

A PASCAL INTERFACE FOR DATA VALIDATION OF CDS/ISIS DATABASE RECORDS

*M. Krishnananda Herala, A. Sreenivasa Ravi,
and T.B. Rajashekar**

CDS/ISIS, an information storage and retrieval package produced and distributed by UNESCO, is widely used in Europe and many developing countries. It supports most of the requirements for designing bibliographic/textual databases. The CDS/ISIS Pascal programming facility can be used to create end-user interfaces for any specific additional requirements. The standard CDS/ISIS package does not provide for data validation with respect to subfielded fields and to check whether any mandatory fields have been missed (left blank) during data entry. This paper discusses the need for data validation and operation of a CDS/ISIS Pascal interface designed for data validation of CDS/ISIS databases. The programmes enable specification of mandatory fields, validation of such fields and inadvertent omission of subfield delimiters in the master records, and obtaining online help about the use of the interface programmes.

1 INTRODUCTION

Micro CDS/ISIS, developed and distributed by UNESCO free of cost to non-profit organizations, is widely used in Europe and many developing countries including India.

Micro CDS/ISIS provides for all requirements for the creation, maintenance and retrieval from structured textual databases. It can handle variable length fields and records, supports subfields and repeatable fields, data export/import in ISO 2709 format, etc. However, data entry must be done carefully, not to miss subfield and repeatable field delimiters where used, as these require simultaneous depression of more than one key on the keyboard.

In bibliographic and functionally related type of databases certain fields may be deemed mandatory, and therefore data must be entered in such fields. However, the data entry operator may fail to enter data inadvertently or the corresponding fields are not in some way identified or recognized as mandatory. Data may also not be entered when importing records from another database due to errors in the conversion FST, or due to deficiencies in imported records. The standard Micro CDS/ISIS package does not provide the means to check such omissions in a database.

* *Contact:* Dr. T. B. Rajashekar, National Centre for Science Information, Indian Institute of Science, Bangalore 560012, India. E-mail: raja@ncsi,iisc.ernet.in

In this paper we describe three interfaces developed using CDS/ISIS Pascal language, to facilitate data validation in Micro CDS/ISIS database records, to overcome these limitations of the standard package.

2 TYPES OF ERRORS IN DATA ENTRY

Micro CDS/ISIS requires specification of subfield delimiters, repeatable field, length of field, type of field, etc., at the time of database definition (in the corresponding FDT). However, it performs data validation only for the type of field (numeric, alphabetic and pattern) at the time of data entry. A subfield delimiter in Micro CDS/ISIS consists of two consecutive elements: the caret sign (^) followed by an alphabetic or numeric character. While entering data, common typographical errors, such as, not typing the caret sign or the following character or both, occur. Omission of the caret (^) sign may lead to non-indexing and non-retrieval of the data in the corresponding subfield. If the character following the caret sign is omitted, CDS/ISIS will consider the first character of the data entered in the subfield as part of the delimiter.

It is a common practice to deem certain fields of a database as mandatory and data must be entered in them. For instance, the CCF (Common Communication Format) for bibliographic and related type of functional databases, suggest a list of mandatory fields. According to the local requirements, one may select some or all of these as mandatory fields, and may also add others to these. Micro CDS/ISIS permits identification of such fields in a data-base in the corresponding worksheet(s) as follows:

1. By providing an appropriate help message (e.g. Mandatory) for the corresponding field(s), which will be displayed when <F1> key is pressed (for each field) at the time of data entry. Remember, there may also be other help messages for each field (e.g. Repeatability; subfield delimiters used and what is to be entered in each sub-field; examples of rendering data, etc.).
2. By choosing a screen attribute for the mandatory field different from that for the non-mandatory fields. This is done at the time of preparing the worksheet(s).
3. By marking in the Field Name of a mandatory field, say with an asterisk (*) or (M). For example: TITLE (M) or TITLE*.

As already mentioned, the type of data and pattern of entry specified in the FDT also helps data validation to some extent.

By adopting one of the above methods the person entering data will be alerted about mandatory fields so that these fields are not left without data. Nevertheless, as mentioned in section 1, data may not be entered in a mandatory field inadvertently or on importing records from another database. Hence the need for data validation of mandatory fields.

3 DATA VALIDATION PROGRAMMES

3.1 Purpose

The CDS/ISIS Pascal programmes described here enable offline data validation and correction after identifying and listing the types of errors mentioned above.

Three programmes are discussed:

- VALIDT.PAS performs data validation of records;
- SPCMAN.PAS permits defining of mandatory fields for a database; and
- VALHLP.PAS displays a one-screen description of what the Pascal programmes can do.

The Pascal Programmes:

Allow declaration of mandatory fields for a database.

Indicate any blank/marked for deletion or deleted records.

If every mandatory field is to be checked

- search for missing subfield identifiers in each occurrence of a subfielded field, which has been actually entered.
- search for any empty mandatory field.

If mandatory fields are not to be checked,

- search only for missing subfield identifiers.
- do not search for mandatory fields, which may be purposefully left blank.

