

# Double-Byte Character Sets

This document is only relevant for Asian countries which use double-byte character sets. It describes all features implemented in Natural to support DBCS terminals and printers and covers the following topics:

- Natural Profile Parameter SOSI
  - Output Format Specification
  - Parameter Definitions for DBCS Support
  - Editor Profile Options
  - Input Data Check
  - Output Data Adjustment
  - Natural Stack Data
  - Application Programming Interfaces for DBCS Handling
- 

## Natural Profile Parameter SOSI

In alphanumeric fields with SBCS and DBCS characters mixed, the DBCS character strings are separated from the SBCS strings by shift codes called SO (shift-out) and SI (shift-in). The Natural profile parameter SOSI is used to pass the values of the shift-in and shift-out codes used in the current environment to Natural.

It is strongly recommended to use the IBM characters X'0E' and X'0F' internally. With this technique, all applications and data can be handled in a compatible manner, which means that a network supporting different mainframe types can still use the same Natural applications and process the same data.

For detailed information on this parameter, see SOSI.

## Output Format Specification

The Natural session parameter PM=D is used to define DBCS-only fields. A DBCS-only field must contain only valid DBCS characters; shift-out/shift-in characters (SO/SI) are not allowed within such a field. To display a field with the session parameter PM=D specified, the screen attribute X'43F8' is added for IBM terminals; for Fujitsu terminals, the field content is enclosed in the required shift-out/shift-in characters (SO/SI).

## Parameter Definitions for DBCS Support

The following parameters must be specified in the setup for Natural for the support of double-byte character sets:

Parameter	Explanation
TS=ON	If Latin lower-case characters are not available, this parameter translates all Natural system output using the translation table defined by the macro NTTABL in the NATCONFIG module.
SOSI=( 0E, 0E, 0F, 0F, 1 )	Defines the DBCS shift-out and shift-in values for IBM hardware.
SOSI=( 28, 28, 29, 29, 0 )	Defines the DBCS shift-out and shift-in values for Fujitsu hardware.
LC=ON	Does not translate all input data to uppercase, which again would destroy possible DBCS input data.

In addition to TS=ON, further parameters to provide for translation of messages into upper case are provided by several Natural components. For detailed information, see *Other Parameters to Provide Upper Case Translation* in the TS profile parameter documentation.

## Editor Profile Options

If you want to enter DBCS or half-width Katakana characters in one of the Natural editors, the following editor general default options should be set in the editor profile to avoid that character constants or field names containing DBCS or half-width Katakana characters are unintentionally converted to upper case:

Option	Value	Explanation
Editing in Lower Case	Y	Lower-case characters in the source code are not automatically converted to upper case. This option is required if you are using DBCS or half-width Katakana characters.
Dynamic Conversion of Lower Case	N	Any source code remains as you enter it. This option is required if you are using half-width Katakana characters.

For detailed information on the editor general default options, see *General Defaults*. For detailed information on the editor profile, see *Editor Profile* in the *Editors* documentation. To avoid the need to change these options for every user, you can modify the default profile for your installation by means of the user exit routine USR0070P, which also supports DBCS; see *USR0070P - User Exit for Editor Profiles* in the section *Configuring Natural*.

## Input Data Check

If the session parameter PM=D is set for a field, it is verified that the input data

- contains an even number of bytes,
- contains only valid DBCS characters,
- does not contain shift-out/shift-in characters (SO/SI).

Because the detection of non-DBCS characters requires ICU, this check will not be performed if ICU is not available (that is, if the profile parameter CFICU=OFF has been set).

## Output Data Adjustment

If a window is to be displayed for user interaction, the window might overlay DBCS characters that are already displayed, or the window might itself contain DBCS characters which are truncated because of the window size. An overlay may also occur if the `NO ERASE` option is used with an `INPUT` statement. In order to prevent screen corruption in case of such an overlay, the following actions are performed to adjust the output data, if necessary:

- if the session parameter `PM=D` is set for a field, an orphan byte (that is, a single byte left at the beginning or end of the data to be displayed as a result of a partial overlay of a DBCS character) is replaced by an attribute; this operation assures that only valid DBCS characters are displayed;
- if the profile parameter `SOSI` has been set, the field contents of an alphanumeric field for which `PM=D` is not specified is examined for shift-out/shift-in characters (`SO/SI`); if a shift-out character (`SO`) is found for which the correlating shift-in character (`SI`) is missing, either the last character of the output data is replaced by a shift-in character (`SI`) or the last two characters are replaced by a shift-in character (`SI`) followed by a blank; if a shift-in character (`SI`) is found for which the correlating shift-out character (`SO`) is missing, either the first character of the output data is replaced by a shift-out character (`SO`) or the leading two characters are replaced by a blank followed by a shift-out character (`SO`); this operation assures that DBCS characters are enclosed properly by shift-out/shift-in characters (`SO/SI`).

## Natural Stack Data

To avoid unintentional interpretation of DBCS characters as delimiter or control characters, the `FORMATTED` option of the `STACK` statement should be used if the data to be placed on the Natural stack contains DBCS characters.

See the *Statements* documentation for further information on the `STACK` statement.

See the *Programming Guide* for further information on the Natural Stack.

## Application Programming Interfaces for DBCS Handling

The following user application programming interfaces (API) are available to support DBCS handling:

- `USR4211N` - Get DBCS Characters
- `USR4213N` - String Handling for DBCS Support

These APIs are contained as subprograms in the Natural library `SYSEXT`. Detailed information on how to use an API is included in the corresponding text member (`USRxxxxT`). See also *SYSEXT Utility - Natural Application Programming Interfaces* in the *Utilities* documentation.

### **USR4211N - Get DBCS Characters**

The application programming interface `USR4211N` can be used to obtain information on the availability of DBCS support and the defined `SOSI` characters.

## USR4213N - String Handling for DBCS Support

The application programming interface USR4213N can be used to perform the following functions:

- Convert a normal Latin character string into the corresponding DBCS character string.
- Convert a DBCS character string that contains Latin data only into a single-byte character string.
- Add the current shift codes at the beginning and at the end of a character string.
- Remove leading and trailing shift codes from a character string.

The last two functions can be used to either produce native DBCS strings or generate mixed-mode data out of native DBCS strings.