INLAND WATERS BRANCH

DEPARTMENT OF ENERGY, MINES AND RESOURCES

A Storage and Retrieval System for Water Quality Data

A. DEMAYO

REPORT SERIES NO. 9

REPORT SERIES NO. 9

A Storage and Retrieval System for Water Quality Data

A. DEMAYO

INLAND WATERS BRANCH
DEPARTMENT OF ENERGY, MINES AND RESOURCES
OTTAWA, CANADA, 1970

Contents

Illustrations

Tables

Acknowledgements

The author would like to thank Mr. Jeff Moppett, Miss B. Gervais, and Mrs. D. Freinkin of the Computer Sciences Division, for their active and very valuable participation in conceiving this system and writing up the computer programs. The author also wishes to thank Mr. A. Treichel (Computer Sciences Division) who participated in the early stages of this project.

The continuous collaboration and support of Messrs. Thomas, Gale, Lively, Traversy and Peters of the Water Quality Division is very much appreciated.

A Storage and Retrieval System for Water Quality Data

(NAQUADAT System)*

INTRODUCTION

This report describes the Data Processing System being implemented in the Water Quality Division, Inland Waters Branch, for storing and retrieving Water Quality data. Some changes may be made in the system before it becomes fully operational. However, these changes should not be many and should not significantly affect the principles and the important parts of the system.

DESCRIPTION OF SYSTEM

The system is divided into two sections: a storage section and a retrieval section.

STORAGE SECTION

The input data for NAQUADAT consist of:

(i) Sampling station data

(ii) Sample data

(iii) Analytical data

(iv) Dictionary information material

Each of these items is discussed below.

Sampling station data

For key punching purposes, sampling station data are recorded on the blue form, number IWB-WQD-1600-1, inserted in the pocket inside the back cover. This form is divided into the following fields:

1. STATION NUMBER composed of the following subfields:

   a) Type of water - a 2-digit numerical code shown in Table 1.

---

* NAQUADAT is a provisional name for the system. It has its origin in the following title: National wAter QUAlity DATa bank

TABLE 1

TYPE OF WATER

Type of water	Code	Subtype of water	Code
Surface	0	Stream	0
		Lake	1
		Estuary	2
		Ocean-sea	3
		Ponds	4
		Impounded reservoirs	5
		Not known	9
Ground	1	Well	0
		Springs	1
		Not known	9
Waste	2	Industrial	0
		Municipal	1
		Mining	2
		Not known	9
Precipitation	3	Rain	0
		Snow	1
		Ice	2
		Not known	9
Treated	4		

Ref. : 1. Water quality and treatment, Manual American Water Works Association, 1941, p. 1ff.

2. G.M. Fair, J.C. Geyer and J.C. Morris, Water supply and water disposal, John Wiley & Sons, Inc., New York, 1954, p. 28ff.

b) Province or territory - a-2 digit alphabetical code consisting of the first two letters or the abbreviated initials of the name of the province or territory; for example, ON for the province of Ontario, NB for the province of New Brunswick.

c) Major basin - a 2-digit numerical code identical to major basin coding in Water Survey of Canada hydrometric networks.[1]

d) Sub-basin - a 2-digit alphabetic code identical to sub-basin coding in Water Survey of Canada hydrometric networks.[1]

e) Sequential number - a 4-digit numerical code representing the station number. Within each sub-basin, the stations will be numbered sequentially as they are established. Thus, the first station established in a sub-basin receives the number 0001, the second number 0002 and continuing to 9999.

An extra 3-digit subfield is provided in the station number for new data in the future. Initially, this additional subfield will contain zeros only.

It must be emphasized that the STATION NUMBER is the key to the system. The file of water quality data is organized sequentially by station number. Within each station, the data are organized in chronological order. To retrieve data from the system, one needs only to specify the station number and specific time period. Later on, as the system develops, computer programs will be written to retrieve data for stations within specified areas, delineated, for example, by their geographical coordinates. However, in the initial stages of the system, water quality data may be retrieved for all stations in a basin by specifying in the retrieval request a range of stations within a specific time period. For example, by requesting water quality data for stations 01BC08MF0001 to 01BC08MF9999 inclusive, from January 1, 1967, to December 31, 1969, all water quality data for lakes in the sub-drainage basin 08MF[1] in the Province of British Columbia for the period January 1, 1967, to December 31, 1969, will be retrieved. Similarly by specifying the range of stations from 01BC08AA0001 to 01BC08ZZ9999, all water quality data from lakes in the major drainage basin 08[1] in British Columbia will be retrieved (note that the range of dates must be also specified).

Many other combinations of retrieval are possible which add to the flexibility of the system.

2. LATITUDE, LONGITUDE and UTM UNIT fields are self-explanatory. Only one type of geographical coordinate may be specified, that is, LATITUDE & LONGITUDE or UTM UNITS. In NAQUADAT, all geographical coordinates are stored as Latitudes and Longitudes. If the initial information was in UTM units, the system will automatically convert it to Latitude and Longitude before storing it on magnetic tape.

A "P" (Precision) subfield of 1 digit, numerical, is also provided. It indicates the precision with which the location of a station is known. The corresponding code is given in Table 2.

[1] Water Survey of Canada, Reference Index Publications, Inland Waters Branch, Department of Energy, Mines and Resources.

TABLE 2

PRECISION OF LOCATION

Radius from the specified location within the station can be found. (the radius is given in meters)			Code
0 < r ≤	1		0
1 < r ≤	3		1
3 < r ≤	10		2
10 < r ≤	30	or 0 < r ≤ 1 sec.	3
30 < r ≤	100		4
100 < r ≤	300		5
300 < r ≤	1,000		6
1,000 < r ≤	3,000	or 1 sec. < r ≤ 1 min.	7
3,000 < r ≤	10,000		8
r >	10,000		9

3. REFERENCE STATION field serves only one purpose at the present time: it is the origin from which the distances to other stations are measured. These distances are given as station parameters (see below). The reference station provides a more precise identification of a station location.

4. NARRATIVE DESCRIPTION is an alphanumeric field usually containing descriptive information about the location of the station. However, it may contain other information.

The input form, IWB-WQD-1600-1, provides only 5 lines of 38 characters each for the narrative description. However, the storage program will accept up to 99 lines of 38 characters each.

5. Other "STATION PARAMETERS" are contained in the lower section of this form. These parameters can be such items as height above sea level, use of water, distance from the nearest reference station, etc. The distance from the reference station, for example, will be represented by a parameter code number of 6 digits (5 numeric and 1 alphabetic), and by the actual distance in a convenient unit (shown in Table 6). A more detailed discussion of parameter code numbers is given under Dictionary Information Material of this report.

Sample Data

Sample information is recorded on the green form, IWB-WQD-1600-2. This form contains the following fields:

1. STATION NUMBER

2. SAMPLE NUMBER composed of the following subfields:

a) Laboratory identification - a 2-digit numerical field which identifies the laboratory where the sample was analysed.

b) Year of receiving the sample in the laboratory - a 2-digit numerical field consisting of the last 2 digits of the calendar year in which the sample was received.

c) Sequential number - a 5-digit numerical field. The samples will be numbered sequentially from 00001 to 99999, each year, as they are received in the laboratory. This number will be recorded in this field.

Thus, each sample will have a unique identification number composed of the three subfields described above.

3. DATE OF SAMPLING (From). This field is composed of the following subfields:

a) Day, 2 digit, numerical

b) Month, 2 digit, numerical

c) Year, 2 digit, numerical (the last two digits are entered only)

d) Hours, 2 digit, numerical

e) Minutes, 2 digit, numerical

f) Zone, 3 digit, alphabetical. This field indicates the time zones according to the code shown in Table 3.

If the hours d) and minutes e) are not given, the system automatically assigns 12 hours and 00 minutes for these two subfields.

In the actual computer file the date of sampling is stored as year plus minutes relative to January 1, GMT time. Tables which convert this number back to day and month, for various time zones of Canada, are available in the Water Quality Division. A page of these conversion tables is reproduced in Figure 1.

The first three subfields, a), b) and c), must be filled in all cases. If the day of the sampling is not known, then an arbitrary day is assumed (say 15) and the uncertainty in the sampling day is shown in the "P" precision field (Table 4).

4. DATE OF SAMPLING (To)

This field is to be filled only if the sample is a composite sample (i.e. made up from a series of samples collected from one date to another date). It contains the same subfields as "from" with the exception of "zone", which is assumed to be the same for both periods.

TABLE 3

LOCAL TIME ZONE

TIME ZONE	Time Difference from GMT (hrs.)	CODE	PROVINCE
Newfoundland Standard Time	-3.5	NST	Nfld.
Newfoundland Daylight Time	-2.5	NDT	Lab.
Atlantic Standard Time	-4.0	AST	N.S., N.B.
Atlantic Daylight Time	-3.0	ADT	P.E.I., Que., and N.W.T. east of 68th Meridian
Eastern Standard Time	-5.0	EST	(1)
Eastern Daylight Time	-4.0	EDT	
Central Standard Time	-6.0	CST	(2)
Central Daylight Time	-5.0	CDT	
Mountain Standard Time	-7.0	MST	(3)
Mountain Daylight Time	-6.0	MDT	
Pacific Standard Time	-8.0	PST	B.C., and
Pacific Daylight Time	-7.0	PDT	N.W.T. west of 120th Meridian
Yukon Standard Time	-9.0	YST	Yukon
Yukon Daylight Time	-8.0	YDT	Territory

Notes: (a) Alberta remains on Mountain Standard Time all year
(b) Areas other than Alberta operate on daylight time from about May 1st to about October 31st.