If the errors are to be written to the output file,

- errors encountered, as above, are entered into a file dbn.ERR, dbn being the name of the database being validated.

- errors not displayed on the screen, but can exit to modify the records.

If the errors are not to be written to an output file,

- errors found are displayed on the screen only.
- only the record which contains errors will be indicated.
- can exit the search programme for modifying a record at any stage.

Provides a one-screen description of what the Pascal Programmes can do.

3.2 Installation of the Programmes

The three programmes should be copied into the c:\>isis\prog sub-directory or as specified by parameter 1 of the SYSPAR.PAR file of the database(s) to be validated. Each programme should then be compiled to create the corresponding dbn.PCD files (namely, VALIDT.PCD, SPCMAN.PCD and VALHLP.PCD respectively). Option **A - Advanced Programming Services** in the Main Menu is to be used for the purpose.

It will be convenient to call the programmes as Menu exits, that is by selecting an option digit from a Menu. For example, we may create a new line **V - Data validation** in the Menu **EXE1 Data Entry Services** as shown in Fig. 1.

Service ISISENT	Data entry Services	Menu EXE1
	L - Change dialogue language	
	W - Select another worksheet	
	N - Create new record	
	E - Edit record (or range)	
	R - Edit last search results	
	D - Define default values	
	P - Recall last record modified	
	C - Clear default values	
	V - Data validation	
	X - End Data Entry	
	?	-
Data base: JRNART		Worksheet: JRNAR
Max MFN : 134		Format : JRNART

Fig. 1: Modified Data Entry Services Menu

Selecting option V may call another Menu **EXVAL Data Validation** as shown in Fig. 2.

Service ISISVAL	Data Validation	Menu EXVAL
S - Specify mandatory fields V - Validate database records(s) H - Help for data validation X - Exit to Data Entry Services		
? _		
Data base: JRNART		Worksheet: JRNAR
Max MFN : 134		Format : JRNART

Fig. 2: Data Validation EXVAL Menu

Creation of the EXVAL Menu and modification of the EXE1 Menu can be done by selecting option U - **System Utility Services** in the Main Menu. The EXVAL Menu and the modified EXE1 Menu can be obtained along with the Pascal programmes from the authors. In that case they should be copied to the subdirectory C:\>isis\menu or as indicated by papameter 2 of the SYSPAR.PAR for the database to be validated.

3.3 Specifying Mandatory Fields

Select option E - Data entry services in the Main Menu. Type in the name of database to be validated at the prompt 'Data base name' if not already selected. Select option V - **Data Validation** in EXE1 Menu. The system will display EXVAL Data validation Menu.

Fields may be declared mandatory for any CDS/ISIS database. The CCF or MIBIS manuals may be used as guides in selecting the mandatory fields in a bibliographic database. However, depending on local needs additional fields may be declared mandatory or some of those suggested in the manuals may not be declared as mandatory.

Select option S - **Specify mandatory fields** in Menu EXVAL to execute the programme SPCMAN.PCD. The system will display the screen in Fig. 3, listing the first Tag and Field name at the bottom left of the

screen. Press <Esc> key if the field listed is to be specified as mandatory, and it will be listed under the title Mandatory Fields (see Fig. 4). Pressing the <Enter> key will list the next field at the bottom left. In this manner each of the fields in the FDT will be listed and selected as mandatory by pressing <Esc> key or skipped by pressing <Enter> key. When all the fields have been called and decided upon, the message - ***** Press <Enter> to get back** will be displayed. When a field is specified as mandatory by pressing the <Esc> key, the Tag of the field is written in a file dbn.MDT, where dbn stands for the name of the database. This file, written in the \isis\data sub-directory, is used by the system for data validation.

Mandatory Fields

- 100 Authors

<Enter> Next field <Tab> Previous field <Esc> Select as mandatory

Fig. 3. Mandatory fields specification

Mandatory Fields

100 - Authors

200 - Title

- 300 Journal details

<Enter> Next field <Tab> Previous field <Esc> Select as mandatory

Fig. 4. Listing of fields declared mandatory

3.4 Validation of Database Records(s)

The programme VALIDT.PCD performs data validation taking into account the subfield delimiters defined in the FDT at the time of database

definition, and the mandatory fields specified using SPCMAN.PCD. It checks for data integrity against the FDT of the database and the mandatory file dbn.MDT and lists errors, if any. An approximate algorithm for the programme is given below.

Open the desired database

Get the MFN or MFN range for Data Validation from the user

Ask whether the user wants to validate all mandatory fields

If so, then

Find whether the mandatory file exists

If it does not exist, prompt the user to create it and stop

Else get the field tags of mandatory fields from 'dbn.MDT' file

Ask whether the user wants to write errors in an output file

Get the field name, subfield delimiters, field tag, etc. of all fields from the 'dbn.FDT' file

For each MFN (in the MFN range specified by the user)

Get the corresponding record

Check whether mandatory and/or subfielded fields are entered (contain data)

If mandatory fields are left blank, indicate it on the screen or in the output file 'dbn.ERR'

If a subfielded field is entered, then

Check whether subfield delimiters are correctly entered in each occurrence of the field

If not, indicate it on the screen or in the output file 'dbn.ERR'

Ask whether the user wants to exit for modifying the record having such errors

If so, exit the program.

