

USER DOCUMENTATION FOR THE SCREEN PROGRAM (version 2.21)

1 INTRODUCTION

The screening program has been developed to support the handling of data for the ICES Environmental Data Bank. It is meant as a tool for the data managers/originators in the continuous process of validating and error-checking (in the following called 'screening') data submitted to the Data Bank.

Data submitted to the Data Bank should be in agreement with the 'ICES Environmental Data Reporting Formats' (*RF*). The program screens the data at various levels, roughly defined by

- 1) **Syntax check.** It is checked, that the sequence of the records is in agreement with the specifications in *RF*. It is also checked, that data fields for which a numeric value is required, are filled in properly.
- 2) **Internal value check.** The definition of the data fields permits some fields to be evaluated against other fields. An example is the size of a sample against the size of several subsamples. The check is that the sum of the sizes of the subsamples is equal to the total size of the sample.
- 3) **External value check.** For some data fields, their definition implies that they should take only discrete, known values, or that the value should be within a defined range. An example is the concentration of a contaminant in biological tissue. Since the unit is grammes/grammes, the value of the field should only take values in the range 0 to 1.
- 4) **Combination of fields check.** For some data fields, it is known, that they should only be filled in, if certain other fields are filled in. An example is the standard deviation of some measured value. It should only be filled in, if the mean value is filled in.

The **external value check** is done on the basis of predefined values. Valid values for several of the fields are found in files supplied together with the program (file name extension **.lis*).

The program generates three files, a report file (*.rep*), an error file (*.err*) and a warning file (*.wrn*). The report file reports the results of the screening procedure as performed on individual data fields or data records. If a disagreement with *RF* has been encountered this is reported as an *error-message* in the error file or a *warning-message* in the warning file. Both messages are signals, that the data need a more careful inspection. **For a data file to be considered as valid, the program should not generate any *error-messages* when the data file is processed by the program.**

2 THE PROGRAM

The program screens data for agreement with *RF* version 2.2 from December 1994.

The program runs on an IBM-compatible PC with a hard disk under MS-DOS (PC-DOS) version 3.10 or higher.

2.1 Installation

The program is distributed on a diskette from the ICES Environment Secretariat. The diskette contains the self extracting file **ScXXX_YY.EXE**, where

XXX indicates the program version number, and

YY indicates the release number.

These values should be noted for inclusion in the data files.

To install the program, create a directory \SCREEN on the hard disk, and copy the file to this directory. Issue the command **ScXXX_YY** at the DOS-prompt (replace XXX and YY with the appropriate values), and the executable file and the auxiliary files will be extracted.

The **.lis* files can be inspected with an editor, and printouts can be made, **but they should not be modified**. If a file is changed by accident, all files can be reinstalled as described above.

2.2 Run the program

To run the program, go to the directory \SCREEN, and give the following command at the DOS prompt:

```
C:\SCREEN>screen [options] <data-set-qualifier> <input-file>
```

or

```
C:\SCREEN>screen -i
```

2.2.1 Screen a data file

In the first instance, the program is used for data file screening, where

input-file

is the name of the data file to screen. Drive and path specifications could be included,

data-set-qualifier

is either *b*, *s*, *w*, or *d* for biota, sediments, water or fish disease data, respectively,

options

could be omitted or be one or both of

-R <*filename*>: output report to <*filename*>. Default is <*input-file*>.rep

-E <*filename*>: output errors to <*filename*>. Default is <*input-file*>.err

-W <*filename*>: output warnings to <*filename*>. Default is <*input-file*>.wrn

-e <*count*> (default value is **100**)

-r <*report-file*> (default value is **screen.out**).

The *data-set-qualifier* and *options* are case-sensitive.

An example: The data file **ICES91CF.IS** should be screened. It contains data on contaminants in biota. The program should stop, when 200 error-messages have been reported. The following command should be issued at the DOS prompt:

```
C:\SCREEN>SCREEN -e 200 b ICES91CF.IS
```

2.2.2 Report program release number

In the second instance, the program reports the release number of the program. The program produces a list of the *.*lis* files associated with the program. The first line in the list says 'SUPER', and the release number associated with this entry is the same as the release number of the program.

3 INPUT

Input data files should be MS-DOS (PC-DOS) ASCII files with a structure as described in *RF*. It should be noted that the first record should look like this (depending on the actual version and release of the program):

```
00 RF2.2 SV2.21 LR01
```

where

00 indicates that this record is the file header (first record in the data file),

RF2.2 indicates that the data are reported according to *RF* version 2.2,

SV2.21 LR01 indicates that the data is screened using version 2.21, release number 01 of the screening program.

4 OUTPUT

Three output files are generated: a warning file and an error file, both containing messages that SCREEN reported for the data file; the third file, the report file, gives a summary of the data file content.

4.1 The report file

The report file has five subsections, each covering its own aspect of the data file content. The five sections are described in the following. The examples shown are output from the screening of a data file with biota data. The output generated by other data types is similar to this.

4.1.1 The sample/subsample/tissue information

The output in this subsection could look like this:

Sample reference	Sub No.	Tissue	Type 10	Groups	Species	Purp.
ALRP880001	B01	1-LI	27O..	PLAT FLE	T
	H02	1-MU	30	..O..		
		2-LI	4		
		1-LI	27O..		
	B03	1-LI	27O..		