(1) Ontario east of 90th Meridian; Quebec west of the 68th Meridian, and NWT between 68th and 85th Meridian.

(2) Ontario west of 90th Meridian; Manitoba; Southeasterly part of Saskatchewan, and NWT between 85th and 102nd Meridian.

(3) Saskatchewan except in the Southeasterly part; Alberta; and NWT between 102nd and 120th Meridian.

Reference: Canadian Almanac and Directory, 1970 edition, The Copp Clark Publishing Company, p. 5.

MINUTES (GMT) FROM	TO	DATE (EST) DAY	MONTH
438061	439500	1	11
439501	440940	2	11
440941	442380	3	11
442381	443820	4	11
443821	445260	5	11
445261	446700	6	11
446701	448140	7	11
448141	449580	8	11
449581	451020	9	11
451021	452460	10	11
452461	453900	11	11
453901	455340	12	11
455341	456780	13	11
456781	458220	14	11
458221	459660	15	11
459661	461100	16	11
461101	462540	17	11
462541	463980	18	11
463981	465420	19	11
465421	466860	20	11
466861	468300	21	11
468301	469740	22	11
469741	471180	23	11
471181	472620	24	11
472621	474060	25	11
474061	475500	26	11
475501	476940	27	11
476941	478380	28	11
478381	479820	29	11
479821	481260	30	11

MINUTES (GMT) FROM	TO	DATE (EST) DAY	MONTH
481261	482700	1	12
482701	484140	2	12
484141	485580	3	12
485581	487020	4	12
487021	488460	5	12
488461	489900	6	12
489901	491340	7	12
491341	492780	8	12
492781	494220	9	12
494221	495660	10	12
495661	497100	11	12
497101	498540	12	12
498541	499980	13	12
499981	501420	14	12
501421	502860	15	12
502861	504300	16	12
504301	505740	17	12
505741	507180	18	12
507181	508620	19	12
508621	510060	20	12
510061	511500	21	12
511501	512940	22	12
512941	514380	23	12
514381	515820	24	12
515821	517260	25	12
517261	518700	26	12
518701	520140	27	12
520141	521580	28	12
521581	523020	29	12
523021	524460	30	12
524461	525900	31	12

Figure 1. Reproduction of a page of time conversion tables.

TABLE 4

PRECISION OF TIME OF THE MEASUREMENT

Precision (P) of the time of the measurement*	Code
Average observation	0
P ≤ 1 minute	1
1 minute < P ≤ 4 minutes	2
4 minutes < P ≤ 15 minutes	3
15 minutes < P ≤ 60 minutes	4
1 hour < P ≤ 12 hours	5
12 hours < P ≤ 5 days	6
5 days < P ≤ 1 month	7
1 month < P ≤ 6 months	8
6 months < P	9

* The time of the measurement (t) is: $t = t_{obs} \pm P$

5. P FREQ:

a) P (Precision) - a 1-digit numerical field which shows the precision with which the time of sampling is known. The precision is indicated according to the code given in Table 4. If the precision is not indicated, i.e., the field is left blank, the computer program automatically inserts code "5" (Table 4) in this field.

b) FREQ (Frequency) - a 2-digit numerical field which represents the frequency with which the samples are collected. The frequency is indicated by a code number shown in Table 5.

6. SUBMITTER'S IDENTIFICATION NUMBER

This field is divided into two subfields:

a) The first subfield is a 4-digit numerical code assigned by the Water Quality Division to various projects or to various agencies submitting water samples to the Division for analyses.

b) The submitter's sample identification number is inserted in the second subfield and may contain up to 23 alphanumerical digits. The submitter's sample identification number, if indicated on the sample, will be recorded without modification.

This number will be used for special retrieval programs. For example, assume that the Department of Fisheries and Forestry, a submitting agency, has been assigned the submitter number 0008. By using this number and the

TABLE 5

FREQUENCY OF SAMPLING

Frequency of Sampling	Code
Intervals ≤ 1 minute	00
1 < intervals ≤ 2 minutes	01
2 < intervals ≤ 3 minutes	02
3 < intervals ≤ 4 minutes	03
4 < intervals ≤ 5 minutes	04
5 < intervals ≤ 6 minutes	05
6 < intervals ≤ 7 minutes	06
7 < intervals ≤ 8 minutes	07
8 < intervals ≤ 9 minutes	08
9 < intervals ≤ 10 minutes	09
10 < intervals ≤ 15 minutes	10
15 < intervals ≤ 20 minutes	11
20 < intervals ≤ 25 minutes	12
25 < intervals ≤ 30 minutes	13
30 minutes < intervals ≤ 1 hour	14
1 hour < intervals ≤ 3 hours	15
3 hours < intervals ≤ 6 hours	16
6 hours < intervals ≤ 12 hours	17
12 hours < intervals ≤ 24 hours	18
1 day < intervals ≤ 2 days	19
2 days < intervals ≤ 4 days	20
4 days < intervals ≤ 6 days	21
Once a week	22
7 days < intervals ≤ 10 days	23
10 days < intervals ≤ 13 days	24
Every two weeks	25
Every 15 days	26
15 days < intervals ≤ 20 days	27
20 days < intervals ≤ 29 days	28
Once a month	29
Every 2 months	30
Every 3 months	31
Every 4 months	32
Every 4 or 5 months	33
Every 6 months	34
Every 6 to 11 months	35
Every 12 months	36
Irregular	90
Not known	99

appropriate computer program, only data on Department of Fisheries and Forestry samples will be retrieved from the file. Furthermore, the print-outs will show the Fisheries and Forestry sample identification number which will enable the submitter to associate the results of the chemical analyses with other data that the submitter may have on the same sample.

TABLE 6 UNIT CODE NUMBERS

Type of Parameter	Unit	Abbreviation	Code
Analytical results	grams/litre	G/L	01
	milligrams/litre	MG/L	02
	micrograms/litre	UG/L	03
	nanograms/litre	NG/L	04
	picograms/litre	PG/L	05
	parts per million	PPM	06
	parts per billion	PPB	07
	equivalents per million	EPM	08
	moles/litre	MOL/L	09
	millimoles/litre	MMOL/L	10
	relative units	REL. UNITS	11
Specific conductance	mhos/cm	MHO/CM	12
	micro mhos/cm	UMHO/CM	13
Type of water, etc.	integers		14
General	per cent	%	15
Altitude, depth	miles	MI	16
	yards	YD	17
	feet	FT	18
	inches	IN	19
	kilometers	KM	20
	meters	M	21
	centimeters	CM	22
Flow	cubic feet/second	CFS	23
	cubic feet/minute	CFM	24
	cubic feet/hour	CFH	25
	imperial gallons/second	IGPS	26
	imperial gallons/minute	IGPM	27
	imperial gallons/day	IGPD	28
	gallons (U.S.)/second	GPS	29
	gallons (U.S.)/minute	GPM	30
	gallons (U.S.)/day	GPD	31
	cubic meter/second	MCPS	32
	liters/second	LPS	33
Temperature	degrees Centigrade	DEG.C	34
	degrees Fahrenheit	DEG.F	35
Atmospheric pressure	inches of mercury (at 0°C)	IN HG	36
	millimeters of mercury (at 0°C)	MM HG	37
Redox potential	millivolts	MV	38
Specific gravity	grams/cm^3	G/CM3	39
Surface tension	dynes/cm	DYN/CM	40

7. FLOW is usually the stream discharge measurement and is shown as "Flow: Instant" which is the actual discharge rate at time of sampling, and "Flow: Monthly Average", which is the average discharge during the month.

The "Flow: Instant" and the "Flow: Monthly Average" fields are each divided into two 6-digit subfields. The left subfield is the parameter code number and the right subfield is the actual discharge value. The parameter code consists of 5 numerical digits followed by 1 alphabetical digit. The units which may apply to the discharge values are shown in Table 6 and are also entered in the dictionary (see Dictionary of this report).

Additional fields have been provided on the input form to enter other information if desired, such as atmospheric conditions at the time of the sampling.

Analytical Data

The lower half of the form, IWB-WQD-1600-2, and the following forms 1600-3 to 1600-5, inclusive, shown in back of book, are designed for analytical results obtained *in situ*, in the field, in the laboratory, etc. The order in which the various tests appear on these forms is essentially for the convenience of the Divisional laboratories and has no significance in the data processing system. The parameters may be listed in any order and the forms may be modified without disruption to the data processing system.

The forms are designed for two purposes: (1) to serve as laboratory work-sheets and (2) to serve as documents from which data are key-punched on to computer cards. To accomplish this, numerical values to be key-punched are inserted in the white background windows, and other information such as descriptive names and calculations are shown in the coloured background windows. The several forms covering one sample may be submitted for key-punching together or separately. The station number and the date of sampling on each form place the data in its proper location in the computer file.