The programme can be invoked by selecting option **V - Validate database records** in Menu **EXVAL - Data validation**. It will ask for the first MFN or MFN range of the records to be data validated. After typing in the MFN or the range in the usual CDS/ISIS pattern and on pressing <Enter>, the system will display the prompt **Want to validate all mandatory fields (Y/N)?**, that is in addition to checking subfield delimiters.

If the response is **Y**, and if mandatory fields have not yet been specified, the programme will require that they be specified (using option **S**) and exits. Otherwise, the prompt **Want to write errors to output file (Y/N)?** will be displayed. Pressing **Y** and <Enter> keys one after the other, will execute the programme. The MFN of the record being validated will be displayed on the screen. Errors, if any, will be

stored in file **dbn.ERR** in the data sub-directory (visis\data). The file can be printed out and the errors corrected. (See Fig. 5)

If the response is **N**, the programme will be executed as before, but the error list will be displayed with the prompt **Want to exit for modifying the record (Y/N)?**. (See Fig. 6). A **dbn.ERR** will not be stored. The user can respond with **Y** and exit to correct the record, or respond with **N** and continue the validation.

Data Entry Errors

- MFN No.: 5 Tag 400 "Month & Year" -> Subfield Delimiter ^A is missing in occurrence 1.
- MFN No.: 7 -> Marked for deletion.
- MFN No.: 8 Tag 100 "Authors" -> Subfield Delimiter ^A is missing in occurrence 1.
- MFN No.: 10 Tag 400 "Month & Year" -> Subfield Delimiter ^B is missing in occurrence 1.
- MFN No.: 18 -> Marked for deletion.
- MFN No.: 21 ->Marked for deletion.
- MFN No.: 28 Tag 100 "Authors" -> not entered.
- MFN No.: 32-> Marked for deletion.
- MFN No.: 39 Tag 100 "Authors" -> Subfield Delimiter ^B is missing in occurrence 1.
- MFN No.: 52 Tag 500 "Keywords"-> not entered.
- MFN No.: 67-> Marked for deletion.
- MFN No.: 83-> Marked for deletion.
- MFN No.: 88 Tag 500 "Keywords"-> not entered.
- MFN No.: 123 -> Marked for deletion.
- MFN No.: 128 Tag 400 "Month & Year" -> Subfield Delimiter ^B is missing in occurrence 1.
- MFN No.: 129 Tag 400 "Month & Year" -> Subfield Delimiter ^B is missing in occurrence 1.
- MFN No.: 131-> Marked for deletion.
- MFN No.: 133 -> Marked for deletion.
-

Fig. 5. A dbn.ERR file printout

MFN No.: 124

Tag 500 "Keywords" -> not entered.

Tag 400 "Month & Year" -> Subfield Delimiter ^B is missing in occurrence 1.

Tag 100 "Authors" -> Subfield Delimiter ^A is missing in occurrence 2.

Want to exit for modifying the record (Y/N)?

Fig.6. Display of error list on screen

3.5 Help on Use of the Programmes

Pressing option **H - Help for data validation** in the Menu **EXVAL** **Data validation** will display a screen-page of hints on the use of the programmes. (see Fig. 6).

Help for Data Validation

- * Allows declaration of mandatory fields for a database.
- * Indicates any blank/marked for deletion or deleted records.
- * If every mandatory field is to be checked
 - searches for missing subfield identifiers in each occurrence of a subfielded field, which has been actually entered.
 - searches for any empty mandatory field.
- * If mandatory fields are not to be checked,
 - searches only for missing subfield identifiers.
 - does not search for mandatory fields, which may be purposefully left blank.
- * If the errors are to be written to the output file,
 - errors encountered, as above, are entered to <DATA-BASE_NAME.ERR>.
 - cannot see the errors, if any, on the screen and exit in between for modifying the record.
- * If the errors are not to be written to the output file,
 - errors found are displayed on the screen only.
 - only the record which contains errors will be indicated.
 - can exit the search program for modifying a record at any stage.

Fig. 6. Help Screen

4 FEEDBACK

The authors will be glad to receive feedback from users of these programmes for carrying out programme modifications, if necessary. We would also like to learn about other data validation programmes for CDS/ISIS databases.

REFERENCES

1. Simmons, Peter; Hopkinson, Alan (eds.). *CCF: The Common Communication Format*. 2nd ed. Paris: UNESCO General Information Programme and UNISIST; 1988. (PGI-88/WS/2).
2. Di Lauro, Anne; Brandon, Ed. *MIBIS Manual for Preparing Records in microcomputer Based Bibliographic Information Systems*. 2nd ed. Ottawa, Canada: IDRC; 1995 May.
3. UNESCO. *Mini-Micro CDS/ISIS Reference Manual*. Version 2.3. Paris: UNESCO; 1989.