Each line summarises information at the *sample/subsample/tissue* level as follows:

Sample reference gives a reference to the sample. It is a combination of the data fields **RLABO**, **MYEAR**, and **SEQNO**. In line 1 **RLABO**=‘ALRP’, **MYEAR**=‘88’, **SEQNO**=0001.

Sub No. gives a reference to the subsample. It is a combination of the data fields **INORB** and **SUBNO**. In line 1 **INORB**=‘B’, and **SUBNO**=01.

Tissue gives a reference to the tissue type. It is a combination of a sequence number and the data field **TISSU**. In line 1 the sequence number is 1 and **TISSU**=‘LI’.

Type 10 gives the number of parameters/contaminants measured for this particular combination of sample/subsample/tissue. In line 1 the number of parameters/contaminants measured is 27.

Groups gives a reference to the type of parameters/contaminants measured. In line 1 only organics are measured, while in line 2 both organics and inorganics are measured.

Species is identical to the data field **SPECI**. In line 1 the species ‘PLAT FLE’ (*Platichthys flesus*) is analyzed.

Purp. is identical to the data field **PURPM**. In line 1 the purpose code is ‘T’.

4.1.2 The analytical methods information

The output in this subsection could look like this:

ANALYTICAL METHODS:

LinkKey	PARAM	METAN	DETLB	DETLI	ICCOD	CONCH	CRMCO	CRMMB	CRMMV	CRMSD	CRMNM	PE	RBMEA	ZSCOR	PSCOR	Freq
ALRPALRQ88	CB10101	001	L	3.0000E-09	2Z	CRM	CRM01	L	7.0000E-09	5.0000E-09	3	5				289
ALRBALRQ88	CB11801	001	L	4.0000E-09	2Z	CRM	CRM02	L	7.0000E-09	5.0000E-09	4	7				288

Each line summarises the information given in a particular *analytical method* record (21-record) as follows:

Link Key/PARAM gives the label of the analytical method. It is a combination of the data fields **RLABO**, **ALABO**, **MYEAR**, **PARAM**, and **AMLNK**. In line 1 **RLABO**=ALRP, **ALABO**=ALRQ, **MYEAR**=88, **PARAM**=CB101, and **AMLNK**=01.

The headings *METAN*, *DETLB*, *DETLI*, *ICCOD*, *CONCH*, *CRMCO*, *CRMMB*, *CRMMV*, *CRMSD*, *CRMNM*, *CRMPE* have the same definition as specified in *RF*.

Freq gives the number of 10-records that has a reference to the particular 21-record.

4.1.3 The sampling location information.

The output in this subsection could look like this:

AREAS

Coordinates latitude	longitude	JMP area	ICES area	Other area	Freq.
53 15.00 N	06 30.00 E	J10	35F65	47	1
54 10.00 N	06 15.00 E	J15	30F71	50	1

Each line summarises the information given in a particular *sample master* record (01-record) as follows:

Coordinates latitude and *longitude* gives the position of the station/sample. It is a combination of the data fields **LATDG**, **LATMI**, **LATMF**, **LONDG**, **LONMI**, **LONMF**, and **QEORW**.

JMP area is identical to the data field **JMPAR**. In line 1 **JMPAR**='J10'.

ICES area is identical to the data field **ICEAR**. In line 1 **ICEAR**='35F65'.

Other area is identical to the data field **OTHAR**. In line 1 **OTHAR**=47.

Freq. gives the number of 01-records with this particular location.

4.1.4 The summary.

The output in this subsection could look like this:

RECORD COUNT	
Master Record.....	: 46
Specimen/Sub-sample Record.....	: 404
Tissue Record.....	: 601
Parameter/Contaminant Record.....	: 9818
Contaminant Analytical Record.....	: 46
EROD Methods Record.....	: 0
Plain Language Record.....	: 1097
Total no. of records	12012

The table summarises the content of the data file. The number to the right gives the number times a particular record type appears in the data file.

4.1.5 The min/max section.

The output in this subsection could look like this:

PARAM	MATRX	Min.Value	Max.Value
PCB	LI	13 E-07	211 E-06
SDDT	MU	53 E-08	201 E-06

The table gives the minimum and maximum values (concentrations) of all parameters (contaminants) found in the data file.

4.2 The error/warning part of the output

The error and warning files give the results of the screening procedure as performed on individual data fields or data records. If a disagreement with *RF* has been encountered this is reported as an *error-message* or a *warning-message*.

An *error-message* is generated when the data is in disagreement with the specifications as given in *RF*. An example could be that the sequence of data records is wrong. The data file should be corrected before it can be considered as valid. The error messages are collected in the error (*.err*) file.

A *warning-message* is generated if something unusual is found in the data file. An example could be that a measured concentration is far higher than the average known for the contaminant. This is a signal that the data could need a second inspection. The warning messages are collected in the warning (*.wrn*) file.

A note should be given on message number 071: 'invalid BASIS/MATRX/PARAM combination or not previously reported. Contact ICES with valid ranges.' This message indicates that a combination of BASIS/MATRX/PARAM is included in the data file that has not previously been reported to the ICES Environmental Data Bank. The Secretariat should be contacted with information on the range within which the VALSNs of this particular BASIS/MATRX/PARAM is expected to be.