All five forms need not necessarily be used for each sample. If no parameters are measured on any one form, this form will not be submitted for key-punching. Parameters may be added indefinitely to the system and the forms are therefore modular in design, flexible and open-ended.

In the case of on-going surveys when a series of samples are collected from sampling stations, the form containing the station information is submitted only once for each station unless some of the information applied to that station changes. By definition, certain information belonging to a station cannot change without establishing a new station, e.g., location. However, information such as water use and narrative description may change without changing the station number.

Dictionary Information Material

The "Dictionary" is an integral and independent part of NAQUADAT. It is a separate file containing the following information:

1. Conversion factors between the various types of units used in this system. A page of the "Dictionary" containing some of the

conversion factors is reproduced in Figure 2.

2. Information about each analytical test:

   a) Parameter code number (first five digits only).

   b) Unit of measurement.

   c) Format for entering analytical results when the decimal point is implied. The same format is used in the retrieval report unless otherwise specified.

   d) Headings to be used in the retrieval report.

   e) Probable and possible limits for the analytical results.

   f) Conversion factor from mg/l to milliequivalents/l.

   g) Computation code which is an index number to a computation method. It applies to those parameters which are calculated from other parameters (as opposed to being directly measured).

   h) Narrative description of the test.

The form used for the dictionary input is shown inside back cover.

In the item "Limits" one can specify as many ranges as desired; discrete values instead of ranges may also be specified. When the data are stored, the values are compared with the limits indicated in the dictionary to check if they are reasonable.

In the case of calculated parameters, no limits are specified. It is assumed that if the values from which the respective parameters are calculated are correct, the calculated values are then correct. In the case of calculated parameters, instead of validation limits, the parameter code numbers of the values used in the calculation are specified.

Various restrictions which apply to the update of the "Dictionary" are given in Appendix 2. An example of a dictionary page is shown in Figure 3.

Parameter Code Number - The "Station Number" and the "Parameter Code Number" are the most important elements in NAQUADAT. The "Station Number" format is discussed in Sampling Station Data. The "Parameter Code Number" is a 6-digit field composed of the following subfields.

1. Element - a 2-digit numerical code. The numbers in this field correspond with the atomic number of the element in the periodic table.

2. State, group - a 1-digit numerical code which indicates the specific group of atoms in which the element is located (for example, in the case of nitrogen: nitrate, nitrite, etc.)

JAN. 30/70 W A T E R Q U A L I T Y - D I C T I O N A R Y PAGE 1

PARAM	RECORD TYPE		DATA
	CONV FACTOR	UNIT FROM 25 TO 27	0000.1038132
		25 28	00000149.491
		25 29	000.00207792
		25 30	00000.124675
		25 31	00000179.532
		25 33	000.00786557
		26 23	00000.160544
		26 24	000009.63264
		26 25	00000577.958
		26 27	0000000000060
		26 28	0000000086400
		26 29	000001.20094
		26 30	00000072.0564
		26 31	0000103761.2
		26 32	000.00454604
		26 33	000004.54592
		27 23	000.00267573
		27 24	00000.160544
		27 25	000009.63264
		27 26	0000.0166667
		27 28	0000000001440
		27 29	000.02001567
		27 30	000001.20094
		27 31	00000 1729.35
		27 33	0000.0757658
		28 25	000.00667336
		28 27	000.00069444
		28 30	00.000833986
		28 31	000001.20094
		29 23	0000.1336816
		29 24	00008.020896
		29 25	0000481.2538
		29 26	00000.832675
		29 27	0000049.9605
		29 28	0000071943.1
		29 30	0000000000060
		29 31	0000000086400
		29 32	0000.0037854
		29 33	00003.785306
		30 23	000.00222803
		30 24	0000.1336816
		30 25	00008.020896
		30 26	0000.0138779
		30 27	00000.832675
		30 28	0000001199.05
		30 29	0000.0166667
		30 31	0000000001440
		30 33	0000.0630884
		31 25	000.00557307
		31 27	00.000578247
		31 28	00000.832675

Figure 2. Dictionary Conversion Factors.

```
              JAN. 30/70              W A T E R   Q U A L I T Y  -  D I C T I O N A R Y              PAGE  8
PARAM  RECORD TYPE                                                        DATA

06101  RPT HEADINGS         FORMAT 1        UNIT 02            CARBON        DISS. ORG.                          C       MG/L
       VALIDATION  01 R  000500  000050  000500  005000
       DESCRIPTION                          01           ORGANIC MATERIAL IN SAMPLE IS OXIDIZED AND RESULTING CARBON
                                            02           DIOXIDE IS MEASURED BY INFRARED ANALYSIS.

06201  RPT HEADINGS         FORMAT 0        UNIT 02            BICARBONATE                                     HCO3      MG/L
       VALIDATION  01                                                                             016390   03
       DESCRIPTION                          01           CALCULATED FROM THE VALUES OF TOTAL ALKALINITY AND PHENOL
                                            02           ALKALINITY :
                                            03           IF PA=0 THEN HCO3=1.219*TA
                                            04           IF PA<=TA/2 THEN HCO3=1.219*(TA-2*PA)
                                            05           IF PA> TA/2 THEN HCO3=0

06301  RPT HEADINGS         FORMAT 5        UNIT 02            CARBONATE     DISSOLVED                         CO3       MG/L
       VALIDATION  01                                                                             033330   02
       DESCRIPTION                          01           CALCULATED FROM THE VALUES OF PHENOL ALKALINITY OR PHENOL
                                            02           ALKALINITY AND TOTAL ALKALINITY :
                                            03           IF PA=0 THEN CO3=0
                                            04           IF PA<=TA/2 THEN CO3=1.2*PA
                                            05           IF PA> TA/2 THEN CO3=1.2*(TA-PA)

06401  RPT HEADINGS         FORMAT 6        UNIT 02            CO2 FREE                                        CO2       MG/L
       VALIDATION  01 C 10101L  10301L                                                                     05
       DESCRIPTION                          01           CALCULATED VALUE :
                                            02           SET Y = 10**(-PH)   THEN :
                                            03           CO2 FREE = 9.7E10*Y*(TA/5E4+Y-1E-14/Y)/(1+11.22E-11/Y)
                                            04           REFERENCE : A.A.HIRSCH, INDUSTRIAL AND ENGINEERING CHEMISTRY
                                            05                       (ANALYTICAL EDITION), VOL.14, NO.12, P.944 (1942

06521  RPT HEADINGS         FORMAT 1        UNIT 02            OIL &         GREASE                                      MG/L

06531  RPT HEADINGS         FORMAT 4        UNIT 02            PHENOLIC      MATERIAL                                    MG/L

07001  RPT HEADINGS         FORMAT 1        UNIT 02            NITROGEN      TOT. KJELD.     07001L              N       MG/L
       VALIDATION  01 R  000000  000001  000050  005000
```

Figure 3. Dictionary Printout.

or, if it is dissolved, total, etc.

3. Analytical method - a 2-digit numerical sequential code. The numbers in this field identify the analytical method used for the test.

4. Place of analysis - a 1-digit alphabetical code indicating where the analysis was done, for example: situ (S), field (F), laboratory (L), or boat (B). Other letters can be used to indicate other analytical locations.

Two examples of the "Parameter Code Number" breakdown are illustrated as follows:

Parameter Code Number 07501L:

07 - nitrogen (periodic table),

5 - dissolved ammonia,

01 - direct nesslerization,

L - laboratory

The element is nitrogen as dissolved ammonia, analysed by the direct nesslerization method in a laboratory.

Parameter Code Number 20103L

20 - calcium,

1 - dissolved,

03 - atomic absorption,

L - laboratory

The parameter is calcium, dissolved, analysed by the atomic absorption method in a laboratory.

For the physical tests, colour, turbidity, etc., or for those chemical tests which cannot logically be associated with one particular element, the first two digits of the atomic numbers of the noble gases are used. It is not anticipated that these elements will be determined in water quality survey work and, therefore, their atomic numbers can be used without disrupting the data processing system described in this report. At the present time the coding adopted is as follows:

00xxx - calculated parameters

02xxx - general physical tests

10xxx - general chemical tests

Other elements in the periodic table which are not likely to be examined in water quality studies include elements numbered 96-99, and these numbers may be available for defining new groups of tests as they arise.

Analytical Results - Six digits are available for entering the analytical results when the decimal point is implied, five only when the decimal point is present. The presence of the decimal point overrides any format specification in the dictionary. When the decimal point is entered, the result can be written anywhere in the six-digit block provided for it on the laboratory forms. The results are stored as they are entered, i.e., if a value "1.2" is written, then it is stored as 1.2 and not as 1.20 or 1.200. This feature assures that no significant digits are lost or added to the file.

The system also provides for analytical results reported as "Greater than", "Less than" and "Trace". "Greater than" is indicated by a "G" followed by the respective number. For example, G0.01 means greater than 0.01. Correspondingly, L0.01 means less than 0.01. Trace is indicated by writing the word TRACE in the block provided for the results.

