaries are composed and delivered.

DateTime	GHI	GHI Flag	DNI	DNI Flag	DHI	DHI Flag
4/5/93 08:00	270.4	91	307.76	C100	110.1	90
4/5/93 08:30	423.3	91	526.74	C200	148.8	90
4/5/93 09:00	522.2	91	669.99	C200	145.8	90
4/5/93 09:30	607.2	91	689.3	C200	149.2	90

Part of a DQMS Data Table



A DQMS Graph

DQMS provides several tests, including missing record and range checks, field-to-field and record-to-record comparisons, and the SERI/QC test developed at the U.S. DOE's National Renewable Energy Laboratory (NREL) for assessing solar radiation data quality. It can also automatically insert blank records when missing records are found. It summarizes test results, and identifies possibly erroneous data in a form that facilitates remedial action. DQMS pro-

DQMS™

vides daily, monthly and yearly statistical summaries in printed or spreadsheet compatible format, and can perform several special solar irradiance calculations.

DQMS can track instrument calibrations, and ties specific instruments to data logger channels and to collected data values. It also has a network operator's log to record events and other information related to data processing.

DQMS lets you modify bad or missing data. It archives data values prior to changes, allowing traceability of changes and access to original data. The two part data quality flag identifies test results and the methods used to modify data.

Sixteen modification types are provided including replacement, addition, interpolation, factoring, copying from elsewhere in the same table or from other sources, and a correction method for linear calibration drift.

Station and record definitions are contained in easily edited *Configuration Tables*, as are test parameters for each field and record. Program functions are controlled via menu operations and with *Driver Tables*, which are lists of operations to be performed for specific stations over specific time periods.

Driver Tables allow precise control of program processes, and provide detailed records of changes to individual data values and records. The simplest *Driver Table* is a list of *Data Tables* and associated start and stop times to control testing, archiving, or exporting. *Modify Driver Tables* also include the modification type, the field to be modified, and space to report the date and time the modification was actually made. DQMS can automatically make partly completed *Modify Driver Tables* comprised of all contiguous time sequences with out of range fields. This greatly facilitates the process of correcting data sets.

DQMS *Data Tables* contain collected data values and associated flags. When *Data Tables* are created, DQMS also creates a customizable graph which can be viewed for any day using simple navigation controls. These graphs can be customized through Paradox for Windows or printed.

DQMS is an outgrowth of the *Data Verification System (DVS)*, developed in 1988 by Augustyn + Company Consulting Engineers in Berkeley, California which has been involved in environmental monitoring since 1978. DVS was enhanced under contract to NREL for use in several domestic and international R&D programs.

DQMS Version 2.0 requires Microsoft Windows 3.1 or higher, and can be purchased for \$295 for a copy limited to one station definition, or for \$895 for a copy allowing unlimited stations. Contact Jim Augustyn for a demo version or further information at 510-525-0464 (voice), 510-525-9410 (fax), aci@ccnet.com (E-mail), or by mail to:

Augustyn + Company, 1029 Solano Avenue, Berkeley, California 94706 510-525-0464

Station Code	Station Number	Network Name	Station Name	Latitude (+N)	Longitude (+E)	Elevation (M)	Time Zone
ABC			ABCville	38.00	-121.00	100	-8
BAL		PacRim	Bali, Indonesia	-7.50	115.02	107	8
CP	911	SIMP	Carissa Plains	35.35	-120.05	610	-8
KMB	1	DOD_L1	Kerman Incinerator	29.33	-142.33	-27.3	-9

DQMS Station Configuration Table

Station Code	Datalogger Record ID	Logger Field #	Datalogger Field Name	Record Code	Inst Code	Data Table Field Name	Data Type	Transformation	Factor
KMT	108			1				TimeStep Minute	1
KMT	108	1	Record ID	1				Record ID	
KMT	108	2	Julian Day	1		RecDate		JulianDay	
KMT	108	3	HHMM	1		Time		HourMinute	
KMT	108	4	WS	1		WS			
KMT	108	5	WD	1		WD			
KMT	201			D				TimeStep Hour	24
KMT	201	1	Record ID	D				Record ID	
KMT	201	2	Julian Date	D		RecDate		JulianDay	
KMT	201	3	AVG12VBV	D		Avg 12V Battery			
KMT	201	4	AVG6VBV	D		Avg 6V Battery			
KMT	201	5	MIN12VBV	D		Min 12V Battery			
KMT	201	6	MIN6VBV	D		Min 6V Battery			
KP	101			1				TimeStep Minute	1
KP	101	1	Record ID	1				Record ID	
KP	101	2	Station Number	1				Station Number	

DQMS Record Configuration Table

Station Code	Record Code	Start Date	Start Time	End Date	End Time
HPP	A	1/1/84	5	1/31/85	2400
AX1		12/23/89	15	5/30/95	2345
CP	H	1/1/84	30	12/31/94	2400
CP	H	1/1/82	30	12/31/83	2400
CP	D	6/12/91	30	6/12/91	1200

DQMS Test Driver Table

Station Code	Rec Code	Start Date	Start Time	End Date	End Time	Field Name	Mod Type	Factor	Date & Time Done	Comment
KMF		10/26/93	1100	10/26/93	1300	WS	Replace	2	9/26/94 19:23	99 5
KMF		10/26/93	1330	10/27/93	800	WS	Multiply	.01	9/26/94 19:23	248.1 08 38
KMF		10/27/93	830	10/27/93	1000	DBT	Linear1		9/26/94 19:23	99 4
KMF		10/27/93	830	10/27/93	1000	DNI	Linear1		9/26/94 19:23	99 4
KMF		10/27/93	830	10/27/93	1000	GHI	Linear1		9/26/94 19:23	99 4
KMF		10/27/93	830	10/27/93	1000	WS	Replace	2	9/26/94 19:23	99 4
KMF		10/27/93	1030	10/31/93	2400	WS	Multiply	.01	9/26/94 19:23	83.2 08 220
KMF		12/2/93	700	12/2/93	700	GHI	Multiply	.5	9/26/94 19:38	12.0 08 1
KMF		12/8/93	730	12/8/93	730	GHI	Multiply	.5	9/26/94 19:38	43.0 08 1

DQMS Modification Driver Table