A minus sign is also permitted and if it is used, it should be immediately before the first digit or decimal point but after "G" or "L" when these are used. Examples:

-11.2; -.2(=-0.2); G-0.2; L-.5

As indicated previously, the units in which the results are entered are specified in the dictionary. Each parameter code number allows for only one type of unit. However, in the retrieval part, conversions from one type of unit to another may be made.

Diagnostics

A shortcoming of some data processing systems now used for storing water quality data is that they accept all data on the file without preliminary checking of its validity. The NAQUADAT system described in this report has an extensive checking procedure built-in, which will assure as much as possible, that only "good" data go on to the file.

The computer prints out "improbable value" and the "impossible value" messages when analytical data do not lie within the limits specified in the dictionary. The system will not accept "impossible values"; it will accept "improbable values" but warns that these should be checked. Other checks which are incorporated in the system are:

1. A check on the anions and cations balance. This check assures that no major ion was left out by misplacing a computer card and also checks the accuracy of the analysis. This check will not reject data; it will give a warning when the expression:

$$\frac{(\text{sum of anions}) - (\text{sum of cations})}{(\text{sum of anions}) + (\text{sum of cations})} \times 100 \qquad \text{(all in m-equiv/l)}$$

is greater than a pre-specified number.

When updating the file, it must be remembered that only new additions to the file are used for anions-cations balance calculations. Previous data placed on the file will not be retrieved for inclusion in these calculations. An extreme case occurs when the analytical results for only one ion is entered. In this case the anions-cations balance calculation will show a difference of 100 per cent. The "Update" program prints out the number of cations and anions used in the calculation which provides a simple check to ascertain if all the ions required in the balance calculation were indeed present.

In the computer subroutine which calculates the balance, only the first three digits of the parameter code are used to indicate a particular ion. Consequently, if two different values having the same three digits in the parameter code number are entered, both will be used in the balance calculation.

2. In the lower right-hand corner of the analytical result forms, IWB-WQD-1600-2 to IWB-WQD-1600-5, the total number of tests recorded on each form is inserted by the laboratory technician and is checked by the computer to assure that no test was omitted during the key punching operation.

3. The fields which should contain only numerical information are checked by the computer for the presence of alphabetical information; fields which should contain only alphabetical information are checked for the presence of numerical information.

4. The sampling dates are checked for validity.

A list of the diagnostic messages for "File Update" and their meaning is given in Appendix 3.

RETRIEVAL SECTION

The retrieval section of the system at the completion of the first phase will consist of printing-out water quality data by stations in tabular form. A retrieval request form is shown inside back cover. Each computer printout page can contain up to eight columns of data, and about 50 printed lines.

An example of a retrieval report is shown in Figure 4.

Some points of interest regarding the retrieval section are:

1. Requests for retrieval reports are made by specifying the range of station numbers and the dates. To retrieve data from one station only, the same station number is entered in the "FROM" and "TO" fields on the Retrieval Request form inside back cover. This capability of the system to accept a specified range of stations and dates opens ways to a larger variety of retrieval requests which can be satisfied.

2. In the print-out format, the order of columns is not fixed. The desired order must be specified in the Retrieval Request.

3. In the Retrieval Request, the "Parameter Code Numbers" of the

WATER QUALITY DATA REQUEST 0001 PAGE 1

STATION 05AC05EC001 LATITUDE 52.95000 LONGITUDE 109.88333

BATTLE RIVER NEAR UNITIN,SASKATCHEWAN

SASKATCHEWAN & QU@APPELLE BASINS
WATER QUALITY DATA FOR YEARS 1967-69

SAMPLE DATE			SAMPLE TIME CST		02061F TEMPERATURE	02041F SPECIFIC CONDUCTANCE	02061L TEMPERATURE	02041L SPECIFIC CONDUCTANCE	02071L TURBIDITY TOTAL		02011L COLCUP	10301L PH	10601L HARDNESS TOTAL	
D	M	Y	H	M	DEG.C.	UMHO/CM	DEG.C.	UMHO/CM	TURBIDITY UNITS		REL. UNITS	PH UNITS	CACO3 MG/L	
03	01	67	12	00	1		22	956			10.	8.3	03L	268.
01	02	67	12	00	1		20	1033			15.	8.6	03L	218.
01	03	67	12	00	1		21	1006			10.	8.2	03L	334.
01	04	67	12	00	1		19	988			5.	8.3	03L	238.
01	05	67	12	00	2		20	490			120.	7.8	03L	118.
01	06	67	12	00			24	544	73L	35.	45.	8.3	03L	175.
22	06	67	12	00	14		25	750			40.	8.3	03L	206.
04	07	67	12	00	21		25	646			40.	8.5	03L	213.
18	08	67	12	00	19		22	878			25.	8.4	03L	209.
01	09	67	12	00	22		26	1032			20.	8.5	03L	227.
02	10	67	12	00	8		23	1043			20.	8.7	03L	165.
02	11	67	12	00	3		22	977			10.	8.2	03L	239.
22	11	67	12	00	2		23	1110			5.	8.2	03L	243.
02	12	67	12	00	1		25	1550			10.	8.2	03L	295.
29	12	67	12	00	1		22	1380			20.	8.7	03L	312.
30	01	68	12	00	1		22	1310			20.	7.9	03L	458.
29	02	68	12	00	1		25	1540			30.	7.4	03L	489.
31	04	68	12	00	1		24	498			60.	7.7	03L	131.
07	05	68	12	00	9		25	796			30.	8.1	03L	228.
30	05	68	12	00	16		26	1031			30.	8.1	03L	242.
29	06	68	12	00	14		24	1082			20.	8.1	03L	235.
30	07	68	12	00	18		22	815			30.	8.3	03L	166.

Figure 4. Sample Computer Printout.

tests must be specified. The "Parameter Code Numbers" include, among other things, a two-digit code for identifying the method used in the analyses. In the Retrieval Request, the user has the option of specifying back-up methods, i.e., if the computer finds no data stored for a particular test under the first analytical method, it will go to a second analytical method, etc. Up to three back-up methods can be specified. If no data are found under the first method specified in the retrieval request, but data are found under a back-up method, the code number for this backup is printed out on the left side of the value representing the analytical result. The first parameter code number requested in the retrieval appears in the heading of a column.

4. The headings of the columns are specified in the "Dictionary".

CONFIDENTIAL DATA

If certain water quality data are of a confidential nature, these should be stored on a separate tape file.

COMPUTER PROGRAMS

The programming was done by the staff of the Computer Sciences Division of the Department of Energy, Mines and Resources. The computer programs were written in COBOL for an IBM 360 System. It is anticipated that only minor changes will be necessary if a different computer is used to run the system. No computer programs are given in this brochure. The following is a list of programs comprising the initial NAQUADAT system. The programs are being documented and will be made available to interested organizations. A flow chart of the system is given in Figure 5.

List of Programs - Phase 1 (Computer Science Division)

J61000	Card to tape.
J61010	Sort transactions.
J61020	Dictionary update.
J61030	Dictionary tape to disk.
J61040	Station file and data file update.
J61050	Request record generator.
J61060	Sort request records.
J61070	Retrieval.
J61080	Sort parameter calculation arguments.
J61090	Parameter calculations.
J61100	Sort retrieved values.
J61110	Provisional report.
J61120	Sort retrieval requests.
J61130	Auxiliary listings.
J61140	Reorganization of search records.
J61150	Merge calculated values, retrieved values, etc.
J61160	Dictionary dump from disk.
J61170	Dictionary print from tape.
J61180	Data file print.
J61190	--number not assigned--
J61200	Station match program.

J61210 Station change validation.
J61220 Sort valid station changes.
J61230 Station change program.
J61240 Sort renumbered station file.
J61250 Sort renumbered data file.
J61260 Resequence and check for duplicates on station and data files.

Phase 2 will be initiated after phase 1 is completely operational, and will consist of additional specialized retrieval programs for special reports and studies. Modifications which may be required to programs completed in phase 1 will be carried out in phase 2.

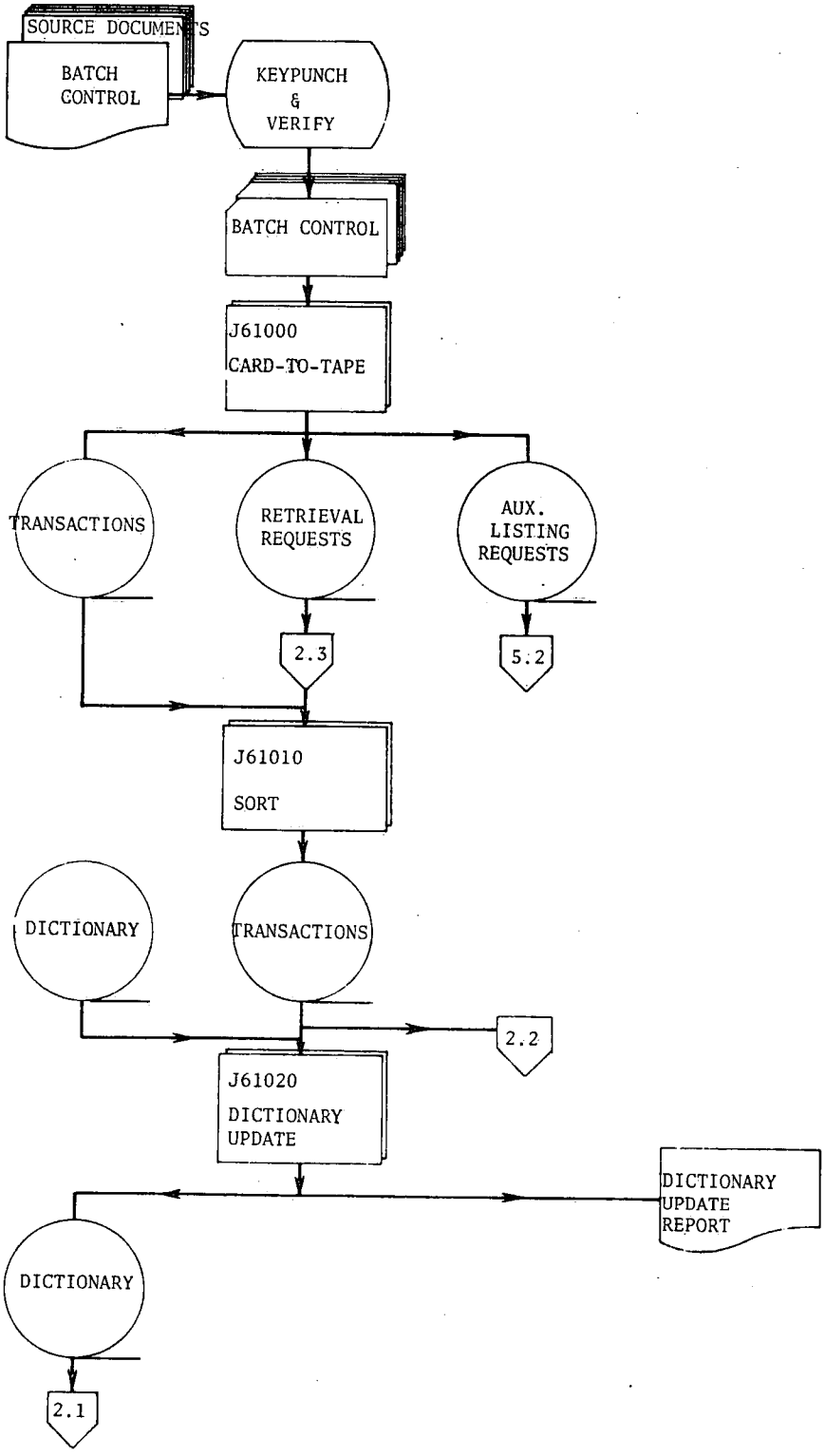

Figure 5. Computer Flow Diagram (Sheet 1).

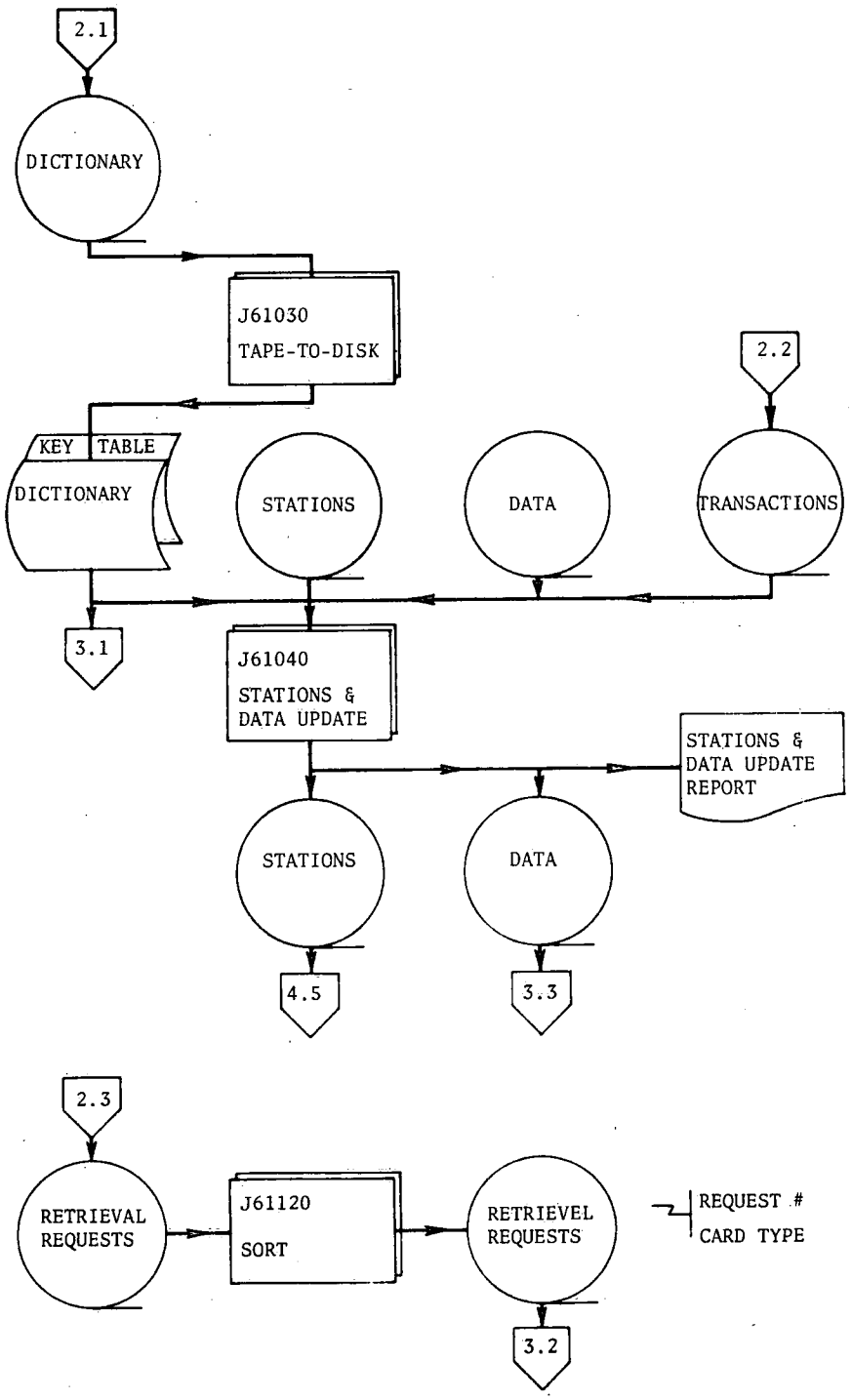

Figure 5. Computer Flow Diagram (Sheet 2).

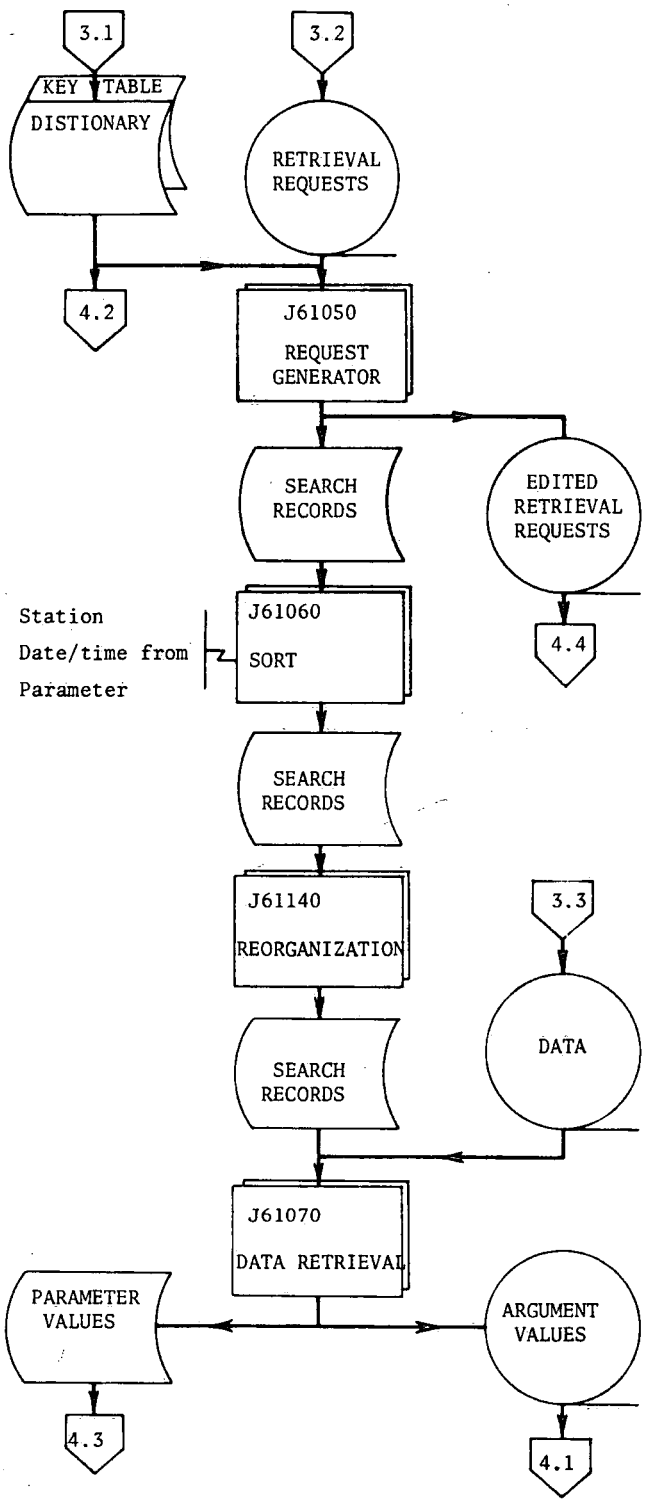

Figure 5. Computer Flow Diagram (Sheet 3).

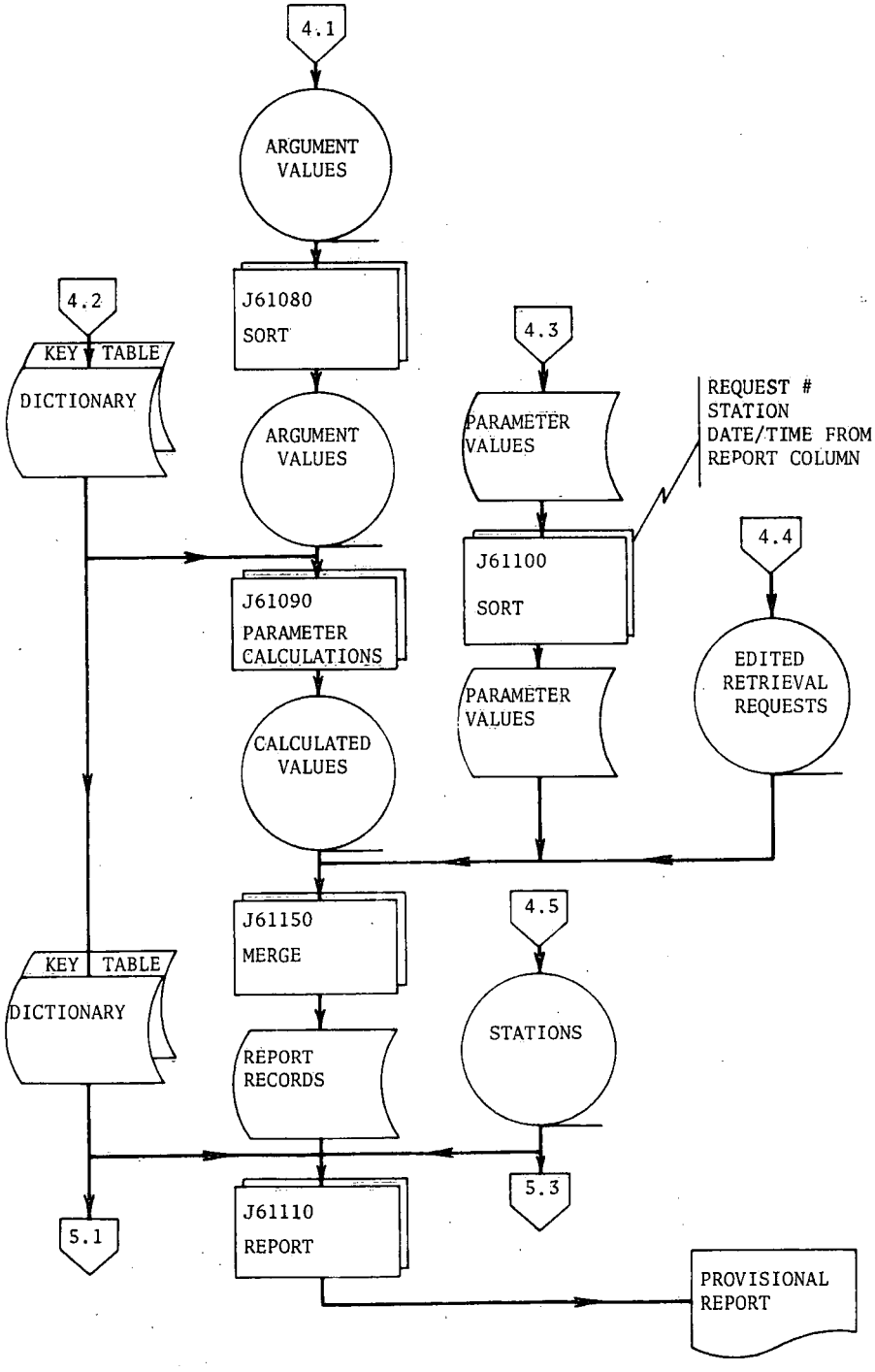

Figure 5. Computer Flow Diagram (Sheet 4).

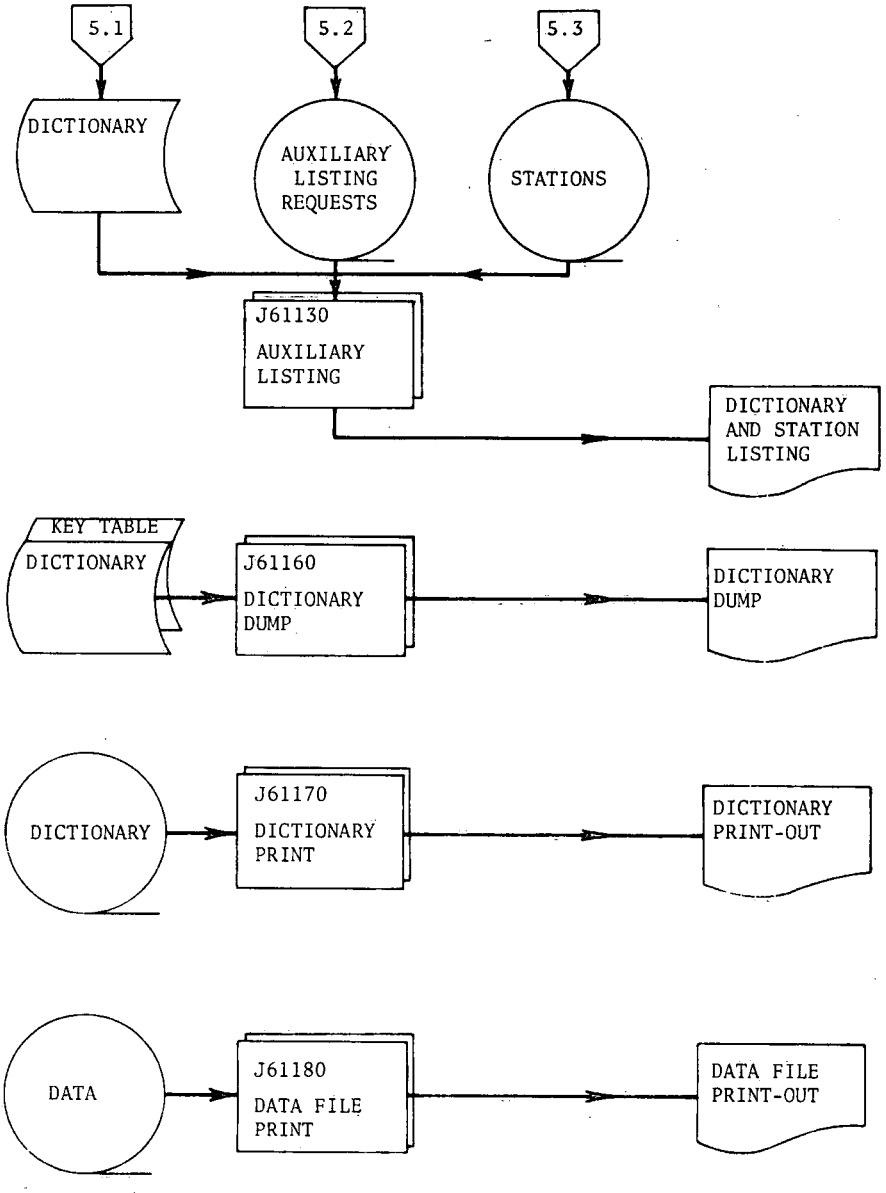

Figure 5. Computer Flow Diagram (Sheet 5).

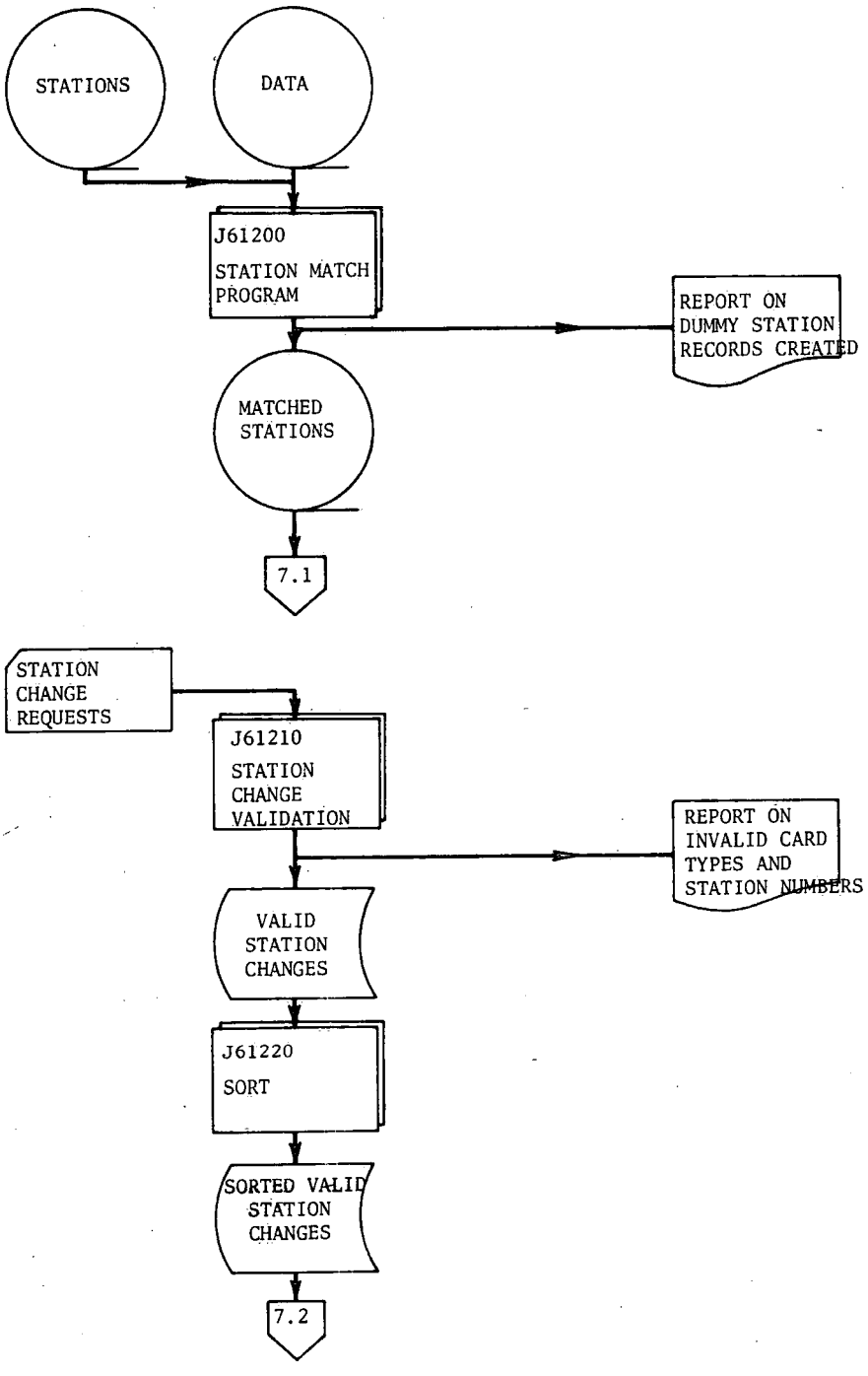

Figure 5. Computer Flow Diagram (Sheet 6).

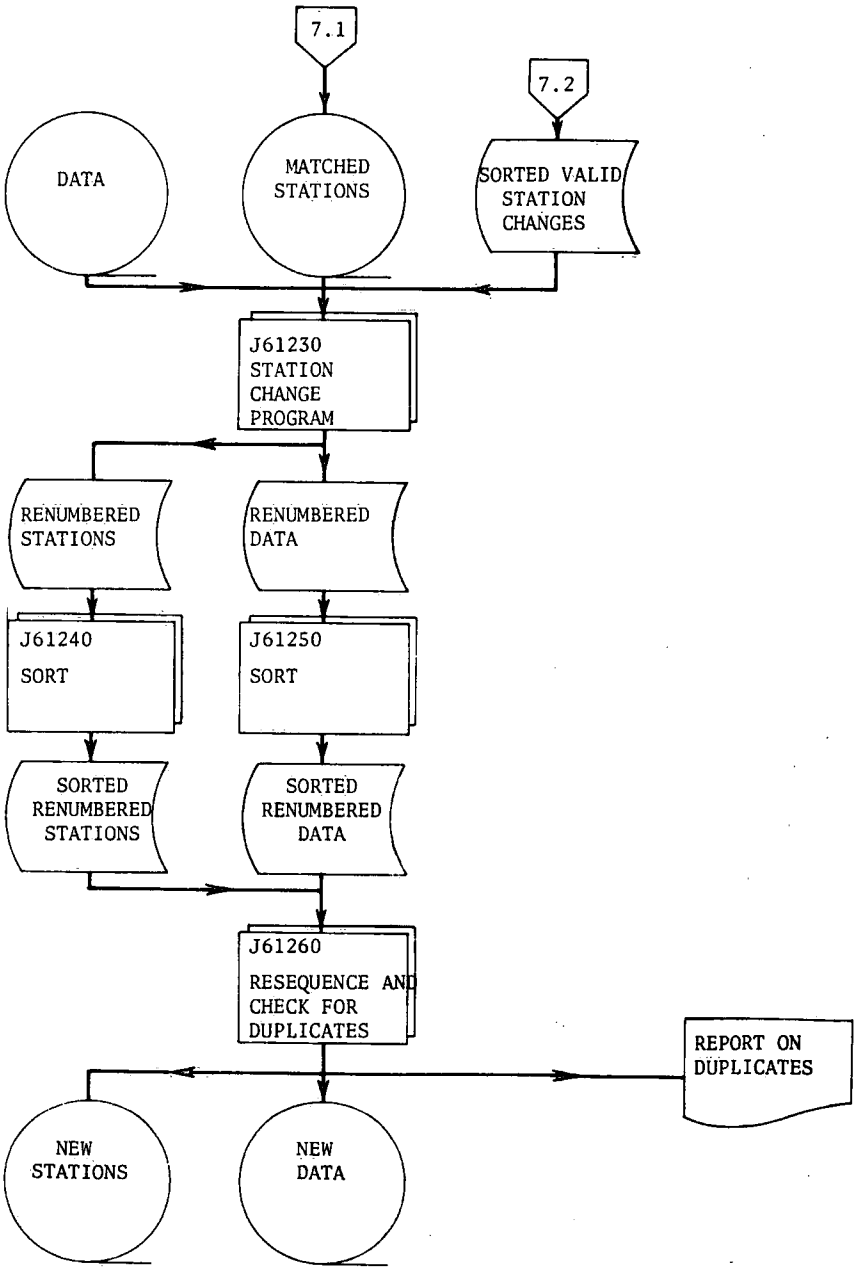

Figure 5. Computer Flow Diagram (Sheet 7).

Completing Laboratory Forms

This appendix contains details for completing laboratory forms.

STATION DATA (Form IWB-WQD-1600-1)

Station Number. All digits must be entered according to the convention explained in the text under station number.

Latitude and Longitude. All digits must be entered. In the "seconds" subfield leading blanks are not allowed; trailing blanks are allowed.

Examples:

LATITUDE (N=+)				LONGITUDE (W=+)				Comments
S	Deg	Min	Sec	S	Deg	Min	Sec	
	17	07	00.32		092	16	32.00	allowed
	17	07	.32		092	16	32.00	not allowed
	17	07	00.32		092	16	32.	allowed
	17	07	.32		092	16	32.	not allowed

Notice the implied decimal point in the "seconds" subfield

It is strongly recommended that "0's" (zeros) be entered in all cases instead of leaving blank spaces. This will avoid any misunderstanding of whether blanks are allowed or not.

The following rules apply to the "+" sign.

(i) If no sign is present the "+" sign is assumed.

(ii) If an IBM-360 system is used for processing, the 029 code must be used when key-punching the "+" sign

UTM units. When these types of unit is used to indicate the location, the whole field must be completed. Thus, leading zeros must be used; blanks are not allowed. The decimal point is implied and is shown on the respective form.

SAMPLE DATA

Sample Number. The whole field must be completed. If this is not done, the data will be rejected.

Date of Sampling. The day, month, year and zone must be completed. If "hours" and "minutes" subfields are not completed then 12.00 hours is

assumed. If the "hours" subfield is filled, then the "minutes" subfield must be completed.

When "hours" and "minutes" subfields are blank, the precision, "P", subfield must be either "5" or blank; in both cases a "5" will be carried in the file.

Frequency. If left blank a warning will be given.

Submitter's Identification: The first subfield (columns 54-57) must be completed and must be numeric, otherwise the data will be rejected. The remainder of this subfield is not checked.

APPENDIX 2

Dictionary Update Listing (From Computer Science Division)

Either all "Dictionary Update Listing" transactions may be printed or only those which are incorrect and rejected. (According to the Listing Option Code on the Batch Control Card).

Rejected transactions are printed with an asterisk to the left and one or more messages stating the source of error to the right.

The following are the error messages used:

ADDITION

Apparently an attempt was made to add data already existing on the Dictionary File.

CHANGE

No matching record.

DELETION

No matching data.

Record type 13: Computation code (col. 43-44) cannot be changed directly, i.e., by using "C". To change a "Computation Code" two separate transactions are necessary. In the first transaction the old computation code is deleted and in the second transaction the new computation code is added to the file.

UNIT CODE

Associated with record type 11. Both "From" and "To" fields must be numeric.

CONVERSION OR EPM FACTOR

Associated with record type 11: The Conversion Factor must contain a) one or more digits, b) not more than one decimal point and c) no imbedded blanks.

Associated with record type 13: The EPM Conversion Factor must consist of 6 digits.

PARAMETERS

Associated with records type 12, 13 and 14: A non-numeric character was found.

DATA DESCRIPTION

Associated with records type 12: a) the Format and/or Unit code is invalid, b) the Format must be numeric or 'A' and c) the Unit code must be numeric.

LINE NUMBER

Associated with records type 13 and 14. It may appear for two reasons: a) the number is not numeric and b) if the data are an addition, the line number was not one greater than the last one on the Dictionary.

ARC

Associated with records type 13. A character other than an A, R or C was found. A = Actual, R = Range, C = Calculated Parameters.

VALIDATION SPECS

The data are rejected when any one of the following requirements is not met: a) the data must be Calculated Parameters or Values and b) the first 5 characters of a Calculated Parameter must be numeric and the last character B, F, L or S.

If the Format specifies the values are A (alphanumeric), no validation is possible and they are accepted as is.

All other values should adhere to the following requirements: a) no imbedded blanks, b) at least one digit and c) not more than one decimal point.

If a RANGE of values is specified, none of the four values may be left blank and each one must be equal to or greater than the preceding one.

NO FORMAT AVAILABLE

Associated with record type 13 this message is a warning only. The transaction is accepted and no asterisk appears to the left of the printed line.

The Data Validation Specs cannot be used when updating the Data File - hence the warning.

COMPUTATION METHOD CODE

Record type 13: Must be numeric. If it contains more than a one line computation method code, it must appear on the first line only.

NOTES

1. Record type 13: Individual items can be D, A, C.

2. In the "Dictionary Listing": Record type 12 has the following lay-out: Unit/6 blanks/5 lines; each line = bb 12 digits bb

APPENDIX 3

Station and Data Update Listing (From Computer Science Division)

There are two types of messages printed out on the listing. Warning messages indicate possible errors or unchecked values but acceptance of the data; in this case, the data should be sight-checked carefully and if it is decided that they are invalid, a deletion should be submitted to the system to delete them from the file. Error messages are printed out with an asterisk at the beginning of the line and indicate that the entire record is rejected and must be corrected and re-submitted. If both types of message are present, the record is rejected; the errors must be corrected before re-submission but the data causing the warnings may be left unchanged if desired.

SEQUENCE ORDER

The print-out usually occurs in order, sequencing being from major to minor fields: station number, time from*, card-type, parameter* or line-number*.

Exceptions to this occur for invalid card-types or invalid time-from or to. Records containing these errors are grouped together at the end of all other data (i.e., valid or with other errors) and printed out in card-image format. Note that the invalid cards may also be dictionary, or requests, as well as station or data-file entries.

In the following explanations the messages are error messages unless otherwise stated.

INVALID CARD-TYPE

The card-type is not one of the values 01 to 05, 11 to 14, 21 to 24 or 31.

STATION NUMBER

The station number is not composed according to the given convention or, for type 04, 05 records, the given station number does not exist on the station file.

GEOGRAPHIC CO-ORDINATE

One or both of the geographic co-ordinates is not valid.

* Not every one of these fields is present on each card.

UTM CO-ORDINATE

One or both of the UTM co-ordinates is not valid.

NO LOCATION

Neither geographic nor UTM co-ordinates have been given. This is a warning if a valid reference station number is present, and an error if otherwise. Even if a warning, rejection of the record may occur if no record type-0 exists on the station-file for the given station number.

REFERENCE STATION NUMBER

The reference station number is not validly composed. If the rest of the record is valid, i.e., contains a valid station number and valid co-ordinates, only the reference station will be rejected. Otherwise the entire record is rejected.

TIME FROM

One or more of the day, month, year, hour, minute or precision is non-numeric or out of range.

TIME TO

As for time from.

FREQUENCY

Warning only. The frequency is non-numeric. A blank field will be entered on the file.

SAMPLE NUMBER

Non-numeric.

PARAMETER

Either not composed according to convention or is not present on dictionary file.

ADDITION

An attempt was made to add data already on the file.

CHANGE

No record with matching key.

DELETION

No record with matching key or data are not identical.

LINE NUMBER

For card-type 2. It may appear for two reasons: a) the number is not numeric and b) if an addition, the line number is not one greater than the last entry on the station file.

IMPOSSIBLE VALUE

1) The value associated with the given parameter is not numeric when the dictionary format specifies numeric. Valid numeric values may be 1-6 digits, the word 'TRACE' or contain at least one digit and one or more of the following special characters:

a) a leading character G or L,

b) a minus sign (preceding any digits or decimal point),

c) a decimal point.

There may be no imbedded blanks within the field.

2) The value specified does not conform to the validation requirements on the dictionary file:

a) if the dictionary format specifies an alphanumeric value, the given value does not match an actual value on the dictionary file,

b) if numeric, and there exist validation specifications in the dictionary, the given value does not match actual values and/or does not fall within the outer limits (of any) of the given range(s).

Note the following:

i) if the value does not contain a decimal point the dictionary format gives the number of decimal places.

ii) if the value contains G or L, .000001 is added to or substracted from the numeric part of the value respectively for the purpose of the comparison.

iii) if the value is 'TRACE' a numeric value of .000001 is assumed for the sake of numeric comparison.

IMPROBABLE VALUE

Warning only. The numeric value falls within one of the outer ranges but not within the inner range.

NO FORMAT

There is no format in the dictionary file for this parameter.

NO VALIDATION

Warning only. The value is a valid numeric but the dictionary contains no validation specifications; hence it has not been checked for accuracy. This applies only to parameters having a numeric dictionary format. If the dictionary format is 'A', the value will be rejected as 'Impossible'.

Completing the Retrieval Request

This appendix contains details for completing the "Retrieval Request" form.

Request Number. A request number can be any four-digit number, but no two requests can have the same number in any one batch.

Station. The range of station numbers is entered in the "From" and "To" fields. If the request refers to one station only, the two fields contain the same number.

Date. The period of time is entered in the "From" and "To" fields using numerals for day of month, month, last two digits of year, and hours and minutes. The time zone is entered in the "Zone" field.

Report Heading. This is an alpha numeric field and is printed as the report heading in exactly the same format as entered on the "Retrieval Request" form.

Column. A single request can contain up to eight columns. Some columns may be left blank.

Parameter Code Number. The main "Parameter Code Number" and up to three secondaries can be specified. If the result is not found under the main "Parameter Code Number", the computer searches for a result under the first secondary "Parameter Code Number", then the next, and so on. The main "Parameter Code Number" is printed as the first line of heading of the column. When used, the secondary "Parameter Code Number" is printed to the left of the result.

Format. Any "Format" can be specified for the analytical results printout of the retrieval report. If no "Format" is specified in the "Retrieval Request", the "Format" specified in the "Dictionary" is used.

Unit. Any unit, see Table 6, can be specified for the analytical results printout of the retrieval report, provided that the unit which is used for storing the result on the tape can be converted to the unit specified in the "Retrieval Request". Conversion factors stored at the beginning of the "Dictionary" are used for this purpose.

If the "Unit" field is left blank in the "Retrieval Request", the unit which appears in the "Dictionary" is used in the retrieval report.

Limits of Significance. If "limits of significance" are specified, then the results which are higher than the "upper" limit are printed as "High" and those which are below the "lower" are printed as "Low". If the field is left blank then the "Possible" limits stored in the "Dictionary" are used as "limits of significance".

REPORT SERIES

No. 1 PLAIN FORTRAN - A guide to compatibility in computer programming. J.J. Therrien.

* *A restricted but compatible FORTRAN intended for the scientist who does not wish to become involved in the comparison of different versions of FORTRAN to determine their compatibility.*

No. 2 North-Central Baffin Island Field Report 1967.

A synopsis of the 1967 Baffin Island operation performed by the Geographical Branch.

No. 3 The Federal Groundwater Program - Annual Project Catalogue 1968-1969.

A catalogue of current groundwater studies summarizing the objectives and progress of each project.

No. 4 Glacier Survey in Alberta. P.I. Campbell, I.A. Reid and J. Shastal.

A report on the method used and the results obtained in computing the volumetric change of two glaciers in Alberta.

No. 5 Glacier Survey in British Columbia. P.I. Campbell, I.A. Reid and J. Shastal.

A report on the method used and the results obtained in computing the volumetric change of five glaciers in British Columbia.

No. 6 Analysis of wave motion in a rectangular channel using electrical network analogy. L.F. Ku.

A clear and systematic approach to solving problems of wave motion using the method applied to electrical network analysis.

No. 7 Ice Studies in the Department of Energy, Mines and Resources - 1969.

A catalogue of current projects and a general introduction to the ice studies carried out during 1969 by the Department of Energy, Mines and Resources.

No. 8 The Federal Groundwater Program - Annual Project Catalogue 1969-1970.

A catalogue of current groundwater studies summarizing the objectives and progress of each project.

* - Out of print.

Copies of this publication may be obtained by writing to:

Director,
Inland Waters Branch,
Department of Energy,
Mines and Resources,
588 Booth Street,
Ottawa, Ontario.